

# 图 象 理 解 及 其 推 理 机 制

丁德恒 吕维雪 (浙江大学科仪系)

## 摘 要

在研究实用图象理解系统过程中,人们逐步认识到推理机制的重要性。本文从推理方式的角度讨论图象理解中存在的问题及其产生原因。重点分析多重信息融合方法和假设推理过程。

图象理解过程一般可分成两大步骤:

(1) 分割 (Segmentation) 从输入图象中抽取图象特征 (边缘或区域), 形成图象的结构描述。

(2) 匹配 (Matching) 将图象特征与待识别对象的模型进行匹配, 形成对图象所表示的景物的描述。

当我们要构造一个图象理解系统 (Image Understanding System), 对复杂景物能进行稳定分析时, 仅仅孤立地研究分割或匹配的个别算法是不够的, 还必须将分割与匹配这两个处理过程有机地结合起来考虑<sup>[1]</sup>。也就是说, 必须研究贯穿于这两个过程中的推理方式和控制机制。由于在一个复杂的场景中, 各种物体交错排列, 照明环境难以描述, 阴影部分或高光强部分受相互反射的影响极大, 所以各算法所基于的模型 (如边缘检测的Step-Edge model, Shape from shading的漫反射模型) 只适用于图象的局部。对于理解复杂景物, 必须使用局部不同的图象、照明、物体模型。为此, 不仅要在匹配过程中计算图象特征与识别对象的

相似度, 还必须生成对图象的哪一局部应采用何种分析模型的控制信息和解析规则, 以指导整个分割过程。

## 一、图象理解中存在的问题

图象理解中产生的主要问题有:

(1) 分割问题

(1a) “有意义”的图象特征有时被漏检;

(1b) “无意义”的信息有时被误检测为图象特征。

(2) 匹配问题

(2a) 与某一图象特征相匹配 (往往是部分匹配) 的识别对象模型有多个;

(2b) 由于遮挡和阴影, 使识别对象的某些局部被消去, 故只能进行不完全匹配;

(2c) 因图象特征的模糊性, 需要有评价匹配优劣的标准。

一般来说, 分割过程中存在的上述两个问题是不可避免, 因为:

为复杂对象。

这就引发了人们对面向对象的程序设计与复杂对象数据库综合、逻辑程序设计与复杂对象数据库综合的研究兴趣。这两方面的研究对开发强功能的程序设计语言、强功能的知识系统具有重要意义。

事实上, O<sub>2</sub>中就已强调了前一方面, IRIS及ALGRES中则强调了后一方面; 尽管如此, 这两方面的工作, 尤其是后一方面的工作, 仍只是刚刚起步。

(参考文献共106篇略)

(1) 图象信号是否表示“有意义”的特征, 仅靠信号自身所具有的信息是不能决定的, 它还随周围环境、处理目的而变化。因此, 若用纯粹的信号分析方法来处理分割问题, 必然会产生 (1a)、(1b) 两个问题;

(2) 图象的结构或质量并非均匀一致。若用基于单一模型所设计的算法, 对图象进行均匀处理, 也必然导致 (1a)、(1b) 问题产生。

在限制识别对象种类的情况下, (2a) 就变成了怎样以最好的效率来寻找与图象特征最佳匹配的对象识别效率问题, 亦为检索问题 (Indexing Problem)。另外, 二维图象特征与三维模型的匹配, 一般会存在无限多个对应点, 形成多种可能的对应, 若采用一般的效率问题求解是不够的。尤其是从二维图象抽取三维信息的计算机视觉 (Shape from Shading、Contour、Texture等) 问题中, 必须求解不定解 (Underconstrained) 方程。由于在图象理解中, 图象数据所包含的信息量往往比理解景物所需的信息量少 (输入信息的不充分性), 所以也导致了 (2a)、(2b) 问题的产生。

(2c) 问题不仅在图象理解中, 而且在基于定量测量数据进行推理的专家系统中也是一个非常重要的问题。但图象理解过程需要进行基于对象物的几何学关系的空间推理, 所以既要表示和处理图象特征和识别对象属性所包含的模糊性, 还必须表示和处理这种空间关系所具有的模糊性。

综上所述, 在图象理解中必须充分考虑不确定性, 进行基于不完全信息的分析推理。仅仅采用搜索算法、纯粹的符号处理、传统的逻辑推理来构造图象理解系统是远远不够的。

## 二、基于不完全信息的分析推理

为了解决第一节所描述的问题, 我们应

该从图象信息的处理、获取等各阶段入手, 采用一系列综合方法。至今人们所采用的手段可归纳为:

(1) 输入信息的多样化: 利用多重图象 (体视图象、动态图象、彩色图象)、多传感器采集;

(2) 采用评价函数的最优化法: 将从图象中获得的信息作为约束条件, 来求得使表示对象形状和匹配优劣的评价函数最大的解释;

(3) 知识的利用: 在获得景物描述的过程中, 不仅要利用物体的形状模型, 还要利用其它知识来控制匹配和分割过程。

### 2.1 多重信息输入与融合

用传统的摄像机获得的灰度图象, 能抽取的信息很有限, 而且在这种单一数据上分析求解结果, 其精度和可靠性均不高。因此, 人们考虑采用多种传感器, 从景物中获得多重信息, 通过多重信息的融合, 最后得到高可靠性、高精度的结果。这种方式被称之为多传感器信息融合 (Multi-sense-fusion) [2], 已被用于移动机器人的导航系统。

任意一种传感器的检测信息可表示成  $I(T, L, t, P, E)$ , 其中

$T$ : 传感器检测的信息类型 (光的波长, 光强度或三维距离等);

$L$ : 传感器位置;

$t$ : 检测的时间;

$P$ : 传感器内部参数 (焦距、光圈等);

$E$ : 检测环境 (光源位置、个数、种类等)。

在多传感器信息融合过程中, 通过这五个参数得到多种关于景物的不同信息, 形成多重图象。例如, TV摄像机和RangeFinder对应于 $T$ 的变化; 体视图象对应于 $L$ 的变化; 动态图象对应于 $t$ 的变化; 自动调焦摄像机对应于 $P$ 的变化; 灰度体视法对应于 $E$ 的变化 [3]。

目前对多重图象的分析方法产生了如下

结果:

T: 在多光谱图象中, 利用各频段信号之比来检测不同质区域 (水域或植被);

L: 使摄像机连续直线运动过程中拍摄的图象序列按图1所示排列, 形成三维图象。物体上各点的象在三维图象的水平断面上呈直线形态, 利用该直线的斜率, 可求得其对应物点的三维距离<sup>[4]</sup>;

t: 在动态图象分析中, 可以从 $di/dt$  (亮度随时间的变化) 求得可见区域的速度矢量<sup>[5]</sup>;

P: 使摄像机的焦距连续变化, 从所获取图象清晰度的变化情况计算出三维距离<sup>[6]</sup>;

E: 在 $n$ 个 ( $n \geq 3$ ) 不同的位置设置光源, 获得与各光源相对应的 $n$ 幅图象。根据各图象象素 (pixel) 亮度的差异, 可求得物体公共可视点的法线矢量<sup>[3]</sup>。这就是灰度体视法的原理。

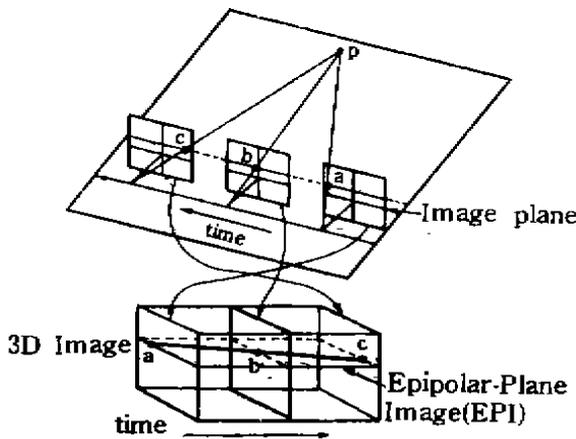


图1 移动体视图象融合 (引自[4])

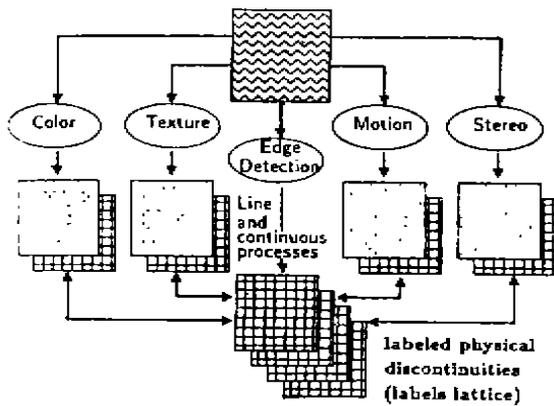


图2 MIT Vision Machine的结构 (引自[7])

Poggio的研究小组开发的MIT Vision Machine系统实现了抽取多种特征信息进行

融合, 如图2所示。该系统的硬件基础是连接机 (Connection Machine), 在信息融合过程中采用了所谓正则化的最优化算法<sup>[7,8]</sup>。

实际上, 可以从传感器级、算法级和推理级这三个层次来讨论多重信息的融合问题。在传感器级人们开发了稳定的彩色信息获取方法<sup>[9]</sup>, 它主要利用摄像机参数连续变化所拍摄的多重图象。在算法级多是有机地复合利用基本的图象处理算子来实现高可靠性<sup>[10]</sup>。在推理识别级, 人们开发了利用部分信息进行融合的所谓空间推理方法<sup>[11]</sup>, 关于这一点后面将进行进一步讨论。

多重信息的融合是今后图象理解研究的重要课题。很显然, 仅采用多种传感器实现输入信息的多样化是有限的。必须开发信息的融合算法, 研究弥补输入信息不完备的推理机制。

### 2.2 最优化方法的不足

为了有效地处理含噪声的信号, 人们引入了各种最优化方法。例如, 采用动态规划法来检测边缘<sup>[12]</sup>、识别脸谱<sup>[13]</sup>、进行体视匹配等<sup>[14]</sup>。

计算机视觉领域的许多问题, 一经数学形式化, 就是一个不定解 (Underconstrained) 方程。正则化方法 (Regularization) 就是求解这类不适定问题的有效方法之一, 它已被用于边缘检测、体视匹配和光流 (Optical flow) 计算等方面<sup>[15]</sup>。实际上 Rosenfield提出的松弛法 (Relaxation method) 和Hopfield提出的神经网络模型也被有效地用于求解初级视觉问题<sup>[16,17]</sup>。

采用最优化方法的首要之点是采用什么形式的评价函数。在计算机视觉领域, 要求评价函数的物理意义从信息处理观点来看是可行的。例如, Hildreth假设一个连续边界上的速度矢量是平缓变化的, 他通过使表示速度矢量变化的评价函数最小化, 来求解边界上的速度矢量<sup>[18]</sup>。笔者认为使用最优化方法时存在如下不足:

(1) 当处理对象较单一时, 给出一个评价函数比较容易。但当对象结构复杂或者需要对整体景物进行识别时, 很难找到一个有效的评价函数;

(2) 由于最优化处理是均匀的, 所以不能变更和控制与对象结构相应的局部处理。

因此, 在处理复杂景物时必须先通过一些别的推理和控制机制, 求得各个结构均匀的局部区域, 然后再用最优化方法来进行处理。

### 2.3 基于知识的视觉处理

所谓的基于知识的视觉 (Knowledge-based Vision) 是指充分利用关于景物的知识 (如识别对象的形状模式和识别过程的控制信息) 来有效地分析复杂景物。这些知识包括:

(a) 关于识别特征 (distinguishing feature) 的知识, 用于识别多个对象的有效特征;

(b) 关于识别特征重要性的知识, 往往是根据重要程度来顺序识别或搜索;

(c) 关于如何从失败中返回的知识, 用于当推理和分析处理失败时, 检查其原因并找出代行方案;

(d) 关于如何生成假设的知识, 用于从现有信息中未测检到的对象生成相应的假设;

(e) 关于局部信息一致性、协调性的知识, 用于检查在处理过程中所得到的局部信息之间的一致性和协调性, 并给出其判别准则。

构造基于知识的视觉系统的关键之一, 是要保证各类知识在整体上逻辑相容。Reiter等人曾用谓词逻辑来描述地图理解所需的知识, 尝试将知识和推理机制形式化<sup>[19]</sup>。引入的知识有三类:

(1) 图象公理 例如

$$(\forall X) \text{image-object}(x) \equiv \text{Chain}(x) \vee \text{region}(x)$$

$$(\forall X) \neg (\text{chain}(x) \wedge \text{region}(x))$$

(2) 景物公理 例如

$$(\forall X) \text{SCENE-OBJECT}(x) \equiv \text{LINEAR-OBJECT}(x) \vee \text{AREA}(x)$$

$$(\forall X) \text{LINEAR-OBJECT}(x) \equiv \text{ROAD}(x) \vee \text{RIVER}(x) \vee \text{SHORE}(x)$$

(3) 图象与景物映射公理 例如

$$(\forall i) \text{image-object}(i) \supset \text{SCENE-OBJECT}(\sigma(i) \wedge \Delta(i, \sigma(i)) \wedge (\forall S) \Delta(i, S) \supset S = \sigma(i))$$

$\Delta(i, j)$  表示图象对象*i*与景物对象*j*对应关系的谓词。在推理过程中, 不仅要利用这些形式化知识, 还要用一系列隐含知识。例如要用到封闭假设 (Closure assumption) 和单名假设 (Uniquename assumption)。理论上说, 这种形式化逻辑系统的一个推导延拓就是相应地图的一个合理解释。但是, 实际上各种假设在分割过程中不一定成立, 如何求得一个代表解释的逻辑延拓也非顺利, 这些都将成为构作系统的障碍。所以必须研究新的逻辑推理方式以及如何利用更复杂的知识的方法。

## 三、解释推理与结构推理

图象理解的关键之一就是必须从不完全的信息中求得景物的描述。下面就解决这一问题的推理方式进行讨论。

### 3.1 松弛法与Hough变换

松弛法与Hough变换是图象分析和理解中广泛应用的有效算法。两者在实现对模糊性和误差信息的稳定分析上具有相同的功能。但从推理方式上来看, 前者属于解释型推理, 后者属于结构型推理。

所谓解释型推理是指对从给定的图象中抽取的图象特征进行解释 (Labeling), 被解释对象的图象特征集是固定的。因此, 当输入信息 (图象特征集) 不充分时, 无法对其弥补。

结构型推理是根据推理/分析过程中动态形成的新的信息生成假设, 以这些新的假

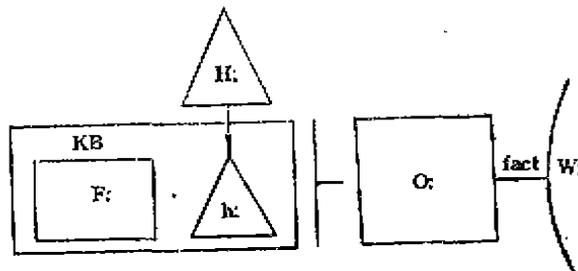
设为基础来推导出合适的解释/结论。例如，在Hough变换过程中，先以边缘点的各图象特征为基础，生成识别对象（直线或圆）的假设。一个图象特征信息可以同时形成多个可能的假设，将假设集合以参数空间的轨迹形式表示，然后分析它们与由别的图象特征生成的假设之间的相容性，以可靠性高的假设作为识别结果，因此最终的解释都是一些通过推理得到的假设，即使图象信息不充分也可以获得对象的完整描述。

### 3.2 图象理解与假设推理

众所周知，基于Robison归结原理的演绎推理系统只能推导出由给定的谓词公式集所规定的信息，它不能生成新的信息（知识）。即这类逻辑系统是封闭的，信息（知识）也必须是完备的。这对图象理解任务来说是很大的不足。

在AI领域人们已认识到，要预先给出客观对象的完整信息是困难的。所以必须研究基于不完整知识的高次推理

假设推理<sup>[20]</sup>就是高次推理方式之一。当给定的知识不能完全解释观察事实时，系统生成新的假设，以达到能说明事实之目的。假设推理框架如图3所示，它属于前述的结构型推理范畴。



- H, a set of possible hypothesis
- KB, knowledge base
- F, a set of known knowledge
- h, hypothesis
- O, a set of given observed facts
- W, real world

图3 假设推理框架 (引自[20])

由景物信息和关于对象的知识生成假设，通过图象数据验证假设的可行性。在这种推理方式中，只要所生成的假设与输入数据不矛盾，就认为假设是正确的。所以即使信息不完整也可以得到若干有效结论。

下面以谓词逻辑形式举例说明假设推理过程。

例1 设识别对象 $O_1$ 和 $O_2$ 满足关系REL, a、b分别是 $O_1$ 、 $O_2$ 的实例，且已被识别。这些信息以谓词逻辑可表示为

$$\text{公理1 } \forall x [O_1(x) \supset \exists y (O_2(y) \wedge \text{REL}(x, y))] ]$$

$$\text{事实1 } O_1(a)$$

$$\text{事实2 } O_2(b)$$

那么结论： $\exists z (O_2(z) \wedge \text{REL}(a, z))$ 在一般的推理方式下可以得证，且有约束条件  $z|f(a)$ ，这里f是公理1变换成子句形式所引入的Skolem函数。

若采用假设推理，则先由公理1和事实1生成假设f(a)，搜索与f(a)相容的别的假设或实例。若f(a)与b相容，则系统认为它们表示同一对象，于是可给出新的假设f(a)=b，由此可得到一个新的事实REL(a, b)，并有约束条件z|b。实际上这是一种自底向上的推理过程。

例2 设有如下公理和事实

$$\text{公理1 } \forall x [O_1(x) \supset \exists r \{O_3(r) \wedge \text{REL1}(x, r)\}] ]$$

$$\text{公理2 } \forall y [O_2(y) \supset \exists t \{O_3(t) \wedge \text{REL2}(y, t)\}] ]$$

$$\text{事实1 } O_1(a)$$

$$\text{事实2 } O_2(b)$$

从所给信息用传统推理方式不能证明以下结论：

$$\exists z \{O_3(z) \wedge \text{REL1}(a, z) \wedge \text{REL2}(b, z)\}$$

但采用假设推理是可行的，首先从公理1、2和事实1、2可生成假设f(a)和g(b)，验证f(a)、g(b)的相容性。如果相容性被确认，则可生成假设f(a)=g(b)。从而结论得证且得到约束条件z|f(a)。进一步检查满足假设f(a)的图象特征并搜索受z约束的实例。这是一种自顶向下的推理过程。

在图象理解过程中，由于分割的不完全，往往会抽取错误信息。为此，理解系统不仅要具有纯逻辑的假设推理功能，还需要具备检查错误观测事实的功能。即在错误信息与正确信息必定产生矛盾的假设之下，一

个系统应具备如下功能:

(a) 矛盾发现: 检查外部输入信息与系统生成的假设之间的相容性、协调性, 检测矛盾之所在;

(b) 搜索并消除产生矛盾的原因: 分析所发现的矛盾, 找出并消除产生矛盾的原因信息, 以达到消解矛盾。

Herman和Kanade在他们的城市航空照片三维结构识别系统<sup>[21]</sup>中, 首先以照片图象得到的部分“面”、“棱”、“顶点”为基础, 生成建筑物整体形状的假设, 以弥补信息不充分。当由新的观测照片中获得的“面”、“棱”、“顶点”与前面生成的假设矛盾时, 搜索并去除产生矛盾的原因, 生成新的假设。这是一种典型的假设推理。这种方式在真值维护系统中也被利用。

图象理解的关键问题是信息量不足, 当分析过程中发现了相互矛盾的解释时, 立即作出“谁是正确的”判断是危险的, 因为没有足够的信息支持这种判断往往会产生失误。可取的方法是在系统发现了矛盾时, 并不消除产生矛盾的解释, 而是记录多个(矛盾的)可能的解释, 保留多种可能性, 继续往下推理。图象理解系统ACRONYM正是采用这种手法<sup>[22]</sup>。

以上我们从推理方式的角度讨论了图象理解中存在的各种问题。现总结如下:

1. 图象理解问题本身存在不确定性, 纯粹的符号处理、传统的逻辑推理手段不足以解决问题。必须研究基于不完全信息的分析推理。

2. 多传感器信息输入与多重信息融合是解决图象分割中不确定性的主要手法, 可以有效地提高结果精度与可靠性。在三种不同层次上利用多重信息融合, 是至今为止图象理解研究中的主要任务, 取得了一些令人满意的成果。但在推理识别层次上应用多重信息融合还很不成熟, 还有许多值得探讨的问题。

3. 对于最优化方法, 由于其全局一致均匀性、评价函数的不易给定, 因此只适用于简单的或局部的区域。

4. 基于知识的视觉处理是目前的热门课题, 而

知识收集、知识表达等问题本身是知识工程领域中较难有效解决的问题。

5. 解释型推理有其固有的静态特征, 而结构型推理是动态生成假设并验证。所以后者是解决不完全性信息推理的有效工具。假设推理方式是属于结构型推理的一种。

Marr给出了视觉信息处理的框架。最近, Tsotsos又从视觉信息处理的计算复杂性出发, 研究视觉信息处理系统应具备的功能。并指出了利用知识、视觉信息的层次结构以及自顶向下分析的必要性<sup>[23]</sup>。笔者认为要建立一个实用的图象理解系统, 必须综合研究信息的融合算法、识别过程的控制机制, 尤其是弥补信息不充分的推理机制。另外, 既要借助于传感器技术和并行计算理论的发展, 又要充分利用人工神经网络等相关学科的研究成果。

### 参考文献

- [1] Nagao, M., "Control Strategies in Pattern Analysis", Proc. of 6th International Conference on Pattern Recognition, pp.996-1006 (1982)
- [2] Odin, K.E., et al., "Development in Knowledge-Based Vision for Obstacle Detection and Avoidance", Workshop on Image Understanding, pp.78-88 (1987)
- [3] 池内克史, "反射率地図に基づき, 二次元濃淡画像より三次元形状を再構成する手法", 電子通信学会論文誌, Vol.J65-D, pp.842-849 (1982)
- [4] Bolles, R.C. and Baker, H.H., "Epipolar-Plane Image Analysis, A Technique for Analyzing Motion Sequence", 3rd Int. Symp. of Robotics Research, pp. 192-199 (1985)
- [5] Horn, B.K.P. and Schunk, B.G., "Determining Optical Flow", Artif. Intell., Vol.17, pp.185-203 (1981)
- [6] Pentland, A.P., "A New Sense for Depth of Field", 9th IJCAI, pp.988-994 (1985)
- [7] Poggio, T., et al., "The MIT Vision Machine", Autumn School, Ecole Normale Supérieure (Sept 1988)
- [8] Hillis, W.D., "The Connection Machi-

- ne", MIT Press (1985)
- [9] Boyer, K.L. and Kak, A.C., "Color-Encoded Structured Light for Rapid Active Ranging", IEEE Trans.PAMI, Vol. PAMI-9, No.1 (1987)
- [10] Horn, B.K.P., "Robot Vision", MIT Press (1986)
- [11] 松山隆司, "画像理解システムSIGMA", 情報処理学会論文誌, Vol.26, No.5, pp.877-889 (1985)
- [12] Montanari, U., "On the Optimal Detection of Curves in Noisy Picture", Comm.ACM, Vol.14, No.5, pp335-345 (1971)
- [13] Fischler, M.A. and Elschlager, R.A., "The Representation and Matching of Pictorial Structures", IEEE Trans., Vol. C-22, No.1, pp.67-92 (1973)
- [14] Ohta, Y. and Kanade, T., "Stereo by Intra and Inter-Scanline Search Using Dynamic Programming", IEEE Trans., Vol.PAMI-7, No.2, pp.139-154 (1985)
- [15] Poggio, T., et al., "Computational Vision and Regularization Theory", Nature, Vol.317, pp.314-319 (1985)
- [16] Hopfield, J.T. and Tank, D.W., "Neural Computation of Decisions in Optimization Problems", Biol. Cybern., Vol. 52, pp.141-152 (1985)
- [17] Hummel, R.A. and Zucker, S.W., "On the Foundations of Relaxation Labeling", IEEE Trans., Vol.PAMI-5, No. 3, pp.267-287 (1983)
- [18] Hildreth, E.C., "Computation Underlying the Measurement of Visual Motion", Artif.Intell., Vol. 23, pp. 309-354 (1983)
- [19] Reiter, R. and Mackworth, A.K., "The Logical of Depiction", Tech.Rep.RC-BV-TR-87-18, University of Toronto (1987)
- [20] 特集, "高次人工知能へ向けてのパラダイム", 人工知能学会誌, Vol.2, No.1 (1987)
- [21] Herman, M. and Kanade, T., "Incremental Reconstruction of 3D Scenes from Multiple Complex Images", Artif. Intell., Vol.30, pp.289-341 (1986)
- [22] Brooks, R. A., "Symbolic Reasoning among 3-D Models and 2-D Images", Artif.Intell., Vol.17, pp.285-348(1981)
- [23] Tsotsos, J. K., "A Complexity Level Analysis of Immediate Vision", Int.J.of Computer Vision, Vol.1, No.4, pp.303-320 (1988)

### 下期主要内容预告

面向对象数据库系统的研究方向

关于PCTE的浅析

计算反射和面向对象程序设计

Stream和Object

——面向对象的并行计算模型及其语言探讨

并行程序设计语言Occam及其运行机Transputer

DSS发展的一个新领域

——执行信息系统

缺席推理的逻辑框架

专家系统间的知识获取与专家系统兼并