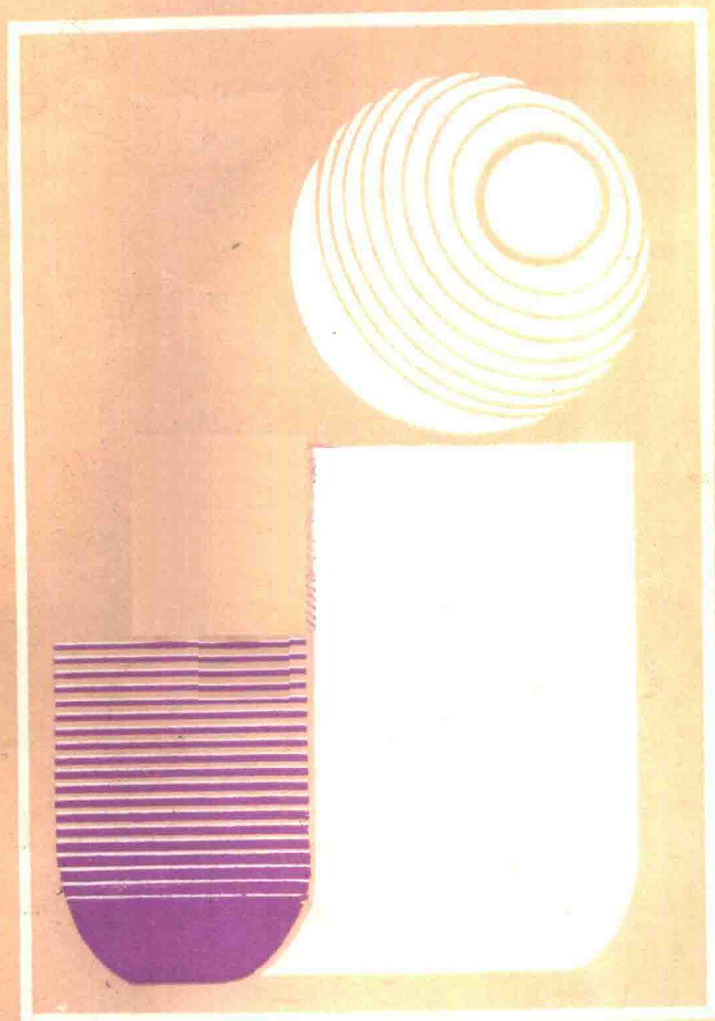


计算机视觉的 理论和实践

李介谷 编著



上海交通大学出版社

计算机视觉的理论和实践

李介谷 编著

上海交通大学出版社

沪新登字 205 号

内 容 提 要

本书介绍了有关计算机视觉的主要的计算理论,包括三维初始略图和 LOG 滤波器, $2\frac{1}{2}$ -维理解及 shape from x 问题,光流计算,数学形态学及三维信息提取,三维结构的表达,序列图像及运动估计,图匹配和松弛算法及人工神经网络理论等。应该指出,三维视觉理论由于问题的复杂性,至今很少有完整的体系,但本书的内容包含了这方面最主要的动向。为了使计算理论能够付诸实践,在本书的各章节中还列举出在各种图像模式识别领域中的实践结果。

本书可作为模式识别与智能控制学科的研究生教材,也可供有关专业的技术人员参考。

计算机视觉的理论和实践

出版:上海交通大学出版社

(淮海中路 1984 弄 19 号)

发行:新华书店上海发行所

印刷:江苏太仓印刷厂

开本:787×1092(毫米) 1/16

印张:16

字数:400000

版次:1991年11月 第1版

印次:1991年12月 第1次

印数:1—2500

科目:263—279

ISBN 7-313-00983-8/TP

定价:4.15 元

前 言

自1985年起,由于形势发展的需要,作者开始为上海交通大学模式识别与智能控制学科的研究生开设计算机视觉课程。每年讲授一次,六七年来陆续积累了一些资料。与此同时,由于“七五”重点攻关、高科技任务和国家自然科学基金等方面课题的开展,在这个领域内进行了若干方面的研究,并在不断接触之下,建立起了一些自己的想法和看法,当然由于磨砺不够,这些想法和看法可能是相当肤浅的。这是本书成书的基本背景。

就自动控制、人工智能等方面来看,计算机视觉是一个特别令人感兴趣的前沿领域。其原因也许在于,这是一门交叉性很强的学科,当从交叉点上向外发掘时,总会给人以一种新奇感。然而,更重要的还在于这种交叉的环境还没有成熟到可以使计算机视觉的理论体系和结构实现能够达到瓜熟蒂落的地步。计算机视觉只是一个处于孕育期的胎儿,还有待于各门不同的学科提供其成长的养料。“路漫漫其修远兮”却正好值得去上下求索。“八方风雨会中州”,这大概是各方人士(包括国内和国外)都对之感兴趣的主要原因吧!

尽管这是一个热门题材,但是在国内却至今还没看到一本稍成体系可以用作教材的书。它的确是不太容易写的,因为它尚未成熟到已经成型的地步。但是,为了使上课时能有教材可循,也为了使众多感兴趣的同志能有一些前进的踏蹬,这样一本书又是迫切需要的。

“十年辛苦不寻常”,这本书实际上还汇集了我们研究所其他许多同志的辛劳。他们有的是我的同事,有的是我过去或现在的学生。他们的姓名,一部分可以从本书所列的参考文献中见到,有的则并没有列进去。我只是在酿蜜,他们是辛勤地协助采集花粉的人。我在这里向他们表示由衷的感谢。

李介谷

于上海交通大学图像所

1991年国庆

目 录

第一章 引论	(1)
1-1 二维的计算机视觉系统.....	(1)
1-2 由二维向三维推广.....	(9)
参考文献.....	(12)
第二章 图像与现实空间	(13)
2-1 投影和透视.....	(13)
2-2 三维视觉的可计算性问题.....	(25)
2-3 对于静止灰度的理解.....	(30)
2-4 关于运动着的灰度.....	(37)
参考文献.....	(46)
第三章 三维基本素描表达	(47)
3-1 LOG 滤波器.....	(47)
3-2 零穿与小面模型.....	(59)
3-3 地形性基本素描表达.....	(70)
参考文献.....	(77)
第四章 $2\frac{1}{2}$ 维理解	(79)
4-1 由明暗度提取外形信息(单目图像).....	(79)
4-2 由明暗度提取外形信息(双目、多位置图像).....	(89)
4-3 由纹理畸变求外形.....	(94)
4-4 单目图像的线性化解.....	(100)
4-5 表面高度的恢复.....	(102)
参考文献.....	(103)
第五章 数学形态学与图像形态分析	(104)
5-1 二值图像的数学形态学运算.....	(104)
5-2 灰度图像的数学形态学运算.....	(127)
5-3 数学形态学与地形性基本素描表达.....	(137)
参考文献.....	(139)
第六章 基本素描的符号标记	(140)
6-1 角度的符号标记.....	(140)
6-2 线段的符号标记.....	(143)
6-3 曲面体的符号标记.....	(146)
6-4 有遮蔽交点的符号标记(推理技术的采用).....	(152)
6-5 对工程图的理解.....	(157)

参考文献.....	(163)
第七章 三维结构的表达	(164)
7-1 表面分解法.....	(164)
7-2 穿心轴表达(或规范圆柱体).....	(165)
7-3 体素表达法.....	(174)
参考文献.....	(184)
第八章 看到的运动	(185)
8-1 正交投影条件下图像运动分析的可能性.....	(186)
8-2 透视投影条件下图像运动分析的可能性(单目图像).....	(192)
8-3 目标轨迹跟踪和预报.....	(196)
8-4 自运动情况下的运动识别.....	(198)
参考文献.....	(204)
第九章 图像的匹配理解	(205)
9-1 关系匹配.....	(205)
9-2 线图同构.....	(209)
9-3 松弛算法.....	(218)
参考文献.....	(222)
第十章 人工神经网络原理	(224)
10-1 神经元及神经元模型.....	(224)
10-2 多层感知器.....	(229)
10-3 感知器的识别功能.....	(231)
10-4 感知器的训练算法.....	(234)
10-5 其他几种人工神经网络线路.....	(241)
10-6 应用举例.....	(245)
参考文献.....	(247)

第一章 引 论

1-1 二维的计算机视觉系统

计算机视觉系统是指用计算机来实现人的视觉功能,也就是用计算机来实现对客观的三维世界的识别。按现在的理解,人类视觉系统的感受部分是视网膜,它是一个二维采样系统。三维的物体的可见部分投影到视网膜上(当然是左右两眼的视网膜,它们带有估计距离的功能),人们按照投影到视网膜上的二维的像来对该物体进行三维的理解。所谓三维的理解是指对被观察对象的形状、尺寸、离开观察点的距离、质地和运动特征(方向和速度)等的理解。人们在实际生活中,除了处于睡眠状态外,无不每时每刻在对客观三维世界进行这种性质的理解。例如行路,人们得不断地根据眼前的景物在视网膜上的投影以进行识别理解,才能对行进的道路、方向和速度作出抉择。计算机视觉希望解决的是完全相同的问题。计算机视觉系统的输入装置可以是摄像机、转鼓等,它们都把二维的影像作为输入源;换言之,输入计算机的就是三维客观世界的二维投影。如果把三维客观世界到二维投影像看作是一种正变换的话,则计算机视觉系统所要做的是从这种二维投影像到三维客观世界的逆变换,也就是根据这种二维投影像去重建三维的客观世界。从三维客观世界到二维投影像,有一维信息从显露成为隐含,计算机视觉系统的工作首先就是要把这种隐含的三维信息提取出来,随后才能进行理解工作。目前的困难是这种隐含信息的显露化,由于原始条件的不足,带来的都是病态的解题过程(例如以后要介绍的 $2\frac{1}{2}$ 维表达和光流计算等所遭遇到的都是病态方程的求解),因而难于

得到唯一解。或者即使能得到解,也需要花费大量的运算时间。因而除理论价值外,暂时还得不到实用的意义。这就是说,人们轻而易举地一天进行千百万次的三维识别工作,用计算机来实现时,却是困难重重。实际上,目前还不存在具有一定通用性且有一定实用价值的三维计算机视觉系统。

在这种情况下,讨论三维计算机视觉系统只能按其可能的有关组成部分局部性地进行。在不存在这类系统以前,如何使对它的局部性讨论带有隐含的系统性呢?为此,这里把讨论的起点放在已有的具有相当通用性的二维计算机视觉系统上。从80年代初期起,已经出现了商品化的这样的类型的系统。从某种意义上来说,二维计算机视觉系统为三维系统提供了一定的发展基础,可以从已有的二维系统隐约地眺望到三维系统的大致骨架。这样对已有二维系统的介绍对本书的整体内容来说,希望能起到穿轴线的的作用,尽管从二维到三维的变化,不是简单的量的丰富,而是质的飞跃。

下面介绍的是在80年代初期问世的VS-100二值化视觉系统,它是由美国机器智能公司生产的。二值化视觉系统的输入装置为电视摄像机,但摄入的灰度图像先被二值化后再输入计算机进行识别运算,因而三维的信息丧失殆尽,这只能是一个二维的计算机视觉系统。

二值化系统的各主要计算环节如图1-1所示。

作为一个带有通用性的系统来说,VS-100的主要工作特点是先按输入的样本模式来训

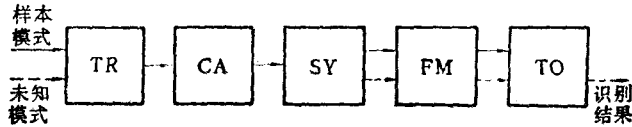


图 1-1 二值化计算机视觉系统 VS-100

练系统，然后再用训练好的系统去识别未知的模式。整个系统由五个主要的环节所组成（见图 1-1），它们是：二值化和行长编码(TR)，连接分析(CA)，符号化(SY)，特征提取(FM)和训练、分类(TO)。TR 模块是附设于计算机外的一个插件，CA、SY、FM 和 TO 则都是计算机的软件模块。

TR 模块的功能是把摄像输入的灰度信息进行二值化处理 and 压缩编码，它们的作用可用图 1-2 来说明。

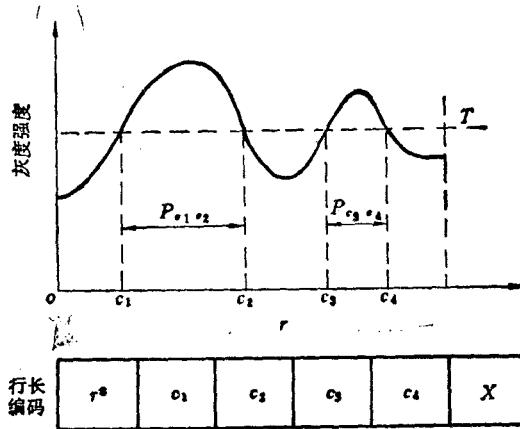


图 1-2 取阈二值化和行长编码

设图像的灰度强度按 256 级离散化(即占据一个字节的空问)，图像的采样尺寸为 256×256 (字节)²。采用阈值 T 对图像进行二值化，凡超过 T 的灰度部分都作为“1”，而低于 T 的灰度部分都作为“0”，则图像信息的每一行都成为一串高度为“1”的脉冲。如图 1-2 所示，如灰度强度与 T 的交截点为 c_1 、 c_2 、 c_3 和 c_4 。则脉冲宽度为 P_{c_1, c_2} 和 P_{c_3, c_4} 。图像的每一行都按脉冲的出现和消失来进行编码，这就是行长编码。在行长编码中，第一个码 r^* 为行数，末尾的码 X 为结束符。对于图像处理的任务来说，由于一幅图像所包含的信息量极大，对于上例来说为 $256 \times 256 \times 8$ bit，因而压缩信息量，减少内外存交换的次数以提高处理速度一直是第一要考虑的问题。80 年代初期，八位微处理机刚登场不久，其内存都不大，即使是 DEC 11/23 型的微型机其内存也不过 256 KB，一个系统要具有实用的价值，非走信息压缩的路径不可。这里二值化和行长编码就起着这样的作用。就上例来说，原来一行图像的存储容量为 256 B，采用二值化和行长编码后所占的存储容量，以一个码占一个字节计，为 6 B。压缩比达到 6/256，只是原来存储容量的 1/42。经编码后，输入计算机的将是一串行长编码，现在所有的计算将建立在对行长编码进行处理的基础上。

第二个环节，连接分析 CA，把一行一行的行长编码连接成块体，后者实际上是建立一个表格型的数据结构，把所有的行长编码按同一个块体的隶属原则分列开来。

CA 模块的功能有四方面：列表(用算符 I 来表示)，延续(用算符 C 来表示)，混和(用算

符 M 来表示)和消除(用算符 D 来表示)。

I 功能为建立如下形式的块体表:

Blob*1	Blob*2	Blob*3	...
RUN x_1	RUN y_1	RUN z_1	⋮
RUN x_2	RUN y_2	RUN z_2	⋮
⋮	⋮	⋮	⋮

在上表中 Blob*1, Blob*2, Blob*3 为块体 1 号, 2 号和 3 号。RUN x_1 , RUN x_2 为隶属于同一块体 Blob*1 的编码; y_1, y_2 和 z_1, z_2 则依此类推。

设一条行长编码为

$$r^*, x_1, y_1, z_1, X;$$

由图 1-2

$$x_1 = c_1 c_2,$$

$$y_1 = c_3 c_4,$$

$$z_1 = c_5 c_6 \text{ (图中未示出)}。$$

当然

$$(RUN x_1 \cap RUN y_1 \cap RUN z_1) = \emptyset.$$

和

$$I(RUN x_1) = \text{Blob}^*1,$$

$$I(RUN y_1) = \text{Blob}^*2,$$

$$I(RUN z_1) = \text{Blob}^*3.$$

C 功能为把相邻行的码按延续的原则组合到一个块体中去。设

$$r^* = A \text{ 和 } r^* = B,$$

且

$$B = A + 1,$$

它们为前后相邻的行, 并且

$$(RUN x_1, RUN y_1, RUN z_1) \in A,$$

$$(RUN x_2, RUN y_2, RUN z_2) \in B;$$

并且

$$RUN x_1 \cap RUN x_2 \neq \emptyset,$$

$$RUN y_1 \cap RUN y_2 \neq \emptyset,$$

$$RUN z_1 \cap RUN z_2 \neq \emptyset;$$

则当然

$$C(RUN x_2) = \text{Blob}^*1,$$

$$C(RUN y_2) = \text{Blob}^*2,$$

$$C(RUN z_2) = \text{Blob}^*3.$$

M 功能为把原来孤立的各块体在出现把它们接续在一起的码以后, 把各孤立的块体混

和成一块,它如图 1-3 所示。

这就是说,如果

$$\text{RUN } x_n \cap \text{Blob}^*1 \neq \emptyset,$$

$$\text{RUN } x_n \cap \text{Blob}^*2 \neq \emptyset,$$

则

$$M(\text{RUN } x_n, \text{Blob}^*1, \text{Blob}^*2) = \text{Blob}^*4.$$

算符 D 的作用为对通过混和建立起来的新块体,消除其中老块体的编号,即

$$D(\text{Blob}^*1, \text{Blob}^*2) = \text{Blob}^*4.$$

经过 CA 模块的作用以后,原来以行为单元的编码,现在分散集结到各自应该归属的块体。这些块体就是输入以供训练的样本模式或待识别模式。

第三个环节为符号化模块 SY,它把以行长表示的各块体转变成八方向链码(或被称为弗里曼(Freeman)链码)。

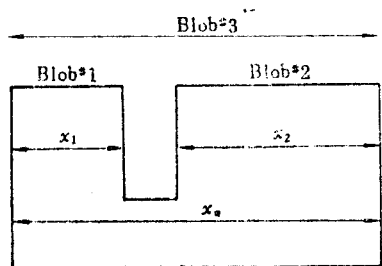
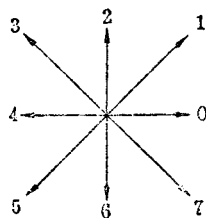
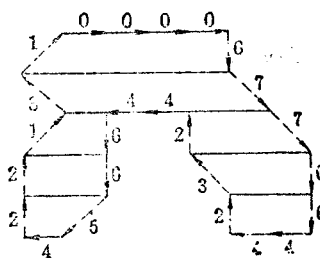


图 1-3 M 算符的作用



(a)



(b)

图 1-4 链码及其表达方式

(a) 链码, (b) 块体。

在图 1-4(a)上 0、2、4 和 6 四根指针的长度为一个像素的宽度, 1、3、5 和 7 则为一个像素的对角线长度, 它们各相应地代表了由原点出发的八个不同方向。图 1-4(b) 中用实线代表图像每一行中出现对象的行长, 其外周用虚线框起来的范围即为对象的外型。如果从最低最左开始, 对该外型用八方向链码编码, 则可得代表外型链码为 $2^21310^467^26^42^2324^26^254$ 。这串链码包含了该对象外型的几何信息。从行长编码到八方向链码是又一次信息形式的转换, 前者的目的是为了信息压缩, 后者的目的则是为了便于提取几何信息。

第四个环节为特征提取 FM, 它对用八方向链码表达的对象外型提取有关的几何特征。作为一个通用系统来说, 该系统提供尽可能多的特征提取手段以供用户选用。这些特征为:

F_1 ——颜色(指 1 或 0 的聚块, 目标中的一个孔为 0 的聚块);

F_2 ——一块中的孔数;

F_3 ——“1”区的面积, 为物体的实体;

F_4 ——“0”区的面积, 为孔;

F_5 ——“0”区和“1”区的总面积;

F_6 ——块体的重心位置, x 轴位置;

F_7 ——块体的重心位置, y 轴位置;

F_8 ——用椭圆来密合非圆形物体时, 该椭圆的长轴;

F_9 ——同上椭圆的短轴;

- F_{10} —— F_8 的长轴与 x 轴的夹角;
- F_{11} ——周长;
- F_{12} ——形状系数, 它等于 F_{11}^2/F_5 ;
- F_{13} ——当块体为非正规形状时的最长半径;
- F_{14} ——同上, 但为最短半径;
- F_{15} —— F_{13} 的最长半径与 x 轴的夹角;
- F_{16} —— F_{14} 的最短半径与 x 轴的夹角;
- F_{17} —— F_{13} 与 F_{14} 之间的夹角;
- F_{18} ——块体的宽度;
- F_{19} ——块体的高度;
- F_{20} ——块体重心至图像中心的二价矩。

已经说过, 这 20 个特征的计算完全是按八方向链码来进行的, 这些算法的设计有时也是颇有意义的问题。以计算块体的面积为例来说明一下这一问题。

设链码为 i , 且 $0 \leq i \leq 7$; 坐标的起始点为 (x, y) , 对应于 $i = 0, 1, 2, 3, 4, 5, 6, 7$ 的坐标序; 列各相应地为 $(x+1, y), (x+1, y+1), (x, y+1); (x-1, y+1), (x-1, y), (x-1, y-1), (x, y-1)$ 和 $(x+1, y-1)$ 。

如以函数 $M(i)$ 来表示八方向链码的坐标特点, 则在

$$i(\text{MOD } 8) = 0, 4 \text{ 时} \quad M(i) = 0,$$

$$i(\text{MOD } 8) = 1, 2, 3 \text{ 时} \quad M(i) = 1,$$

$$i(\text{MOD } 8) = 5, 6, 7 \text{ 时} \quad M(i) = -1.$$

链码 i 从 x, y 出发到达新坐标 x', y' 的位置为

$$x' = x + M(i+2)(\text{MOD } 8), \quad (1-1)$$

$$y' = y + M(i)(\text{MOD } 8). \quad (1-2)$$

面积可以从多边形包围面积的关系去求取, 它是

$$S = \frac{1}{2} \left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \dots + \begin{vmatrix} x_{n-1} & x_n \\ y_{n-1} & y_n \end{vmatrix} + \begin{vmatrix} x_n & x_1 \\ y_n & y_1 \end{vmatrix} \right), \quad (1-3)$$

在上式中, $x_1y_1, x_2y_2, \dots, x_{n-1}y_{n-1}, x_ny_n$ 为多边形各顶点的坐标, 毫无疑问, 它们形成封闭的区域。由于该面积是有向的, 因而 S 取其绝对值。

对式 1-3 来说, 如 x_1y_1 为链码的起点位置, 则整个链码序列的坐标可由式 1-1 和 1-2 求得。这样整个链码的包围面积就可由式 1-3 求得。

最后一个环节 TO 是训练和分类。训练实际上是指对于给定的识别任务作分类器的设计。如果现在的任务是对形状不同的 A 和 B 两种物体进行识别, 则分类器的设计就是从上一环节所提供的特征向量

$$F = (F_1, F_2, \dots, F_{20})^T,$$

式中 T 为转置符号, 从中选择一个特征向量

$$F_s = (F_{s1}, F_{s2}, \dots, F_{sn})^T, \\ F_s \subseteq F,$$

使得可以按照 F_s 来对 A 和 B 进行分类。

众所周知, 对于一定的样本集来说, 它在一定的特征空间出现的概率是正态分布的。所选

择的特征集应该使这些样本在该特征空间是各别地按不同类别分布开来的(见图1-5)。

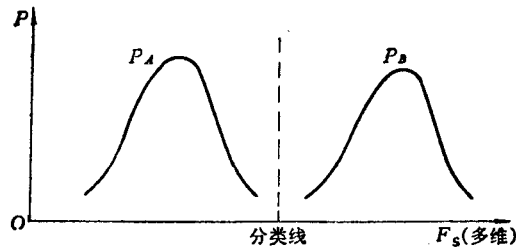


图1-5 理想的分类器

在图1-5上,在 F_s 为多维时,分类线是一个超平面,它实际上是无法用图来表示的。图1-5所示只是一种最理想的情况,概率分布曲线总要向 $+\infty$ 和 $-\infty$ 两个方向延伸,因而不同对象的概率分布曲线总是要相交的。分类器将按最小风险的原则来进行设计。

经过训练以后,系统设定在按 F_s 进行分类的状态下,在输入未知模式 ω 时,系统通过计算其 F_s ,最后作出这是物体A或B的判断(仍旧延袭前面所说的区别A、B两类的问题)。当然类别数,并不只限于两类。

稍后一些,大约在1984年左右出现了另外一种通用型的以边缘检测为基础的二维计算机视觉系统。这个系统从硬件配置来说因采用了解决专门计算任务的专用插件板而改善了系统的实时性,从软件设计来说则采用了结构模式识别的手段,由于这两个特点,因而也值得提出来作一介绍。正是由于专用插件板的思想,随着大规模集成电路技术的发展,近几年来(1987年以后)所出现的多种图像处理(或数字信号处理)专用芯片,使图像处理技术的实用化进入一个新的阶段。属于这类专用芯片的有美国的TMS-320系列和日本东芝的T9506等。它们都采用流水线和多处理机的手段以及附在芯片上的动态存储器,使一些常用的图像处理算法例如频域中的FFT,IFFT和空域中的几何变换和滤波等达到每帧秒级以下的速度。关于这些专用芯片的应用,由于不属于本书的范围,在这里将不再深入地加以讨论。

图1-6上示出以边缘检测为基础的二维系统的框图,图1-7上示出其边缘检测用的插件板的线路图。

这个系统也是一个二值系统,系统中的处理机采用Motorola 68000或Intel 8086。经摄像机采样输入的图像在数字化以后存放在帧存储器中。两块专用的插件板实际上是两具处理器,一具用来做图像平滑和边缘检测;另一具则用作直线提取和特征提取。插件板之间有专用的总线,使大量的数据可以直接连通。主处理器的工作是匹配和识别。

第一块插件板按 3×3 的邻域范围执行平滑、边缘检测和边缘检出的任务。图1-7中的九个寄存器就起着存放这一邻域范围的作用。图中的两个行存储器用来存放两行图像数据,并随着运算过程的进行,把这些数据顺次送入九个寄存器。图像数据先从帧存储器直接进入第一组三个寄存器,它们离开这一组寄存器后进入第一个行存储器。以后再由第一个行存储器进入第二组三个寄存器。由第二组三个寄存器出来的图像数据再进入第三组三个寄存器。通过平滑和边缘检测(它们都采用 3×3 邻域运算的方法)的计算结果重新回进插件板的帧存储器,因此该帧存储器内并不保留原始的数据。通过边缘检出环节得到的结果(它是每一行中提取

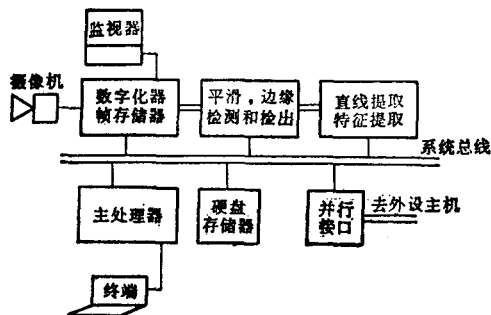


图 1-6 以边缘检测为基础的二维系统

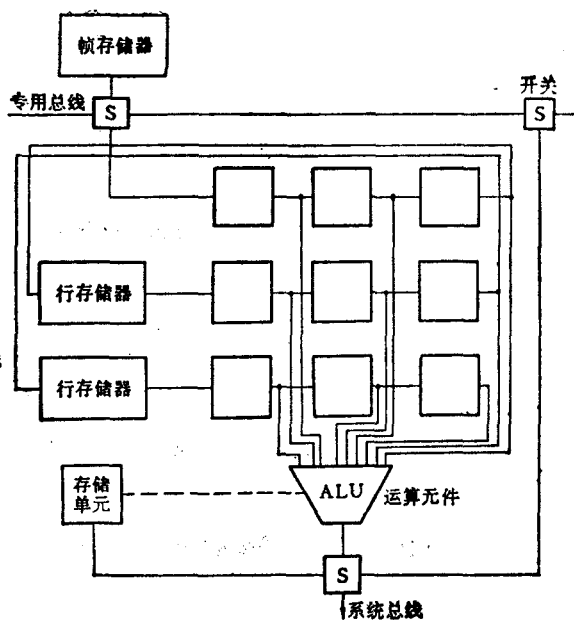


图 1-7 边缘检测专用插件板

出来的边缘点的几何位置)则被送到下一块处理器板作直线提取。

经过一定取阈后从边缘检测的结果中可以检出许多边缘点来，它们应由于所处环境的不同而加以标记，这些标记为

- IP: 四周无其他边缘点的孤立点，
- EP: 一行边缘点末尾的边缘点，
- LP: 一行边缘点中间的边缘点，
- ND: 边缘点聚集成的结点。

检取出来的各边缘点(实际上是它的坐标,占一个字节存储容量)以及附在其上的标记(也占一个字节)被送到下一块插件板作直线提取。

直线提取环节先是在每一行中提取出来的边缘点连成一个边缘表，这一过程基本上与上一系统的连接成块体表完全相似，它共有五个算符，它们是：

算符 *I*(列表)：一个新的边缘的起点将作为边缘表中的一项。所谓新边缘的起点是指在上一行数据内或是在同一行的左侧没有与它邻接的边缘点；

算符 *C*(连接)：如果一个边缘点在上一行内有一个邻居或是在同行的左侧有一邻居，则它将连接到表内的一项中；

算符 *D*(消除)：边缘表内所列的都是待接续的边缘，到达边缘的终点以后，这条边缘就不再列为待接续的项；

算符 *F*(分叉)：如一个边缘点在它的下一行有两个相邻点，则分叉以后的边缘分列为两项，并在分叉点上用指针指出它们相互之间的关系；

算符 *M*(混和)：如两条边缘交于同一边缘点，从该交点起作为一个新的边缘列入边缘表，并用指针联系它们三者之间的关系。

通过这一过程得到的将是许多连接在一起的边缘点，它们通过直线方程

$$Ax + By + C = 0$$

进行迭代拟合以后就得到了许多边缘线段。这些边缘线段的组合形成了被识别对象的轮廓线。众所周知，通过物体外轮廓线的匹配来对物体进行结构识别可以采用线图匹配或造句法。但线图匹配法有很大的计算工作量，而造句法则抗干扰性比较差，这一系统采用的是概率匹配法，它的适应性比较强，因而有其实用价值。

在概率匹配法中，模型是一个三层的符号结构，每一层的符号有它们各自的特征，这三层为：

- 第一层——线段，
- 第二层——顶点，
- 第三层——顶点表。

模型的最低层符号为线段，它由该线段的特征和向其他线段的连接所构成：

线段 LS ，它的附属特征为

L ：线段长度，

θ ：线段取向，

(x_1, y_1) ：线段的起点，

(x_2, y_2) ：线段的终点，

指针表：指向与该线段相连接的其他线段。

多条线段交会在一起形成第二级的符号，顶点。例如两条线段交会一起的顶点，其符号为 **Pair**。一个名为 **Pair** 的顶点的识别特征和属性如下：

顶点 **Pair**：

出现概率： $p(V)$ ；

识别特征：

相对角度： (μ, σ^2) ，

线段 1 的长度： (μ, σ^2) ，

线段 2 的长度： (μ, σ^2) ；

属性：

位置： (x, y) ，

方向： α 。

连接在一起的顶点表：

在上列的识别特征中，相对角度和线段长度都以概率密度曲线来表达， μ 为均值， σ^2 为方差。引入概率密度的概念使匹配具有相当的弹性。一个模型的三层结构中的识别特征和属性等基元是通过输入样本进行训练而建立起来的，它实际上是顶点基元构成的一个网络。

与以前相同，对未知模式的识别实际上是把该模式与模型相匹配。最靠近的模型就是识别的结果。因而识别的过程实际上是一个搜索的过程，当识别特征以概率密度曲线表示时，搜索可先按高概率密度的线段进行，这样可以节省搜索的时间。

上面所介绍的两个系统虽然各有其特点，但归纳起来说，基本上由三个环节所组成。第一个环节是目标提取。在第一个系统中所提取的是二值化的块体，在第二个系统中所提取的是块体的二值化的边缘。如果边缘提取做得好，则后一个系统有可能提取到运动物体的交接部

分,如单纯提取二值化的块体,则这类信息是得不到的。第二个环节是符号表达,在前一个系统中采用链码来表达图义。在后一系统中采用线段、角的符号来表达图义。因为现在的冯诺伊曼(von Neumann)机器所能处理的只能是数字和符号,因而通常用符号阐述图义以便进行识别大概是不可避免的。当然还可以有以后将要介绍的人工网络的方法,这将从另外一个角度去考虑识别问题。第三个环节是匹配,依靠预先存储在计算机中的模型去识别输入的未知模式。这预先存储的过程是一个训练的过程。看起来,除非另辟蹊径,作为一个视觉系统来说,都必须含有这三个环节。

1-2 由二维向三维推广

上面所介绍的两个二维系统都对输入图像信息进行了二值化处理,正是这二值化处理使包含在图像中的三维信息丧失殆尽。二值化处理所提供的物体的外型轮廓仅仅只是三维对象的一个剖面的投影。并且为了使二值化处理能得到良好的结果,对象的光照条件必须考虑,光线要均匀,不应该出现物体的阴影,背景与物体应有足够的反差。事实上有些轮廓上的线条不管怎样使用确当的光线都是无法使之分明起来的。因而只有保留灰度的层次,对这些原始资料进行处理方有可能进入三维视觉系统的领域。上面所介绍的两个系统所做的工作实际上还只是分类,还不能进到理解图义的足够深度。例如,要把物体的阴影与物体区别开来,要把交叉叠置的物品区别开来,这些问题都不是两值化所能解决的。也许可以这样说,到分类为止,这还是一个信号处理的问题,是通过数学途径可以解决的问题;而理解则是牵涉到人类智能的问题,有许多看起来是非常简单的人类智能是否可能在计算机上复现,至今还处于争论之中,很难有定论。视觉理解也属于这种类型的智能。因而就目前来说,要在现有的计算机上实现带通用实用意义和具有理解智能的三维视觉系统,看来还不能有太乐观的估计。但是,结合一个局限性的应用范围,设计和构成三维视觉系统,即使使用现在的微机系统还是完全有可能的。因而对于三维视觉系统,不太过于奢求它的模拟人的智能的通用性,也许是比较明智的决策。下面就结合若干实际应用介绍几个三维视觉系统,藉以说明由三维系统带来的若干新的问题。

六腿漫步机器人

六腿漫步机器人是1987年由美国国家宇航局所提出,并由卡内奇-梅隆大学机器人研究所研制的。研制该机器人的目的是在外星球表面观察和收集有关星球物理、气象和生物状况的资料。要求机器人具有自动行走、导航和采集样本的能力。机器人还应该能在极恶劣的环境下进行工作,其中包括低温、强风和尘埃等。作为机械机构来说,机器人的每三条腿结成一组,它可以跨越一米高的障碍并可在 30° 的斜坡上爬行。按照观察前方、规划行进道路和挑选样品的不同要求,机器人的视觉系统具有两种不同的功能;其一为对近旁范围的细致的视觉理解,其二为把崎岖的地形表达成等高线地图(elevation map)和障碍物的大范围的三维描述。这种立体的多分辨率的视觉特征,从某种意义上来说,多少复现了人的视觉功能。

机器人的视觉系统带有两种视觉传感器,第一种是激光扫描器,它用来直接测量距离;另外一种彩色摄像机,它用来从色彩和纹理识别物体以及进行立体视觉。

视觉系统从距离图像(range images)建立起多分辨率的等高线地图和在其上以符号标出的障碍物的位置,这是储存于机器人视觉系统知识库中关于周围环境的理解。在行进时,

机器人根据等高线地图来选择每一步的位置和路线。作为行进方向的引导，机器人把自己的位置也标到等高线地图上以设计行进的道路。在采样时，机器人根据等高线地图和在其上标出的目标以选取样品和确定如何对样品进行处理。

介绍这一个机器人，其目的在于举例以说明一种专门的三维视觉系统的功能，它是多分辨率地理解地形和目标(包括样品和障碍物)，所采取的检测手段为激光测距和距离图像(类似于人类左右眼所得的图像)。

SDI 中的三维视觉问题

SDI(strategic defense initiative) 战略主动防御(又称星际大战)是美国五角大楼自1983年起已经耗费大量资金并将继续投入大量资金(以百亿美元计,每年的投资数居最重要科研项目之冠)以进行研究的项目。SDI 项目包括很多方面,这里仅着重于指出其中的三维视觉问题。SDI 企图采用拦截对撞的形式来销毁“敌”方带攻击性的洲际飞行器,使“敌方”飞行器所带的攻击弹头不能到达“己”方的重要目标。拦截将在三个阶段中进行(见图 1-7)。第一阶段为发射阶段,敌方的攻击弹头在火箭 M 的助推下发射起飞,历时约 180 s。第二阶段为后发射阶段,攻击舱将与火箭分离,此后攻击舱将释出弹头和诱饵,历时也在 180 s 左右。第三阶段为中制导阶段,弹头在制导之下向攻击目标行进,这阶段历时在 1300 s 左右。图 1-8 为第一阶段的示意图。第一阶段是最有利的攻击时间,因为在这阶段被攻击目标只是一个。在图 1-8 中, SBI 为空间拦截器, BSTS 为地球同步监视卫星, SSTS 为第二阶段监视卫星, CC 为指令中心, GBR 为地面雷达站。

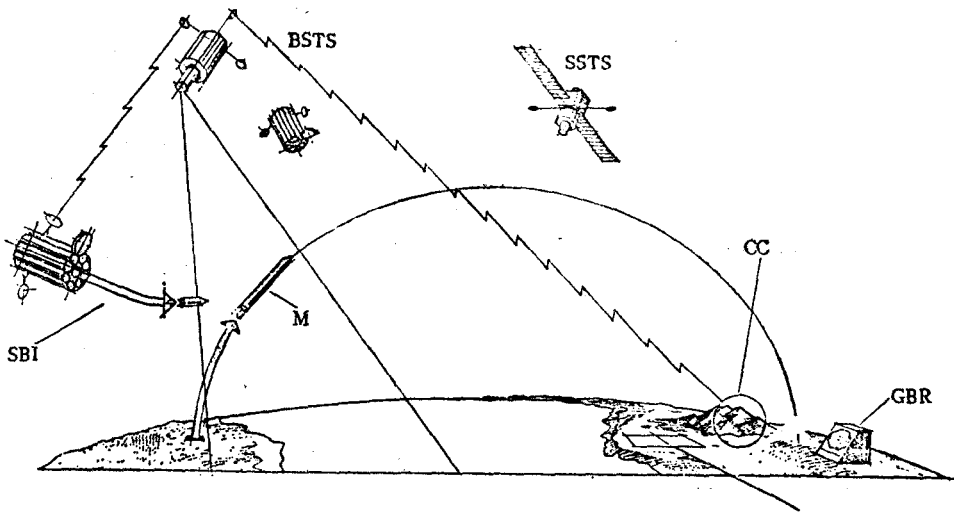


图1-8 SDI 示意图

BSTS 运行于地球同步轨道,因而是一族定点卫星。BSTS 的监视作用有四个方面,这些作用都是通过多波段近红外传感器来实现的。第一是检测出所监视地区有否发射事件发生。第二是对敌方火箭的尾喷进行多波段分析和模式识别,从燃料识别敌方火箭的类型。第三是识别敌方火箭的轨迹(这是三维视觉问题)。第四是预报敌方火箭的位置,这实际上是识别轨迹的继续。数据由 BSTS 传送到控制中心 CC,由它进行决策和对 SBI 下达击落敌方火箭的命令。

经过第一阶段的 180 s 以后,敌箭 M 飞出 BSTS 的监视范围,于是第二阶段监视卫星

SSTS 起监视作用。SSTS 是一个中等高度的轨道卫星,它的监视或识别功能如下:第一是识别敌方火箭及其释出物的轨道并进行预报;第二是识别对方释出的战斗弹头和诱饵,这些都是三维视觉问题。同样地,控制中心 CC 接受数据并作出决策命令其他轨道上的 SBI 对敌方飞行物进行拦击。进入第三阶段时,被监视物只是战斗弹头和诱饵,它们将进入 SSTS 的正下方,这时背景由低温的太空变成带有一定温度的地球表面。进入这种状态时, SSTS 上的多波段长波红外传感器失效,在控制中心 CC 的命令之下,基地卫星 GSTS(图上未示出)将起飞,它将在较低的轨道上继续检测和预报敌方飞行体的轨道并对弹头和诱饵进行识别。与此同时,地面雷达站也参与敌方飞行体的轨道监视工作。这时,因敌方飞行体逐步临近,视场中检测出来的像开始成形,因而这时的识别还包括对敌方飞行体的攻击致命点的选择。

我们只是从计算机视觉的应用来研讨这一问题的,可以看出,这里包含着大量的视觉识别问题。由向量(多波段)特征分类进而对三维轨道识别,由运动点的识别进而对运动面的识别,由点的轨迹进而对面的形状识别,其中包含的内容是极为丰富的。

运动员的训练

以计算机视觉技术来辅助运动员的训练是一个非常令人感兴趣的方向,目前已在多项运动训练中得到了使用,也取得了很令人满意的结果。图 1-9 所示为奥运会游泳金牌获得者 M. 比昂迪(Biondi)自由泳训练时对手臂划动一周所作的图像分析。在运动员训练时,用两具水下摄像机摄取手动作的序列图像。由序列图像计算出手的划动速度,求出由手所产生的作用力,它是图 1-9 坐标上的上面一条曲线。再由序列图像求得人体的前进速度,由此求得有效的作用力,它是该坐标上的下面一条曲线。由曲线可看出,在曲线的前端,手的作用力并未全部转化成有效的推进力,这是因为在手臂伸向头部的前端时,手掌或手臂的姿势不对,因而形成了阻力,影响了身体的前进速度。这样应该校正的就是游泳的姿势了。如果作用力和有效力的曲线相当吻合,但速度不够理想,则应该采取的措施,就是肌肉训练,使作用力提高。可以看出,这种训练分析方式为提出正确的训练方案提供了理论上有力的依据。

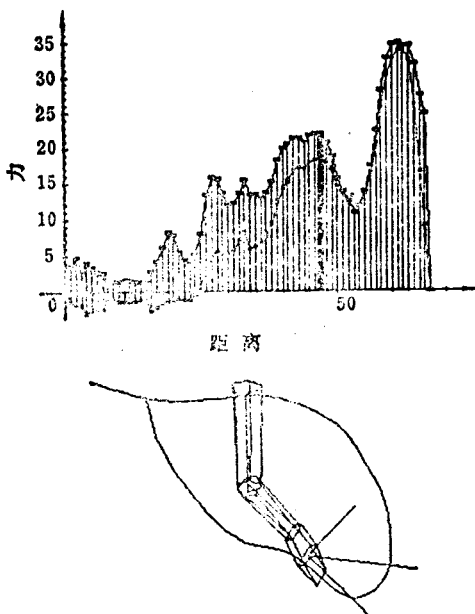


图 1-9 计算机视觉技术用于对游泳运动员的训练

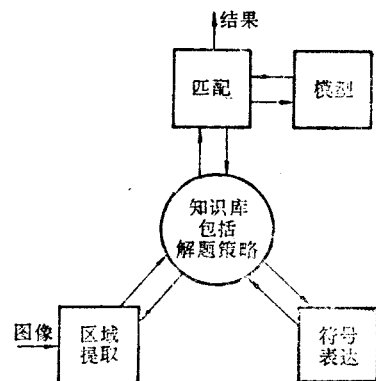


图 1-10 建议的三维视觉系统