

基于级联神经网络的实用型三维复合 不变性多目标识别*

刘 玥 陈 成 郭鹏毅 张文伟 申金媛 张延焯

(南开大学现代光学研究所, 国家教委光学信息技术科学开放研究实验室, 天津 300071)

摘 要 以三种飞机模型作为待识别目标, 模拟真实场景, 对用于多目标分类识别的级联神经网络重新进行了研究。实验发现识别率下降的主要原因是实际采集的目标发生的复杂畸变与计算机模拟产生的效果并不一样。用采集得到的目标图像作为训练样本, 对网络重新构造和训练, 取得了好的实验结果。分析了其中涉及到目标定位、图像分割等图像预处理问题。提出了一种基于二值图像形态学腐蚀运算的快速目标检测定位法, 可快速有效地对目标进行检测定位。

关键词 模式识别, 级联神经网络, 图像预处理, 腐蚀运算。

1 引 言

目标识别是人工视觉的重要分支, 在军事预警、交通管理、目标跟踪、无人驾驶及自主机器人等领域有重要意义。通常识别既要区分相近的不同目标(即识别精度高)又要在同一目标发生畸变时(旋转、尺度伸缩或平移)不至误判(即识别不变性), 导致识别过程非常复杂, 信息量和计算量都很大。然而任何有实用意义的识别又必须在足够短的时间内完成(即识别速度高)。基于人工神经网络的多信息目标识别技术在目标识别方面具有巨大的优越性^[1]。我们已完成了将级联前馈网络用于三维目标模式识别的工作^[2]。但是这些工作着重于方法的研究, 输入的待识别目标模式由计算机模拟产生, 因而在推广应用时遇到了障碍。因为实际应用中目标是实时从外界采集来的, 通常包含多种复合畸变, 与计算机模拟产生的结果不同。本文以实用化为目标, 用采集到的目标图像作为训练样本, 对网络重新进行了构造和训练。

基于光学运算实现的模式识别和分类中令人们感兴趣而且有实现意义的一种方法是利用三维物体的二值投影图像完成对三维目标的识别和分类^[3]。目前实时图像的采集通常由 CCD 摄像机实现。但这样得到的图像是一幅很复杂的多灰度图像。要得到简单的二值投影图, 还必须首先完成图像二值化、目标定位、局域图分割等准备工作, 即图像预处理。有关文献就目标检测定位分割已提出的方法有矩心法^[4]、基于神经网络算法的振动相关法、遗传算法^[5]等等。这些方法的运算量相当大, 不满足实时处理的条件。图像形态学是一种从几何的角度出发进行图像处理的方法, 在数字图像处理中已经有很好的应用。二值腐蚀运算可以在一幅

* 国家自然科学基金(69877005)资助项目。

收稿日期: 1998-11-04; 收到修改稿日期: 1999-03-26

二值化图像中找出具有预期几何形状的目标^[6]。本文在解决基于光电混合实现的人工神经网络多目标分类系统图像预处理问题中,提出了根据目标的基本几何结构,构造新的形态结构元素,并通过修改图像形态学二值腐蚀运算进行目标检测定位的方法。

2 识别预处理

2.1 图像形态学中腐蚀运算基本原理

图像形态学中腐蚀运算的定义如下^[7]:

设 $E^{(n)}$ 中的集合 $X(x \in X)$, 结构元素集合 $B(b \in B)$ 为 $E^{(n)}$ 或其子空间 $E^{(m)}$ 中的一个点集, 那么, 集合 X 关于结构元素集合 B 的腐蚀定义为如下点集:

$$X \ominus B = \{x: B_x^* \subset X\}, \quad \subset \text{表示包含于}, \quad (1)$$

$B_x^* = \{x - b: b \in B\}$, 即集合 B 关于原点对称后沿向量 x 的平移。

本文以下所讨论的图像均指二值化图像。实现二值腐蚀运算的方法有很多, 其中易于计算的一种是通过逻辑与(AND)运算实现。设 $f(x, y)$ 为原始图像, $g(x, y)$ 为腐蚀运算后得到的图像, H 为结构元素集合, 则有:

$$g(x, y) = \begin{cases} 0 & \text{AND}_{i,j \in H} \{f(x-i, y-j)\} = 0, \\ 1 & \text{AND}_{i,j \in H} \{f(x-i, y-j)\} = 1, \end{cases} \quad (2)$$

计算结果可以由图 1 直观地理解。可见, 特定的结构元素可以有效地检测出具有期望几何形状的目标。例如: 若结构元素为圆形, 则相同面积的三角形、矩形或其他不规则形状将被“腐蚀”掉而不会被检测到。

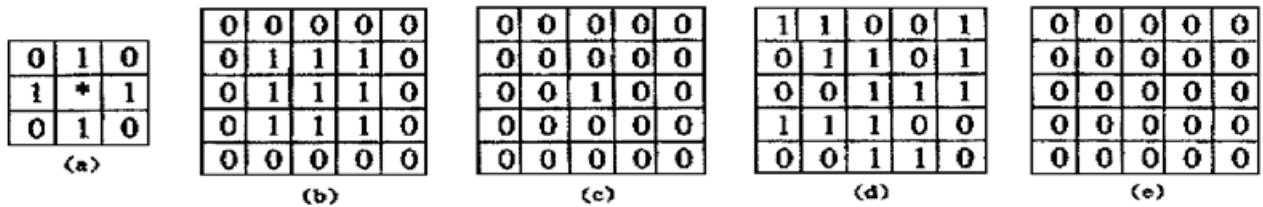


Fig. 1 (a) Structuring element H , (b) Binary image $f_1(x, y)$, (c) Binary image $g_1(x, y)$, (d) Binary image $f_2(x, y)$ and (e) Binary image $g_2(x, y)$. $g_1(x, y) = f_1 \ominus H$, $g_2(x, y) = f_2 \ominus H$ and * represents (0, 0) point

2.2 实现算法

预处理流程图如图 2 所示。所面临的实际问题是实时地用 CCD 从外界摄取目标的图像, 然后用计算机进行数字图像处理得到二维投影图像信息, 最后输送到后续的人工神经网络模

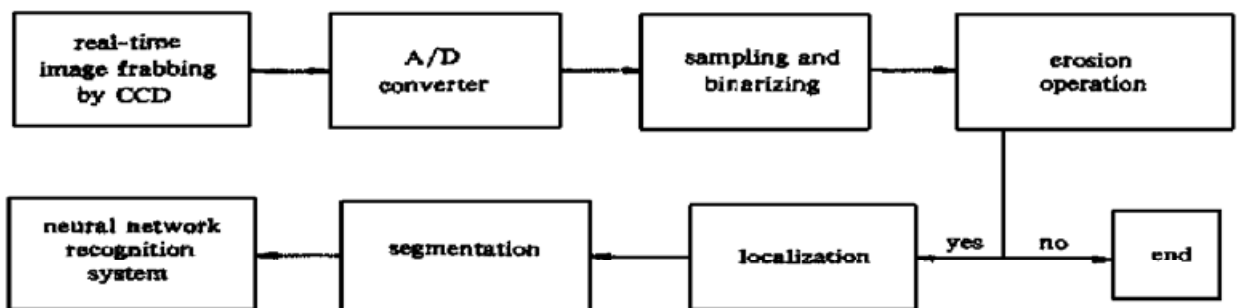


Fig. 2 The flow chart for image pre-processing

式识别系统进行识别。图像的空域统计特性表明二维空间图像相邻像素间具有很强的相关性，因为图像的内容和亮度在空间上总是渐变的。为了减少运算量，可以采取等间距隔点欠采样的方法压缩图像数据。这种压缩方法对于提高处理速度非常有效，而且为粗略定位运算提供了足够的信息，保证了腐蚀运算定位法的可行性。因为后续的认可系统允许目标作平面内 $0^\circ \sim 360^\circ$ 的旋转，即不考虑目标在图中出现的方向，所以结构元素构造成圆对称(离散情况下为“米”形)。算法中以取阈的方法实现逻辑与运算，则(2)式改为(3)式：

$$g(x, y) = \begin{cases} 0 & \sum_{i,j \in H} f(x-i, y-j) < T, \\ 1 & \sum_{i,j \in H} f(x-i, y-j) \geq T, \end{cases} \quad (3)$$

阈值 T 通常由(4)式得出：

$$T = \sum_{i,j \in H} H(i, j), \quad (4)$$

$g(x, y) = 1$ 时，表明可能有目标存在， (x, y) 即为目标中心点。

在实验中，设计了两个尺度不同的“米”字形，以允许目标有一定范围的尺度伸缩^[8]。图 3 给出一个原始采集图像和分割结果的例子。其中原始图像为 256 阶的灰度图，分割的目标图为二值图。

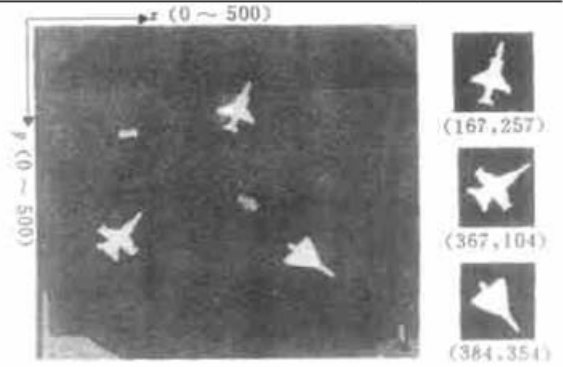


Fig. 3 The original gray-level image (500 × 500 pixel) and binary local images obtained by segmentation (80 × 80 pixel, coordinates of center points are given out in brackets). All images shown are not in real size

3 级联神经网络识别

3.1 级联模型

级联模型是一种多层网络，它将几个独立网络前后连接而成，前一层的输出即为后一层的输入。该模型既保持了多层前馈网络的映射能力，又避免了学习算法的复杂性，网络结构有较大的灵活性。级联模型中每一子网络都有给定的功能，并单独完成其互连权重的学习和选择。用级联模型来解决多目标识别问题时，识别精度和识别不变性的要求可以分别赋予不同的子网络来完成。由于级联网络在输入层完成目标聚类(clustering)和编码的同时也实现了大幅度的降维和信息压缩，因此采用前光后电的级联型硬件结构，其处理速度将是很快的。

本工作构造的级联模型由两级子网组成，如图 4。第一级子网是一个多对一的异联想网络，第二级为一个 WTA 网络。第一级的输入即为所采目标的二值形状投影图，输出应为给

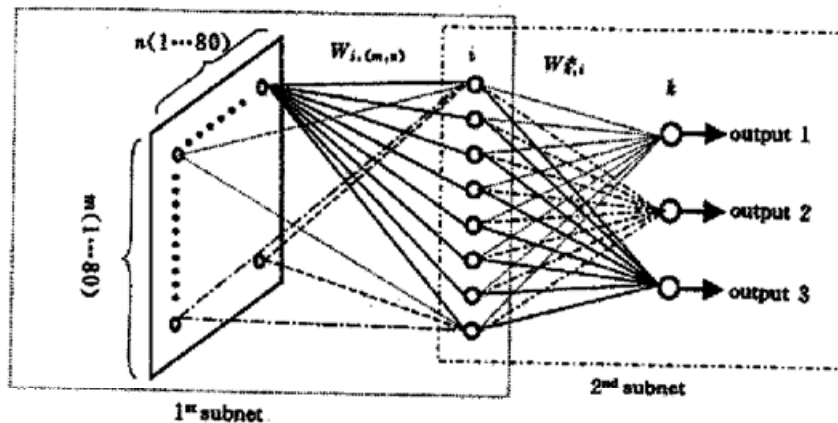


Fig. 4 The cascaded model of neural networks

定的编码值。因所取图像大小为 80×80 pixel，所以输入层网络规模为 80×80 。第一级的输出层的规模取决于给目标编码的码长。对于三类目标，本工作的编码长度为 8 位，所以其输出层的神经元数为 8。编码长度的确定原则和范围，将另文讨论。为了便于光学实现，用 1 和 -1 对三类目标进行编码，编码结果如表 1 所示。 C_i^k 表示第 k 个目标的第 i 位码。第一级子网的输出层亦为第二级子网的输入层。对于三类识别目标，第二级子网的输出层，即最后一层神经元数为 3。这样，上述网络的拓扑结构为 6400-8-3。第一级网络维数高，计算量大，适合用光学方法实现。第二级网络由于计算量小，可在计算机中完成。图 5 为在实验中实现异联想网络的光学结构。

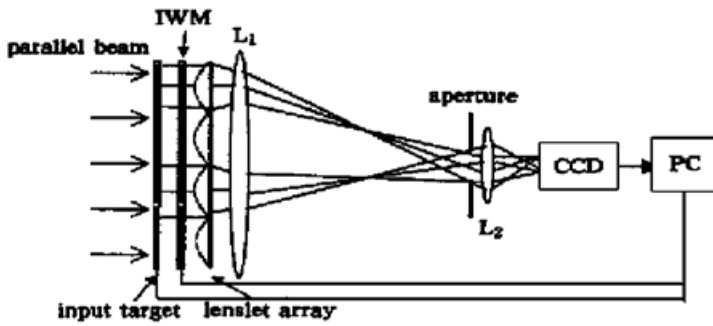


Fig. 5 Optical implementing system of 1st subnet (hetero-associative model). LCDs are used as input devices for targets and interconnection weights

平行光照明输入平面(输入目标被复制 8 次，以与 8 个互连权重子矩阵相对应)，透射光照射到互连模板(IWM)，实现目标与互连权重相乘。透过每个互连子矩阵的光经小透镜阵列上对应位置小透镜及成像透镜 L_1 和 L_2 会聚于 CCD 的靶面上。经转换后输入计算机进行取阈操作，然后由计算机完成 WTA 网络的运算。光阑可部分地消除衍射效应。

Table 1. The codes of three targets

code	C_i^k	C_1^k	C_2^k	C_3^k	C_4^k	C_5^k	C_6^k	C_7^k	C_8^k
plane A ($k = 1$)		1	1	1	1	- 1	- 1	- 1	- 1
plane B ($k = 2$)		1	- 1	- 1	1	1	1	- 1	- 1
plane C ($k = 3$)		- 1	1	- 1	1	1	- 1	1	- 1

3.2 权重学习

设 1、2 两层的互连权重为 $W_{i, (m, n)}$, ($i = 1, 2, 3 \dots 8$; $m, n = 1, 2, 3 \dots 80$)，第 2、3 两层的互连权重为 $W_{k, i}^*$ ($k = 1, 2, 3$; $i = 1, 2, 3 \dots 8$)， $W_{k, i}^*$ 由编码唯一确定。所以关键在于如何训练权重 $W_{i, (m, n)}$ 。对于光学实现，权重通常只取 1, 0, -1 三值。但是为了便于学习，学习时不控制权重灰阶，权重可以取任意实数。设输入样本为 80×80 二维矩阵，表示为 $X_{m, n}^k$ (k 表示第 k 类目标， $m, n = 1, 2, 3 \dots 80$)，则第二层第 i 个神经元的总输入 U_i^k 和输出 Y_i^k 为

$$U_i^k = \sum_{m=1}^{80} \sum_{n=1}^{80} W_{i, (m, n)} X_{m, n}^k, \quad Y_i^k = f(U_i^k - \theta_i), \quad (5)$$

一般 $f(x)$ 为硬取阈函数，即

$$f(x) = \begin{cases} 1 & x > 0, \\ -1 & x \leq 0, \end{cases} \quad (6)$$

对于任意输入样本，它的对应输出应为前面规定的编码值，即

$$Y_i^k = C_i^k, \quad (7)$$

此式不满足时，需要对权重进行学习，迭代算法如下：

$$W_{i, (m, n)}(t + 1) = W_{i, (m, n)}(t) + \alpha(C_i^k - Y_i^k)X_{m, n}, \quad (8)$$

α 为学习步长，取任意小于 1 的正实数。学习开始时，对初始权重矩阵赋予随机值，直到对于每个选定样本，(7) 式的条件都满足，学习结束。最后用截值 (clipping) 的方法将互连权重灰度阶压缩到 (1, 0, -1)，即

$$W_{i,(m,n)} = \begin{cases} 1 & W_{i,(m,n)} > a_1, \\ 0 & a_2 < W_{i,(m,n)} < a_1, \\ -1 & W_{i,(m,n)} < a_2, \end{cases} \quad (9)$$

式中 a_1 、 a_2 为给定的阈值。本工作提出了自适应截值模型寻找最佳截值点的方法, 这种方法能够最大限度地保留多灰阶权重的识别性能^[9]。

人们已经提出偏振编码、颜色编码、方向编码等方法来解决光学计算负值表示的难题, 本文采用正负通道相减的方法来等效表示出正负, 既简化了系统结构, 也可通过正负相减来部分消除光源不稳定的影响。但是如果编码长度很大时, 这种正负通道的方法对空间带宽要求过高, 这是一个缺点。作全双级运算时, 把权重值和输入值 S_i 都表示成正负两部分:

$$W_{i,j} = W_{i,j}^+ - W_{i,j}^-, \quad S_i = S_i^+ - S_i^-, \quad (10)$$

则第 j 个神经元的总输入为

$$U_j = \sum_i W_{i,j} S_i = \sum_i (W_{i,j}^+ - W_{i,j}^-) (S_i^+ - S_i^-) = 2 \sum_i (W_{i,j}^+ S_i^+ + W_{i,j}^- S_i^-) - \lambda_j. \quad (11)$$

对于 8 位码长的编码, 共应给出 8 个权重模板; 若分为正负通道, 将给出 16 个模板。

4 实验结果

图 6 为模拟实时采集-识别系统简图。首先在与将来应用时相似的条件(距离、照明条件等)采集训练样本, 三类模型各采集样本 130 个, 包括面内旋转 $0 \sim 360^\circ$, 面外旋转 $\pm 25^\circ$ 。图 7 为一组样本例子。训练得到的权重模板如图 8 所示(给出 16 个正负通道模板)。对于训练样本, 三类目标识别率均为 100%。对于在模拟环境下实时采集的非学习样本目标, 用本工作设计的小型光电混合多目标识别系统进行识别, 三种目标统计识别率均达到 90% 以上。

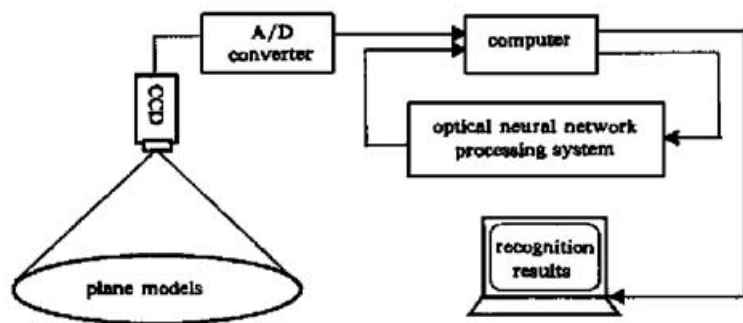


Fig. 6 The schematic diagram of real-time recognition system

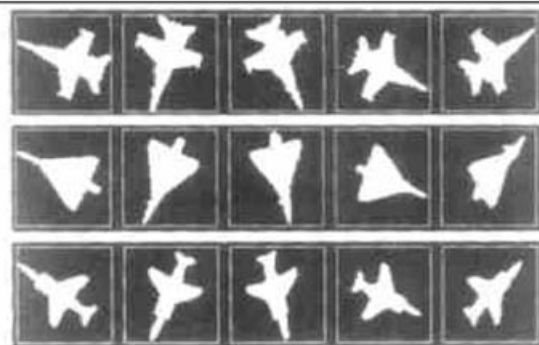


Fig. 7 Binary projective images of three kinds of planes

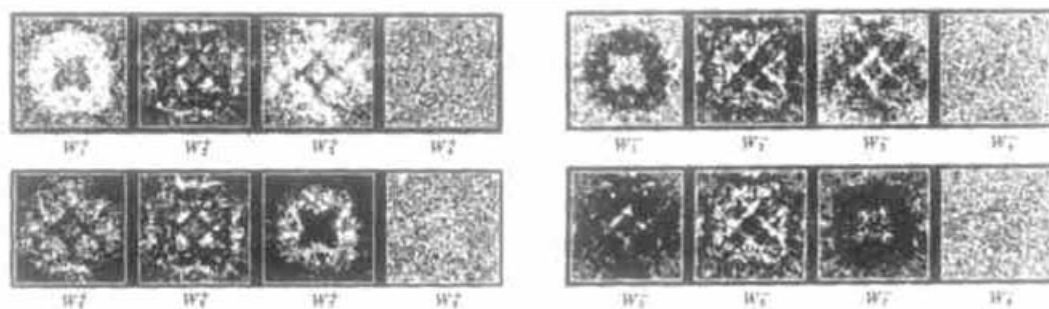


Fig. 8 Interconnection weight masks

从采集到最后给出识别结果,所用时间不超过 2.0 秒。

结 论 三维目标识别,尤其是目标发生旋转、平移、尺度伸缩等复杂的复合畸变时,已有计算机和光学相关方法都很难实现理想的识别。级联神经网络把识别复杂性分散到各子网络,取得了很好的识别效果。本文所构造二值图像形态元的目标检测法算法简洁,速度快。这种预处理方法与级联神经网络相结合,较好地实现了对发生复合畸变的三维目标的实时识别。

参 考 文 献

- [1] 张立明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1994. 1~ 6
- [2] 李豫华. 多目标分类识别的级联神经网络研究(博士学位论文). 天津: 南开大学, 1996.
- [3] Lee S H. Optical implementations of digital algorithms for pattern recognition. *Opt. Engng.*, 1986, **25** (1): 69~ 75
- [4] Freeman M O, Saleh B E A. Optical location of centroids of non-overlapping objects. *App. Opt.*, 1987, **26**(14/15): 2752~ 2759
- [5] Huh H, Pan J K. Optical/digital invariant recognition of two-dimensional patterns with straight lines. *Opt. Engng.*, 1996, **35**(4): 997~ 1002
- [6] Weick Jr. A R. Fundamentals of electronic image processing. Bellingham(Washington USA): *SPIE Optical Engineering Press*, 1996. 316~ 332
- [7] 吴敏金. 图像形态学. 上海: 上海科学技术出版社, 1991. 5~ 12
- [8] 刘 珩, 申金媛, 张延焯等. 模式识别图像预处理中的目标快速定位法. *光电子·激光*, 1998, **9**(5): 415~ 418
- [9] Chang S, Song Z, Jin C *et al.*. Adaptive clipped model and its optical implementation. *Optik*, 1998, **107**(4): 135~ 140

Real-Time Recognition of Multi-Targets with Complicated Distortion Based on Cascaded Neural Networks

Liu Yue Chen Shu Guo Pengyi Zhang Wenwei
Shen Jinyuan Zhang Yanxin

(*Institute of Modern Optics, Nankai University,
Optoelectronics Information Science and Technology Lab, MOE, Tianjin 300071*)

(Received 4 November 1998; revised 26 March 1999)

Abstract Further work was done on the cascaded neural network system for multi-target recognition. Three plane models were used as targets to be recognized to imitate the real scene. It is found that the recognizing rate is lowered dramatically mostly due to complex distortions occurring in applications, which are very different with that stimulated by computers. Utilizing images of the models as training samples, the network was reconstructed and retrained. Good results were obtained. And to approach image preprocessing problems, such as location and segmentation, an efficient method of quick target detection and location is proposed based on binary morphological erosion algorithm.

Key words pattern recognition, cascaded neural networks, image pre-processing, erosion.