

三维运动目标的多尺度智能递推识别新方法¹⁾

张天序^{1,2} 翁文杰¹ 冯 军¹

¹⁾(华中科技大学图像识别与人工智能研究所 武汉 430074)

²⁾(图像信息处理与智能控制教育部重点实验室 武汉 430074)

(E-mail: txzhang@hust.edu.cn, hustwengwenjie@tom.com, funjun@hotmail.com)

摘 要 在实际成像条件下, 运动中的三维目标, 其投影形状 (Silhouette) 是变化的, 因而其可识别性也处于变动中. 为了应对这类困难情况, 本文定义了模式的动态特征空间和模式的动态可识别性等概念. 讨论了建立多尺度三维目标特性视图特征模型的必要性, 以及将目标运动特性一般约束用于目标序列图像识别的合理性. 据此, 提出了处理三维目标运动图像序列的多尺度智能递推识别方法 (MUSIRR). 构造了一种混合神经网络和逻辑决策模块的智能识别器, BP 神经网络和 RBF 网用作识别器的基本构成单元. 在训练阶段, 该识别器使用目标的多尺度二值特性视图模型的规则矩不变量为样本特征向量. 在识别阶段, 算法在递推识别序列目标图像过程中, 充分利用了目标姿态不会突变以及有关成像过程的合理约束, 达到了提高识别率目的. 与文献中的基于单尺度特性视图的三维目标识别方法相比, 本文的方法训练过程简单, 只需较少的目标特性视图模型样本, 不仅能处理单帧图像, 更能有效处理序列图像. 对几类飞机目标的大规模模拟实验结果证实了本文方法的合理性和有效性.

关键词 三维目标识别, 运动目标识别, 动态特征空间, 动态可识别性, 规则矩不变量, 多尺度特征模型, 计算机视觉, 模式识别

中图分类号 TP391.41

A Novel Multi-Scale Intelligent Recursive Recognition Method for Three-Dimensional Moving Targets

ZHANG Tian-Xu WENG Wen-Jie FENG Jun

¹⁾(Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074)

²⁾(State Key Laboratory for Multispectral Information Processing of P.R.China)

(E-mail: txzhang@hust.edu.cn, hustwengwenjie@tom.com, funjun@hotmail.com)

Abstract In practical imaging condition, the silhouette of a 3D moving target is changing, therefore its recognizability is variable. In this paper several definitions are given, such as dynamic feature space of patterns and dynamic recognizability of patterns. Necessity of multi-scale feature models of 3D target and rationality of using the general constraint for recognizing target image sequence are discussed. Based on these discussions, a multi-scale intelligent recursive recognizer (MUSIRR) is proposed for recognizing 3D moving targets, in which BP neural network and RBF neural network are the basic cells. During training,

1) 国家自然科学基金重点项目 (60135020) 和国家重点预研项目 (413010701-3) 资助

Supported by the Key Project of the National Natural Science Foundation of P. R. China (60135020) and the Key Project of the National Advanced Research of P. R. China (413010701-3)

收稿日期 2005-10-26 收修改稿日期 2006-1-23

Received October 26, 2005; in revised form January 23, 2006

regular moment invariants of the multi-scale binary characteristic views of the target model are used as the pattern feature-vector. During recognition, the algorithm sufficiently uses reasonable restrictions of imaging process and target poses which are not changed acutely to achieve a good recognition ratio. Compared with the algorithms based on single-scale characteristic view models in references, the training of the MUSIRR algorithm is easy and needs less samples composed of the target characteristic view models. The algorithm can not only treat single frame images but also treat image sequence more effectively. The rationality and validity of the approach are proved by the results of massive simulation experiments on several kinds of aircrafts.

Key words Three-dimensional target recognition, moving target recognition, dynamic feature space, dynamic recognizability, regular moment invariants, multi-scale feature models, computer vision, pattern recognition

1 引言

由二维图像识别三维物体是计算机视觉研究中的重要研究领域, 解决此问题的一个难点是如何有效地将二维信息与三维目标表示紧密联系. 三维目标表示是识别的基础, 主要分为两类. 一类是以目标为中心的表示, 另一类是以观察者为中心的表示. 神经生理学方面的许多证据支持后者的合理性^[1~3]. 在这方面, 许多文献研究了所谓方位图表示 (Aspect graph). 沿着这种途径, 首先要计算出目标的一组有限的二维投影视图, 希望这一组特性视图集合尽可能小, 而且能代表实际条件下无限数目的可能视图. 于是有如下映射: 扩展的高斯图像映射 (Extended Gaussian images)^[4]. 它从三维物体表面法线向一个称为高斯球的单位球上映射. 球面上的一个点对应于具有同样坐标方向的向量, 也称该球为观察球 (Viewsphere), 球上的点为视点. 具有该观察方向的目标的投影 (正投影或透视投影) 相当于从球面上该点观察该物体. 可将该球面划分为相联通的若干点集, 使得在某个点集上的所有点上观察该目标, 会得到特征相近或拓扑同形的视图, 该点集的代表即所谓特性视图 (Characteristic view)^[5~8]. 而在其它点集上观察, 则其视图是拓扑不同形或特征差异很大的. 本文根据三维飞机目标投影的二值图或其轮廓图矩不变量特征矢量的相似性来聚类产生特性视图.

Eggert 等指出不能仅从完美的分辨率的理论观点来研究和建立方位图模型^[8], 应研究尺度空间方位图, 但该文仅给出了对简单多面体生成尺度空间方位图的例子. 另外, 该文没有给出利用尺度空间方位图识别物体的算法和实验结果.

Seibert 和 Waxman 研究了三维飞机目标的二值特性视图的自动生成和识别问题, 提出并实现了基于大规模模拟集成电路 VLSI 的算法实现^[9]. 其训练方式是采用在高斯观察球面上的大量视点轨迹来实现的, 非常复杂. 另一个缺点是, 该文没有考虑多尺度问题, 没有考虑实际成像条件及模糊对特性视图的影响.

文献 [10] 也研究了基于三维飞机目标特性视图的数字式识别方法, 但该方法使用了数量庞大的特性视图作为模型来识别飞机目标, 而且同样没有考虑多尺度问题及成像条件和模糊对特性视图的影响, 没有考虑运动条件下的序贯识别.

文献 [11~14] 研究了真实三维目标的识别方法, 但它们只是利用了目标有限视点的信息来进行识别, 对于其它任意视点下获取的图像识别能力有限, 而且这些文献只考虑静态条件下的目标识别, 没有将目标运动条件下的序贯约束用于识别. 另外, 文献也没有考虑目标多尺度问题及成像条件和模糊对目标识别的影响.

目前公开发表的文献均从静态观点来考虑识别问题, 没有考虑在运动状态和实际条件

下三维飞机目标图像序列的识别问题. 目标在运动状态和由近及远的实际成像条件下, 若不考虑调焦, 目标图像的模糊和尺度变化非常自然的发生, 它们对识别的影响如何, 没有文献进行过报道^[15~18]. 另一方面, 运动产生的序列图像无疑提供了丰富的信息, 如何充分利用这些序贯信息, 提高运动目标的识别率是非常值得研究的.

本文试图从系统的观点来研究和解决三维运动目标的识别问题, 与已有文献不同的是: 1) 从模式的动态特征空间的观点出发, 审视三维运动目标的建模和识别, 指出可识别性是动态变化的; 2) 指出既然存在实际成像条件下的成像传感器的有限分辨率和模糊等约束, 则应建立基于多尺度特性视图的目标模型, 即三维运动目标的有效表示问题; 3) 提出一种多尺度智能递推识别三维运动目标序列图像的新方法, 该方法利用尽可能少的多尺度特性视图特征模型进行训练, 训练过程简单, 所构造的基于 BP 网和 RBF 网构件的识别器, 可利用运动目标的一般动力学特性约束递推识别目标图像, 能统一处理从模糊的小目标到较清晰的大目标的序列图像, 达到了比静态、单帧图像的更高识别率.

2 动态特征空间

在经典的模式识别理论中, 在一个问题域内, 分类器所运作的特征空间通常是固定不变的^[19]. 这种特征空间的定义和表达方式完全不适应复杂条件和动态场景下的目标识别问题, 应该加以扩展和推广.

定义 1. 动态特征空间

给定一组模式 $(\omega_i, i = 1, 2, \dots, I)$, 每一类的模式向量 $\xi_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iD})$ 在观察空间 Ω 中将构成独特的一个子空间 Ω_i . 令模式向量经特征提取映射 A 将 D 维观测空间变换为 d 维的特征空间, 每一类的模式特征向量 $\xi_{fi} = (\xi_{fi1}, \xi_{fi2}, \dots, \xi_{fid})$ 在 d 维特征空间 Ω_f 中将构成独特的子空间 Ω_{fi} , 若 $\{\xi_{fi}\}, \{\Omega_{fi}\}$ 均为时间的函数, 则我们称 Ω_f 为动态特征空间. 这里 $\{\xi_{fi}\}, \{\Omega_{fi}\}$ 将以某种系统的方式改变 (这里不包括随机干扰或随机噪声所引起的改变). 比如, 模式特征向量概率分布密度的变化, 类别可分性的变化, 模式新特征的产生或消失等等.

模式特征空间的动态变化的例子在成像目标识别中不胜枚举. 例如, 对光学成像的目标识别, 光照条件、气候条件的变化将使同一个场景的目标 / 背景图像发生显著变化. 运动中的三维飞机目标, 其姿态、大小和形状特征 (由远及近, 或由近及远) 在序列图像中呈现剧烈的变化. 因此, 复杂条件下的目标识别问题就是典型的动态特征空间中的模式识别问题. 其处理方法, 包括性能分析与预测^[16~18, 20~22], 与静态特征空间有本质的区别. 有必要以统一的观点, 系统地考察这一问题.

特征向量 ξ_{fi} 随时间的变化, 可写为

$$\frac{d\xi_{fi}(t)}{dt} = \left(\frac{d\xi_{fi1}(t)}{dt}, \frac{d\xi_{fi2}(t)}{dt}, \dots, \frac{d\xi_{fid}(t)}{dt} \right) \quad (1)$$

若把因三维目标运动而引起变化的经纬度观察角 $\theta = (\theta_1, \theta_2)$ 、变化的观察距离 s 作为 (隐含时间 t) 变量, 则式 (1) 的另一种表达可写为

$$\frac{\partial \xi_{fi}(\theta, s)}{\partial \theta_j} = \left(\frac{\partial \xi_{fi1}(\theta, s)}{\partial \theta_j}, \frac{\partial \xi_{fi2}(\theta, s)}{\partial \theta_j}, \dots, \frac{\partial \xi_{fid}(\theta, s)}{\partial \theta_j} \right), \quad j = 1, 2 \quad (2)$$

$$\frac{\partial \xi_{fi}(\theta, s)}{\partial s} = \left(\frac{\partial \xi_{fi1}(\theta, s)}{\partial s}, \frac{\partial \xi_{fi2}(\theta, s)}{\partial s}, \dots, \frac{\partial \xi_{fid}(\theta, s)}{\partial s} \right) \quad (3)$$

根据式 (2), 三维飞机目标投影形状的矩不变量随观察角度 (球面上经纬度坐标) 变化的例子见图 1, 其中图 1(a) 为矩不变量 C3 的曲面, 图 1(b) 为 C3 关于 θ_1 的偏微分, 图 1(c)

为 $C3$ 关于 θ_2 的偏微分.

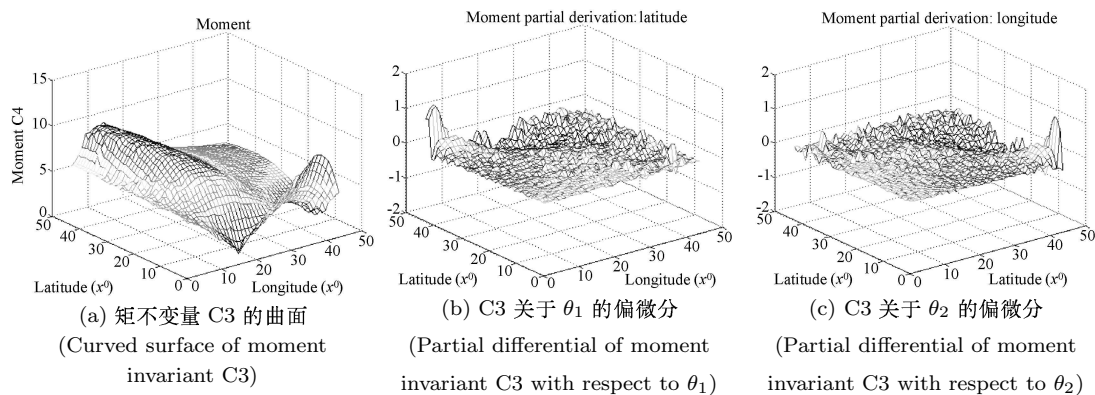


图 1 幻影 2000 飞机目标投影形状的矩不变量 $C3$ 随经纬度 (θ_1, θ_2) 方向的变化曲面图

Fig. 1 Curved surface of variation of the moment invariant $C3$ of the airplane object Mirage 2000's

如图 2(a) 所示, 动态特征空间增加了时间维 $t(\theta, s)$

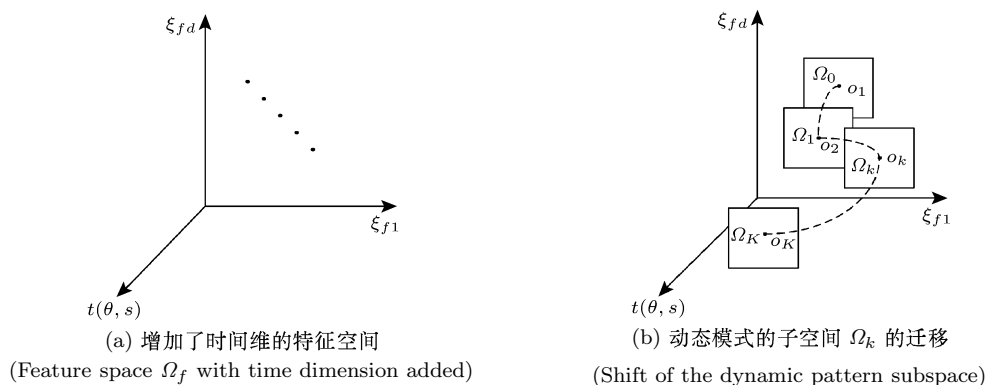


图 2 动态特征空间示意图

Fig. 2 Schematic diagram of the dynamic feature space Ω_f

定义 2. 动态模式

特征向量呈现随时间变化而系统性变化的模式, 称为动态模式.

例如, 飞行中的三维飞机目标, 其侧影图如图 3 所示.



图 3 幻影 2000 飞行的侧影图

Fig. 3 Profiles of Mirage 2000 in flight

如图 2(b) 所示, 我们有以下定义:

定义 3. 暂态子空间

在时刻 t_k 动态模式 ξ_f 在特征空间 Ω_f 中所构成的子空间 $\Omega_{f_k}, i = 0, 1, 2, \dots, K$.

定义 4. 子空间迁移

时间上相邻的一对子空间, 若暂态子空间 $\Omega_{f_k} \neq \Omega_{f_{k-1}}$, 称模式子空间发生了迁移. 因噪声而造成的两个子空间不相等, 是非系统性的, 故此处不算作子空间迁移.

定义 5. 动态模式的子空间轨迹

从时刻 t_0 到时刻 t_K , 动态模式 ξ 在特征空间 Ω_f 中所构成暂态子空间 Ω_{f_k} 的中心点 o_k , 所连接起来的有向曲线. 此处, o_k 为 Ω_{f_k} 中模式的聚类中心.

定义 6. 子空间序列上的可识别性度量

令 $\rho(\Omega_{f_j})$ 表示在某子空间 Ω_{f_j} 上的可识别性度量, 利用子空间集合 $(\Omega_{f_k}, 0 \leq k \leq K)$ 上的信息, 所能达到的关于目标识别分类性能的上界 $\rho_s((\Omega_{f_k}, 0 \leq k \leq K))$, 称为定义在该子空间序列上的可识别性度量. 这里 ρ_s 并非是对各个子空间可识别度量求最大值, 而是一个在子空间序列上能达到的理论最大值.

根据目标 / 背景特征的可变性和问题域的不同, 模式可以驻留在某个子空间, 或动态地从一个子空间迁移到另一个子空间. 从子空间 Ω_{f_i} 迁移到子空间 Ω_{f_k} 后, 模式的可识别性可能出现: 1) 模式的可识别性提高, $\rho(\Omega_{f_k}) > \rho(\Omega_{f_j})$; 2) 模式的可识别性劣化, $\rho(\Omega_{f_k}) < \rho(\Omega_{f_j})$; 3) 模式的可识别性无变化, $\rho(\Omega_{f_k}) = \rho(\Omega_{f_j})$.

鉴于模式的可识别性在变化, 如何从该迁移的子空间序列 $(\Omega_{f_0}, \Omega_{f_1}, \dots, \Omega_{f_k})$ 的角度研究提高模式可识性的策略和方法是非常重要的. 我们可以把子空间可识别性推广到在子空间序列上的可识别性.

3 三维运动目标的多尺度特征模型

基于高斯扩展图像映射而获得的三维目标的特性视图无疑是三维目标二维表示的有效模型. 但该模型的不完备之处是没有考虑实际成像条件下, 当成像传感器定焦, 焦平面敏感元数恒定, 目标运动所造成的观察距离变化等因素所引起的目标图像的模糊和尺度的变化. 常规的方位图是目标模型的一种理想化表示, 而实际应用必须考虑目标识别可能处在不同识别距离, 而不仅是方位的约束. 为此, 本文把常规特性视图表达推广到多尺度特性视图表达, 以反映实际成像条件下的约束. 如图 4 所示, 典型方位 i 指在高斯观察球面的特性视图所对应的观察方位, 尺度 j 对应于不同远近级别的观察距离, 尺度 0 为最近距离级或称分辨率相对最高、最清晰的级, 尺度 S_i 为分辨率相对最低、最模糊、距离相对最远的级.

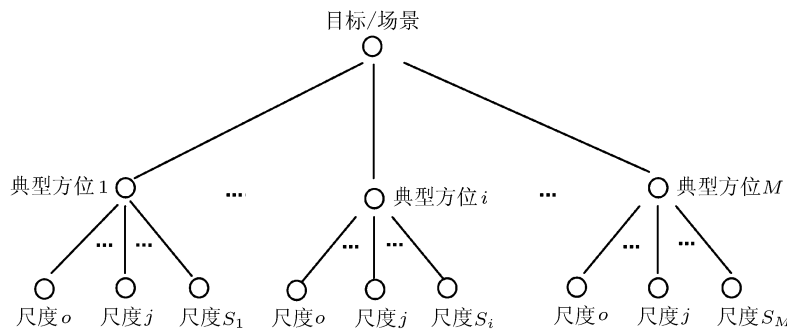


图 4 三维目标的多尺度、多方位表示

Fig. 4 Multi-scale and multi-aspect representation of 3-D objects

高斯观察球面也可以按经纬度的方式将其划分为若干特性视图区, 如图 5 所示.

当三维目标运动时,其姿态的变化所造成的效果等价于观察者视点在高斯球面上的移动.三维目标相对于观察者距离变化所造成的效果相当于目标图像尺度或模糊程度的变化.

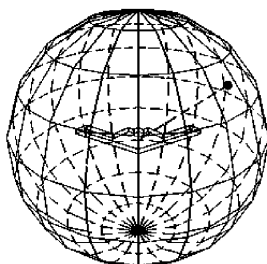


图 5 高斯观察球面的经纬度划分

Fig. 5 Division of longitude and latitude on the surface of the Gaussian viewing sphere

利用仿真软件 MultiGen 建立飞行器的三维模型,使模型尽量逼近目标真实三维数据.

当三维目标运动时,其姿态的变化所造成的效果等价于观察者视点在高斯球面上的移动.如图 5 所示,从不同的角度观测三维目标会得到不同的二维投影侧面视图,确定观测的方位角,仰角的范围和角度间隔结合建模仿真软件 Vega 就能够自动生成任意视点的目标侧视图.每隔 20° 采样得到 162 个视点的侧视图作为标准样本,用于生成目标的特性视图.视觉空间被划分为 162 个分区后,每个分区对应一幅二维侧视图,这 162 幅图像之间存在巨大的数据冗余,可以将标准样本(162 视点划分)中特征变化很小甚至不变化的分区进行聚类,重新构成一些分区,这些聚类后的分区提取出对应的视图即为三维目标的特性视图.

三维目标相对于观察者距离变化所造成效果相当于目标图像尺度或模糊程度的变化.目标侧视图的二维轮廓是闭合曲线,可以对其离散坐标序列进行等步长线性插值,并对插值后的轮廓坐标的 x 分量、 y 分量一维序列分别进行一维高斯核卷积模糊,然后降低采样率,再经过取整,生成该类目标轮廓的某级尺度的模型^[23].高斯核的大小决定了目标的尺度级别,高斯核越大,图像尺度越大,高斯核越小,图像尺度就越小.

将原始尺度级别(最小尺度)的 162 视点图像进行聚类得到目标特性视图,再将特性视图进行上述处理后,即能得到目标的多尺度特性视图.

图 6 为一种飞机目标多尺度特性视图表示例子.

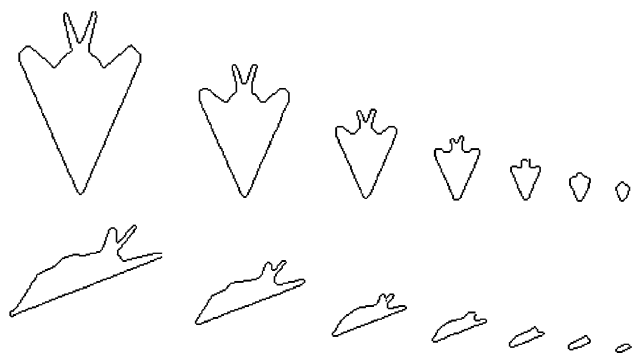


图 6 一种飞机目标在两个视点的多尺度特性视图表示例子

Fig. 6 Multi-scale characteristic view representation of an airplane at two viewpoints

由图 6 可见, 目标的二维形状随尺度和模糊以及方位而变化, 因此其矩不变量是随尺度和模糊而变化的^[17], 应建立多尺度模型用于实际成像条件的目标识别。

对于三维运动目标, 除一般的特征维之外, 还拥有三个既互相区别又互相联系的维: 方位维、尺度维和时间维的模式特征空间, 称为三维运动目标特征空间。

特性视图转移矩阵 T 是规定各特性视图互相转移难度的一个数量矩阵, 其横 - 纵坐标表示各特性视图的编号, 矩阵中某坐标处的取值越小, 表明“视图对”转移越容易。例如, 对角线坐标处全为“1”, 表示同一视图保持自我的代价最小。

由此, 我们可以推断在足够短的成像时间内 (如 50 帧 / 秒), 三维运动目标的图像序列及其对应的特性视图序列帧间发生突变的概率很小。这是我们利用序列图像及转移矩阵 T 识别三维目标可以利用的重要约束。

4 多尺度智能递推识别

通常, 目标并非能同时表现出其所有可能的特征, 而会以一个过程来逐步展示出来。特别对运动中的三维目标, 如果序列图像的信息自然可用, 就没有必要在单帧图像上追求高的识别率。这就是说更聪明的策略应该是通过多帧目标图像有效信息的递推积累来达到更高的识别率。这样, 不仅减少了目标特性视图模型的数量, 而且可简化分类器的训练, 保证更高的收敛性。

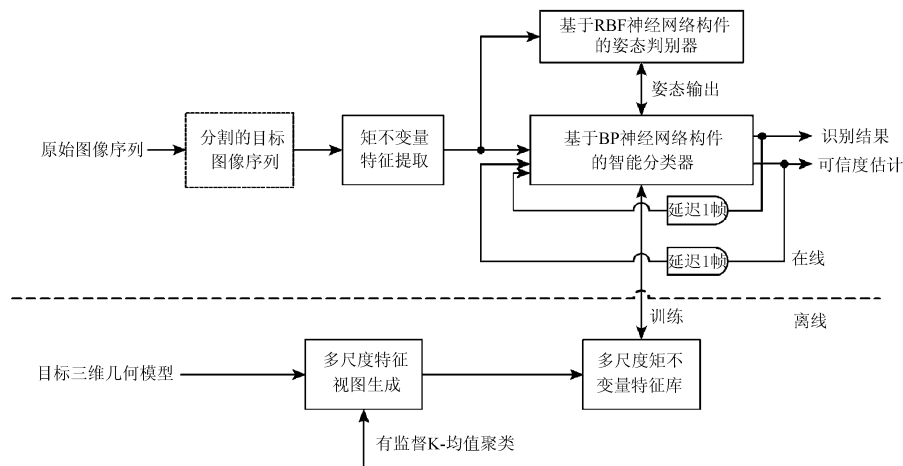
递推识别合理性在于, 由于目标模式是处在动态特征空间的迁移中, 其在单个子空间的可识别性可能不高, 但在多个子空间构成的子空间序列上, 其可识别性能得到极大的改善。

由于三维运动目标可能有无穷多个姿态, 基于其投影的识别, 对每一类目标本质就是一个多类识别问题, 对具有无穷多可能姿态的多类目标识别是非常困难的, 如何构造一个有效的识别器是本节的重点。

基于上述分析, 本文提出并实现了一种有效的多尺度智能递推识别算法 (MUSIRR)。构造的识别系统和算法如图 7 所示, 其中 (a) 为系统框图, 包括建模、训练、识别检验三个基本模块。本文不涉及具体的分割方法, 假定目标已从图像中分割出来, 但可能有缺陷、有噪声影响造成的畸变。

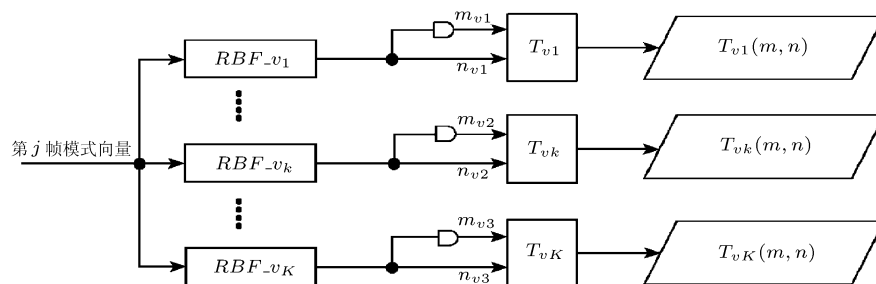
人工神经网络是一种具有学习能力的非线性分类器^[20,21]。本文将 BP 神经网络作为构建智能识别器的一种基本模块。BP 网的输入层节点数为模式样本的特征向量 (矩不变量) 的维数; 中间层为隐含层, 隐含层神经元的最优数目在文献 [24] 中有详细介绍; 输出层节点数为预期的目标类别数。经过多样本实验得出, 在识别输出类别数不大的情况下, BP 网的识别效果比 RBF 好; 而在训练样本不多且识别输出类别较多的情况下, BP 网的识别效果不如 RBF 网络好, 因此本文采用 RBF 网络识别三维目标姿态 (即特性视图所对应的视点)。RBF 是一个前向网络, 由一个输入层, 一个高斯节点组成的隐含层以及一个输出层组成。输入为目标模式的特征向量, 输出为特性视图编号。

图 7(b) 为以 RBF 网为构件组成的多目标姿态识别器。表 1 为利用 RBF 网络获得样本特性视图编号的识别正确率。图 7(c) 为以 BP 网和 RBF 网为构件组成的多尺度识别器, 其中 BP_0 为经 0 级尺度目标模型数据训练后所得 BP 网, BP_i 为经 i 级尺度目标模型数据训练后所得 BP 网。目标特征向量的检验数据输入该系统后, 各级 BP 网构件分别有一个输出类别的判断结果和可信度, 经智能判决模块 (该模块的具体实现在图 10 中已给出), 输出最终的识别判决和相应的综合可信度。以 RBF 为构件组成的三维目标姿态估计模块, 其输出将用于多尺度识别模块的重要约束。以下我们按层次的方式逐步给出算法完整的描述和解释。



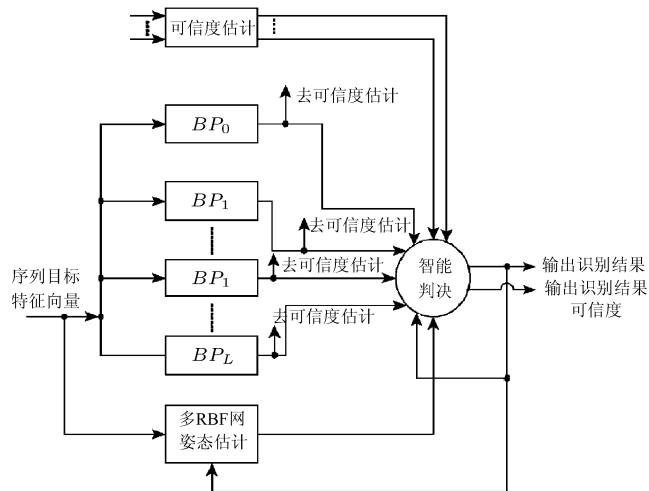
(a) 系统框图

(Block diagram of the recognizing system)



(b) 基于 RBF 网构件的多目标姿态识别器

(Multi-object gesture discriminator based on RBF network components)



(c) 基于神经网络构件的多尺度识别器

(Multi-scale recognizer based on BP neural network components)

图 7 识别系统示意图

Fig. 7 Schematic diagram of the recognizing system

表 1 RBF 网的姿态判别正确率统计 (%)

Table 1 Statistics of the rate of correct gesture judgment of the RBF networks (%)

图像分辨率	未畸变	畸变强度 1	畸变强度 2	畸变强度 3
256×256	94.0	89.3	74.6	63.5
224×224	91.8	86.8	72.1	61.2
192×192	92.8	83.7	70.7	54.4
160×160	90.9	82.9	68.4	51.1
128×128	91.8	81.3	66.5	46.4
112×112	87.9	80.2	64.1	43.5
96×96	88.5	74.1	61.2	42.5
80×80	80.1	71.6	57.7	37.7
64×64	80.7	66.0	53.8	34.7
56×56	71.1	61.4	48.4	33.4
48×48	72.7	56.3	41.6	27.0
40×40	60.6	50.2	37.1	26.8
32×32	62.2	44.9	30.5	24.4

4.1 基于多尺度模型的单帧识别

最基本的识别算法是基于单尺度模型训练的 BP 网来施行的. 设预期识别的目标类数为 K , 即目标类集合为 $\mathbf{f} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K)$. BP 网输出集合为 $O = (o_1, o_2, \dots, o_K)$, 这 K 个输出分别代表 K 类不同的目标. 对 BP 网训练时, 理想情况下, 对某个属于第 k 类目标的样本特征向量, 该 BP 网的第 k 个输出应为 1, 其他所有输出应为 0. 实际情况可能是网络的第 k 个输出为最大, 其他所有输出都应较小.

定义 7. 相对可信度

将输出结果 $(o_1, o_2, \dots, o_K)$ 按从大到小的顺序排列为 $(o'_1, o'_2, \dots, o'_K)$, $0 \leq o'_i \leq 1$, 单帧识别的相对可信度可合理地定义为

$$c = (o'_1 - \frac{1}{1 \times 2} o'_2 - \dots - \frac{1}{K(K-1)} o'_K) / (o'_1 + o'_2 + \dots + o'_K) \quad (4)$$

该定义是直观合理的:

- 1) 可信度的值域变化范围应该是 $[0, 1]$;
- 2) 当且仅当最大输出 $o'_1 = 1$ 且 $o'_2 = \dots = o'_K = 0$ 时, 可信度为最大值 1;
- 3) 当且仅当 $o'_1 = o'_2 = \dots = o'_K = 0$ 时, 可信度为最小值 0;
- 4) o'_1 越大可信度越大, o'_2 到 o'_K 越小可信度越大;
- 5) $\forall m \leq K$, 如果 $o'_1 = o'_2 = \dots = o'_m = 1$ 且 $o'_{m+1} = o'_{m+2} = \dots = o'_K = 0$, 就得到 m 个完全相等的最大值, 而只选取 o'_1 对应的类作为识别结果, 显然可信度应该只有 $1/m$.

若目标图像的尺度是未知的, 假定它可能的尺度范围是 $[s_0, s_{L-1}]$, 则上述基于单尺度目标模型训练的 BP 网就不能应付尺度变化的目标识别问题了. 必须建立由多个 BP 网构成的多尺度识别器, 各 BP 网构件分别由 $[s_0, s_{L-1}]$ 中相应尺度的目标模型训练得到, 由第 l 级目标模型训练所得构件表示为 BP_l . 识别时它们的输入相同, 并将其输出经过特别的融合操作, 得出判决结果 (融合判决过程的具体实现在图 10 中已给出), 如图 8 所示.

设输入特征向量 $\mathbf{f}_j \in (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K)$, 令尺度 l 的 BP 网构件 BP_l 各最大输出值为 $O_k^{(l)}$, $k = 1, 2, \dots, K$, $l = 0, 1, \dots, L-1$, l 为尺度级.

定义 8. 多尺度 BP 网的单帧识别度量可合理地规定为

$$MO_k = \frac{1}{L} \sum_{l=0}^{L-1} O_k^{(l)} \times c^{(l)}, \quad k = 1, 2, \dots, K, l = 0, 1, 2, \dots, L-1 \quad (5)$$

式中 $c^{(l)}$ 是 BP_l 输出的相对可信度作为加权, $O_k^{(l)}$ 是 BP_l 的最大输出值.

定义 9. 多尺度 BP 网单帧识别准则

若 $MO_{k_0} = \max\{MO_k, k = 1, 2, \dots, K\}$, 则称输入特征向量 f_j 属于第 k_0 类, $k_0 \in \{1, 2, \dots, K\}$.

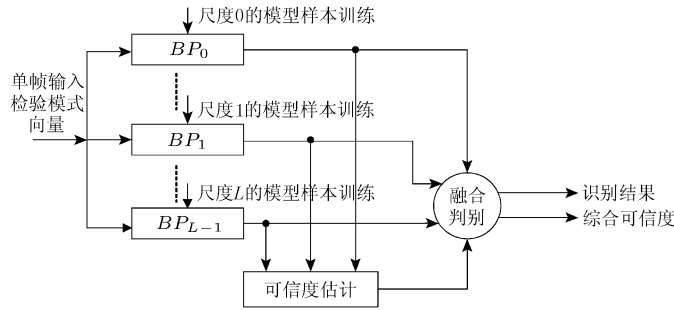


图 8 基于多尺度模型的单帧识别器框图

Fig. 8 Block diagram of single frame recognizer based on multi-scale models

4.2 基于多尺度模型的多帧序贯识别

相对于单帧图像, 三维目标运动产生的序列图像, 含有更丰富的关于目标的信息, 如何充分地利用这些信息提高识别率是非常有价值的, 将图 8 扩展的序贯多尺度识别器参见图 7(c).

设输入某类目标的第 j 帧图像特征向量 $f(j)$, 令 $MO_k(j)$ 为多尺度识别器输出端第 k 个节点对当前帧单帧的响应, 序贯多尺度识别器第 k 个输出端上一时刻的输出信息为 $MMO_k(j-1)$, 则识别器融合当前帧神经网络计算的信息与过去神经网络计算的信息, 可得出当前时刻第 k 个输出端的输出信息 $MMO_k(j)$, 递推公式如下

$$\begin{aligned} MMO_k(j) &= MMO_k(j-1) + w_k(j) \cdot MO_k(j) = \\ &= MMO_k(j-2) + w_k(j-1) \cdot MO_k(j-1) + w_k(j) \cdot MO_k(j) \quad (6) \\ MMO_k(0) &= 0, k = 1, 2, \dots, K, j = 0, 1, \dots, J-1 \end{aligned}$$

式中, $w(j-1)$ 为第 $(j-1)$ 帧权系数, $w(j)$ 为第 j 帧权系数. 由于 $w(j)$ 的值决定当前帧获得的新信息对当前作出判决所依赖的信息 $MMO_k(j)$ 的贡献的大小, 所以采取以相加的方式和前帧的信息进行融合是合理的. 关于该式收敛性的证明尚需进一步研究, 但实验结果表明, 在尺度不是很大的情况下, 利用该式能达到较高正确识别率. 以下是权系数合理取值的分析.

1) 由目标图像尺度信息获得权系数

目标的信息与图像的分辨率即尺度有关, 在不变焦条件下目标图像所占像素数越多, 即分辨率越高, 尺度越小, 信息越丰富. 设 j 第帧图像中目标像素数为 N_j , 该帧加权系数可合理地按式 (7) 计算

$$w^{(1)}(j) = e^{-\frac{\alpha}{N_j}} \quad (7)$$

式中, α 为一常数 (一般来说, 我们所获取的图像分辨率为 $256 \times 256 \sim 16 \times 16$, α 的取值应使分辨率 256×256 的图像目标权系数接近 1, 而分辨率为 16×16 的图像目标权系数接近 0), 图 9 为该权系数与 N_j 关系的图示.

直观地讲, 目标像素越多, 则可信度相对越高, 权重相对越大.

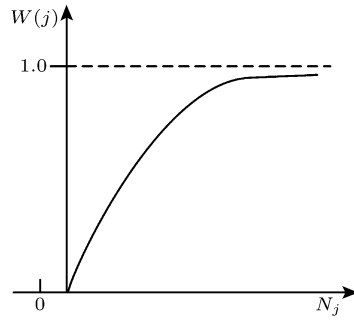


图 9 加权系数随目标像素数的变化关系

Fig. 9 The relationship in variation of weight coefficient with the number of object pixels

2) 由特性视图转移矩阵 T 获得权系数

由 RBF 网构造的姿态识别器估计各类目标的可能姿态. 根据运动目标动力学约束, 目标图像序列的相邻帧一般不会发生大的姿态迁移. 这就是说, 若识别器判断发生大的迁移则表明相邻帧信息的不可靠, 可考虑赋予小的权重. 故转移矩阵的相应元素可用于产生权系数, 即

$$w^{(2)}(j) = 1/[T_k(m, n)] \quad (8)$$

式中, T_k 表示第 k 类目标的特性视图转移矩阵, $T_k(m, n)$ 为 T_k 中的在坐标 (m, n) 处的元素, m 表示第 $(j-1)$ 帧单帧判决的特性视图的标号, n 表示第 j 帧单帧判决的特性视图的标号. $T_k(m, n)$ 越大, 表明相邻两帧发生了较大的姿态迁移, 即较大的三维目标姿态变化. 若第 j 帧图像识别为属于 k_0 类, 则对相邻两帧图像只可能发生以下 4 种情况:

- 情形 1: 第 $(j-1)$ 帧正确, 第 j 帧也正确;
- 情形 2: 第 $(j-1)$ 帧正确, 但第 j 帧错误;
- 情形 3: 第 $(j-1)$ 帧错误, 但第 j 帧正确;
- 情形 4: 第 $(j-1)$ 帧错误, 第 j 帧也错误.

推断出现何种情况的判据为 $T_k(m, n)$ 和当前帧判决可信度 $c_{k_0}(j)$, 以下是一种较合理的选择权值的方法.

1) 若 $T_{k_0}(m, n) < T_t$, 且 $c_{k_0}(j) > c_t$, 则认为第 j 帧单帧识别属于情形 1、情形 3 的可能性较大, 则

$$w_{k_0}^{(2)}(j) = 1/[T_{k_0}(m, n)] \quad (9)$$

其中, T_t 和 c_t 为一恰当的门槛 (根据对一定量训练、学习样本的识别结果确定这两个门槛值, 为经验值);

2) 否则, 情形 2、情形 4 发生的可能性大, 由于类别识别可能错误, 则姿态转移定为各类别中转移矩阵元素的最大值加 1, 即

$$w_{k_0}^{(2)}(j) = 1/\{\max[T_k(m, n)] + 1\}, \quad m, n, k = 1, 2, \dots, K \quad (10)$$

其含义是压低当前帧可能错误的信息的影响.

总结一下, 第 j 帧的综合权系数为两部分组成

$$CW_k(j) = w^{(1)}(j) \cdot w_k^{(2)}(j), \quad k = 1, 2, \dots, K \quad (11)$$

式 (6) 可重写为

$$MMO_k(j) = MMO_k(j-1) + CW_k(j) \cdot MO_k(j)$$

$$MMO_k(0) = 0, k = 1, 2, \dots, K, j = 0, 1, \dots, J-1 \quad (12)$$

定义 10. 多帧序贯识别判据为

若 $MMO_{k_0}(j) = \max\{MMO_k(j), k = 1, 2, \dots, K\}, j = 1, 2, \dots, J-1$, 则输入样本序列

$$\mathbf{f} = \{\mathbf{f}(1), \mathbf{f}(2), \dots, \mathbf{f}(j)\} \in \mathbf{f}_{k_0} \quad (13)$$

一个完整的针对多类飞机目标的混合神经网络和逻辑决策模块智能识别器 (Mixed neural networks and logic decision modules intelligent recognizer — MINNALODMIR) 框图如图 10 所示。

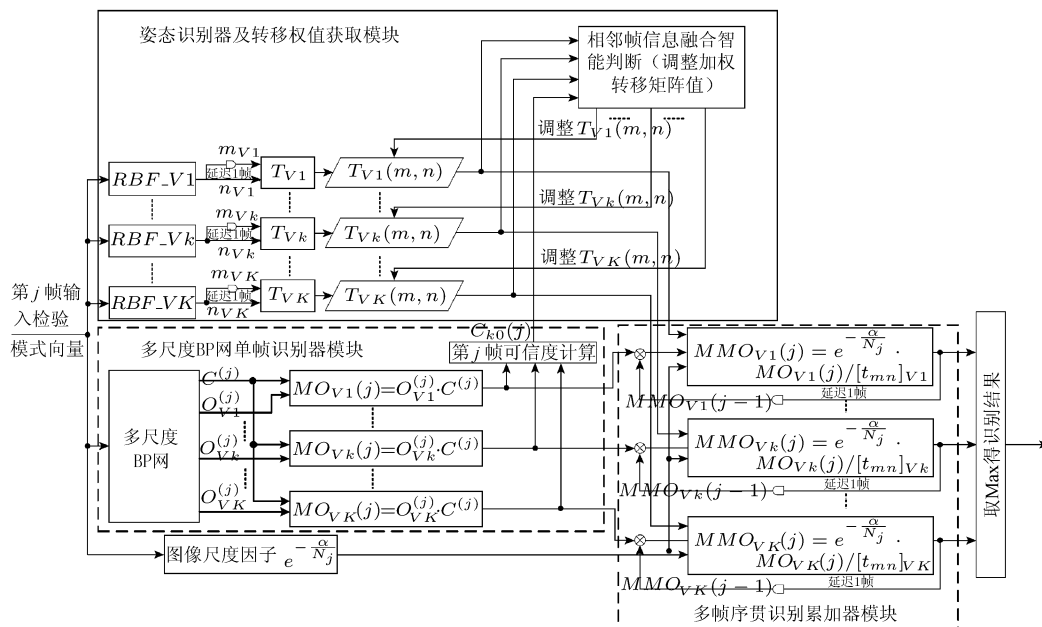


图 10 多类目标识别器框图

Fig. 10 Block diagram of a multi-category object recognizer

其中“识别累加模块”的输入需要“单帧识别器模块”和“权值获取模块”的输出, 由于“识别累加模块”对后两个模块结果的加权相加, 所以其计算速度相对于后两个模块来说是很快的, 整个系统的识别效率取决于后两个模块的计算速度. 在主频 1.8G 的 PC 机上 50 帧图像序列的识别时间为 0.861 秒, 对于 50 帧 / 秒的视频可以做到实时处理。

5 实验结果

5.1 样本的产生和训练

生成大量三维飞机目标模拟图像作为训练样本和测试样本. 针对三类目标: B2, F117, 幻影 2000, 由高斯观察球产生 162 个视点的目标侧影图, 经聚类合并形成少量的特性视图. 每一类目标包括 7 个尺度级别的二维特性视图集. 两个尺度级别的特性视图例子见附录, 每一个尺度分别为 38、39、44 个特性视图, 代表了不同视点、距离、模糊和尺度的影响, 即三类目标分别有 7×38 , 7×39 , 7×44 个特性视图, 由每一个特性视图可得 8 维不变量特征向量^[25~27]. 各尺度模型的特征向量作为训练样本, 训练相应尺度的 BP 神经网络和 RBF 网, 直至收敛。

检验样本产生：在 7 个尺度级范围内随机产生不同视点的样本图像 450 幅。可分别对其轮廓坐标迭加噪声，使轮廓产生畸变，模拟图像噪声和分割的缺陷所造成目标图像分割误差 (本文不讨论分割算法，假定目标已从图像中分割出来)。4 种畸变强度的例子见图 11 所示。

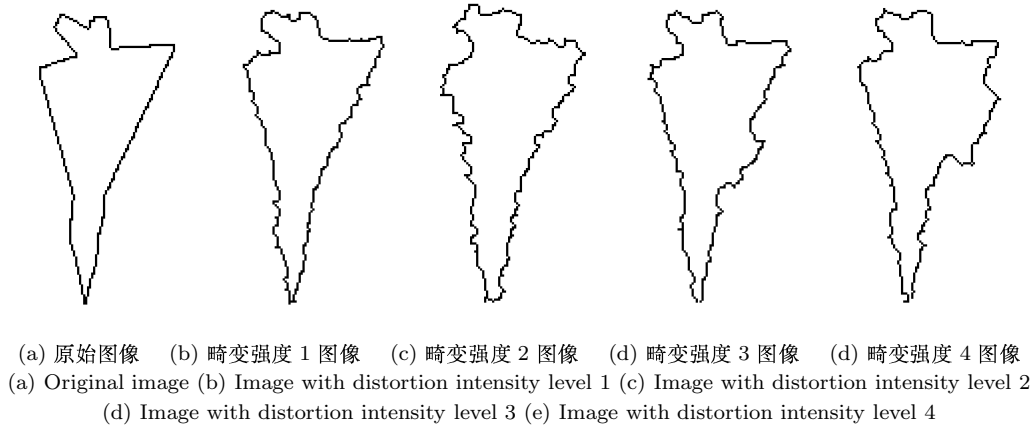


图 11 对目标图像施加 4 种不同畸变强度的例子

Fig. 11 Examples of superimposing 4 different intensity distortions on object contour images

每个 BP 网构件有 8 个输入节点，隐含层为 32 个节点，输出层为 3 个节点，代表三类目标输出。

5.2 单尺度模型训练的多帧序贯识别

图 12 为第 0 级尺度模型训练后的 BP 网识别率与帧数和轮廓畸变强度的关系。统计实验次数，其中 2~18 帧各有 10800 个实验序列，单帧识别实验样本有 450 个。

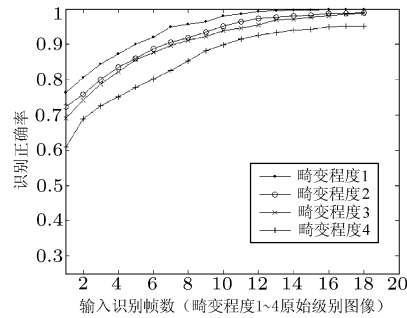


图 12 BP 网识别率与畸变强度、帧数的关系曲线

Fig. 12 Curves of relationship of BP network recognition rate with respect to distortion intensity and number of frames

5.3 多尺度模型训练的单帧识别

表 2 证明多尺度模型训练的多尺度 BP 网识别单帧样本的正确率，平均提高了 9.42 个百分点，特别是对 (模糊的) 小目标。这说明建立多级模型是非常必要的，其性能要优于单尺度模型训练的 BP 网。

表 2 单尺度及多尺度 BP 网对不同分辨率 (即不同尺度) 目标图像的识别结果对比 (识别率 %)

Table 2 Comparison between results of single frame recognition of object images of different resolutions (*i.e.* different scales) with single-scale multi-scale BP networks

BP 网尺度 / 分辨率	256 ²	224 ²	192 ²	160 ²	128 ²	112 ²	96 ²	80 ²	64 ²	56 ²	48 ²	40 ²	32 ²	平均
单尺度识别器	0	90.5	89.7	86.8	81.9	81.7	76.5	73.6	68.9	66.2	56.5	52.2	44.4	73.51
	1	85.6	88.2	88.3	87.4	83.7	82.4	82.1	79.8	77.1	73.0	63.5	50.8	76.71
	2	84.1	88.0	87.0	87.0	87.8	84.8	81.1	80.0	77.4	75.7	66.5	59.0	77.85
	3	79.6	81.0	81.5	83.1	83.5	83.1	83.9	79.8	76.5	73.2	60.3	48.4	74.65
	4	72.4	75.7	77.4	79.2	80.9	81.0	80.0	82.5	84.0	77.3	68.9	64.8	75.10
	5	65.2	66.3	69.1	68.7	68.5	69.1	66.5	72.6	72.4	73.8	77.4	64.8	68.21
	6	49.6	51.0	51.9	52.3	51.0	57.8	58.8	60.4	58.8	60.8	56.6	60.4	56.56
多尺度识别器	88.2	88.0	87.6	87.4	86.8	85.4	85.0	84.3	82.3	81.6	73.4	67.2	59.0	81.22

5.4 多尺度模型训练的多帧识别

使用 450 个视点的各种尺度及轮廓畸变的样本图像对本文的新算法进行检验 (分辨率级别分别为 256×256、224×224、192×192、160×160、128×128、112×112、96×96、80×80、64×64、56×56、48×48、40×40、32×32 共 13 个尺度级别, 其中标准级别 7 个, 中间级别 6 个; 两种畸变强度).

图 13, 图 14 给出多尺度智能递推识别算法的多帧序贯识别率.

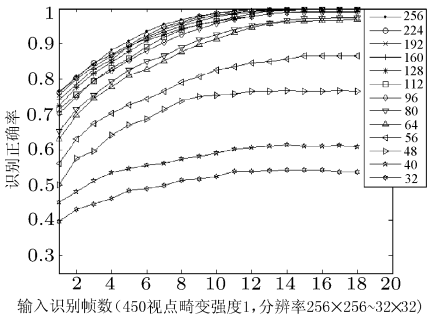


图 13 450 视点畸变强度 1 图像识别率曲线

Fig. 13 Curves of recognizing rate under distortion intensity 1

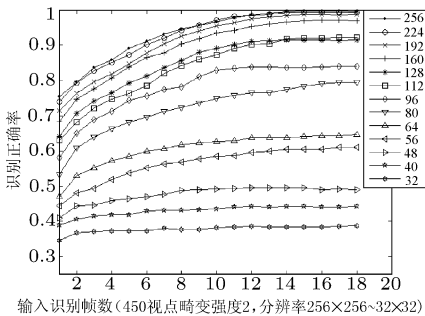


图 14 450 视点畸变强度 2 图像识别率曲线

Fig. 14 Curves of recognizing rate under distortion intensity 2

图 15 为检验图像尺度与姿态同时变化的多帧序贯图像的识别结果. 图 16 为多帧序贯识别可信度变化曲线的例子. 图 15 识别的图像序列中目标由小到大, 图 17 给出其中的一个实例.

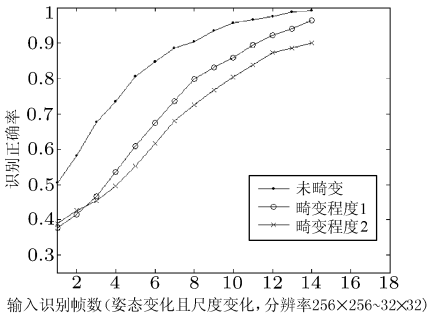


图 15 姿态方位与尺度均变化的多帧序列识别结果

Fig. 15 Curves of recognition rate of multi-frame sequential images with the gesture and scale varying at the same time

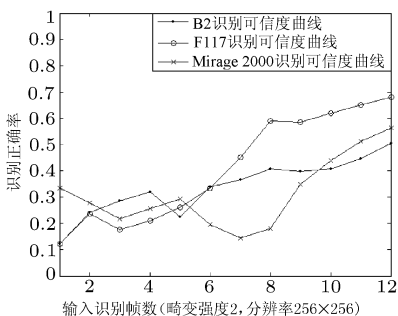


图 16 多帧序贯识别可信度变化曲线

Fig. 16 Distorted sequential object images with the scale and gesture varying at the same time

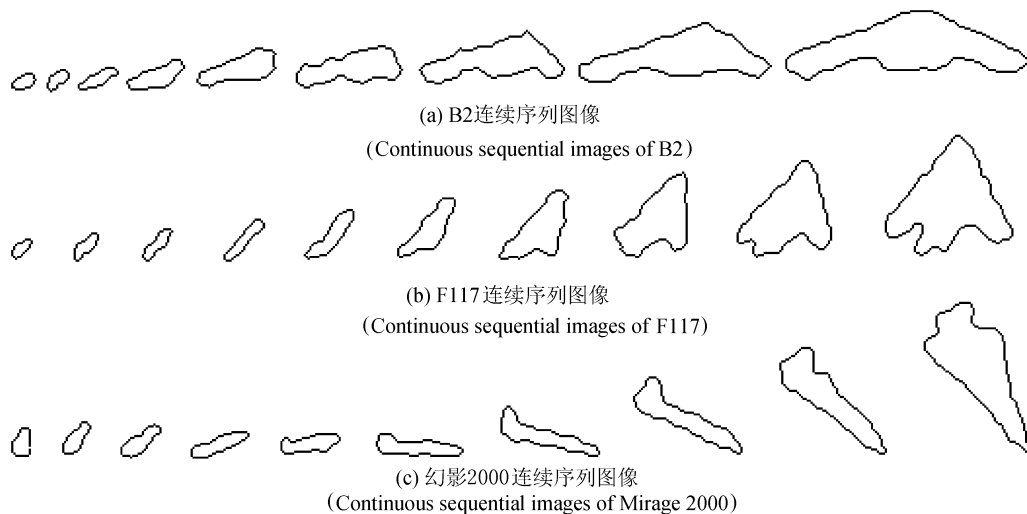


图 17 具有畸变的尺度和姿态同时变化的序列目标图像

Fig. 17 Curves of variation of confidence in multi-frame sequential recognition

模拟实验结果证明: 1) 多帧序贯识别正确率大大优于单帧识别; 2) 图 13、14 表明, 在大尺度子空间 (小目标) 及其子空间序列上的目标识别率一般低于在小尺度子空间 (大目标) 及其子空间序列上的识别率, 即目标可识别性随子空间和子空间序列的不同而变化; 3) 图 15、16 表明, 在由大尺度子空间 (小目标) 逐渐迁移到较小尺度子空间 (大目标) 所形成的子空间序列上, 目标识别率大大提高。

6 结论

目标特征的可观测性是动态变化的, 目标的模型和目标识别算法的研究很自然地要受到特征动态变化过程的约束, 因此目标可识别性的研究本质上应建立在动态特征空间之上。特征空间的概念可推广到包含时间维或隐含了时间因素的其他参数维 (如距离, 视点, 新现象新事件的介入等)。三维运动目标在实际成像和观测条件下的识别问题就是动态特征空间中目标识别的一个有说服力实例。

本文以三类飞机目标在姿态、运动、模糊和畸变强度变化条件下的序贯识别为例建立了多级二值特性视图模型及相应的多级矩不变量特征模型, 提出了智能递推识别框架和算法。该框架和算法能统一地处理从小的模糊目标到大的较清晰目标识别问题, 从单帧目标图像到目标图像序列。其性能优于常规单级目标表示及识别模型和方法。大量的模拟实验表明, 从动态特征空间观点出发来研究和发展相应的目标表达模型和新的识别算法是合理、有效的。

矩不变量特征容易受噪声和畸变的影响。进一步的工作可以在多级动态特征空间的框架内推广到使用其他统计特征和结构特征。另一方面序贯递推识别算法也需要进一步发展和完善以适应不同的复杂情况。

目标 (进一步可包括背景) 的多级特征建模, 可视为动态特征空间的离散化, 有关目标在动态的 (多级) 特征空间可识别性的理论研究应进一步深化。动态特征空间的合理性和新算法的有效性在于: 增加了特征空间时间维数 (隐含了尺度维、观察方向维、距离维), 就增加了识别子空间的可选择性, 从而改善了目标类别的可分性, 试验样本特征矢量的维数并没有增加。

有关文献曾经指出特征维数的增加, 不一定导致识别率和可分性的改善, 甚至可能劣化. 这是由于单纯地增加特征维数而没有动态地选择最优特征向量所致. 相反, 本文提出增加特征空间维数而不增加特征向量维数, 特征空间维数的增加, 导致增加子空间的可选择性. 即在某些子空间, 识别率可能优于其它子空间. 增加特征空间维数与增加特征矢量维数是不同的.

References

- 1 Koenderink J J, Doorn A J V. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 1979, **32**(4): 211~216
- 2 Marr D. Vision. San Francisco, CA: W. H. Freeman and Company, 1982
- 3 Das S, Bhanu B, Ho C. Generic object recognition using multiple representations. *Image Vision Computing*, 1996, **14**(5): 323~338
- 4 Horn B K P. Extended Gaussian images. In: Proceedings of the IEEE. USA: IEEE Press, 1984. **72**(12): 1671~1686
- 5 Ikenchi K, Flynn P J. Recent process in CAD-based vision. *Computer Vision and Image Understanding*, 1995, **61**(3): 114~132
- 6 Chen S, Freeman H. The dominant views of solid objects. In: Proceedings of 11th International Conference Pattern Recognition: (A), USA: IEEE Computer Society Press, 1992. **I**: 332~336
- 7 Munkelt O. Aspect-trees: Generation and interpretation. *Computer Vision and Image Understanding*, 1995, **61**(3): 365~386
- 8 Eggert D W, Bowyer Kevin W, Dyer Charles R, Henrik I. The scale space aspect graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(11): 1114~1129
- 9 Seibert M, Waxman A M. Adaptive 3-d object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(2): 107~124
- 10 Reeves A P, Taylor R W. Identification of three-dimensional objects using range information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, **11**(4): 403~410
- 11 Christopher M. Cyr, 3D object recognition using shape similarity-based aspect graph. In: Proceedings of the Eighth International Conference on Computer Vision, USA: IEEE Press, 2001. **1**: 254~261
- 12 Anshul Sehgal, Desai U B. 3D object recognition using Bayesian geometric hashing and pose clustering. *Pattern Recognition*, 2003, **36**(3): 765~780
- 13 Zhang J, Zhang X. Object representation and recognition in shape spaces. *Pattern Recognition*, 2003, **36**(5): 1143~1154
- 14 Mashor M Y, Osman M K. 3D object recognition using 2D moments and HMLP network. In: Proceedings International Conference on Computer Graphics, Imaging and Visualization. USA: IEEE Press, 2004. 126~130
- 15 Bhann B. Introduction to the special issue on automatic target recognition. *IEEE Transactions on Image Processing*, 1997, **6**(1): 1~6
- 16 Boshra M, Bhanu B. Predicting performance of object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(9): 956~969
- 17 Zhang T X, Liu J. An investigation on the instability of targets. *Journal of IR and Millimeter Wave*, 2004, **23**(3): 197~204
- 18 Zhang T X, Liu J. Blurred image recognition based on complex moment invariants. In: IEEE 2004 International Conference. Singapore: Image Processing (ICIP), 2004. 2131~2134
- 19 Fukunaga K. Introduction to Statistical Pattern Recognition. 2nd edition, London: Academic Press, 1990
- 20 Zurada J M. Introduction to Artificial Neural Systems. New York: West Publishing Company, 1992
- 21 Kaykin S. Neural Networks: A Comprehensive Foundation. New York: Prentice-Hall, 1999
- 22 Bhanu B, Pavlidis I, Hummel R. Special issue on computer vision beyond the visible spectrum. *Machine Vision Application*, 2000, **11**(6): 265~266
- 23 Peter J Burt. The Laplacian Pyramid as a compact image code. *IEEE Transactions on Communications*, 1983, **31**(4): 532~540
- 24 Kaykin S. Neural Networks: A Comprehensive Foundation. New York: Prentice-Hall, 1999
- 25 Hu M K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 1962, **8**(2): 179~187.

- 26 Schau H C. Shape recognition with scale and rotation invariance. *Optical Engineering*, 1992, **31**(2): 268~274
- 27 Liu J, Zhang T X. Fast algorithm for generation of moments invariants. *Pattern Recognition*, 2004, **37**(8): 1745~1756

张天序 中国科学技术大学本科毕业, 于华中工学院获计算机科学硕士学位, 于浙江大学获生物医学工程博士学位。1997 年、2002 年、2004 年法国波尔多第三大学客座教授, 1999 年香港中文大学客座教授。现任华中科技大学图像识别与人工智能研究所所长, 教授, 博士生导师, 图像信息处理与智能控制教育部重点实验室主任, 中国宇航学会光电技术专业委员会副主任委员, 《红外与激光工程》杂志副主编, 航天三院兼职研究员, 国家自然科学基金信息学部自动化学科评审组成员, 总装备部精确制导专业组成员, 湖北省有突出贡献的中青年专家。研究方向为图像分析、模式识别、计算机视觉、医学图像处理等。已在国内外重要学术刊物发表学术论文 100 余篇, 著有专著《成像自动目标识别》。获教育部科技进步一等奖、二等奖和三等奖各 1 项, 航天工业总公司科技进步二等奖和三等奖各 1 项, 国防科工委国防科学技术二等奖 1 项。

(ZHANG Tian-Xu Graduated from the University of Science and Technology of China in 1970. He received his M.S. degree in computer science from Huazhong University of Science and Technology in 1981 and his Ph.D. degree in biomedical engineering from Zhejiang University in 1989. He is currently a professor of the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology. His research interests include image analysis, pattern recognition, computer vision, and medical image processing. He has issued more than 100 papers in important learned periodicals and conferences and published a monograph *Automated Recognition of Imaged Targets*.)

翁文杰 2003 年于华中科技大学获得控制科学与工程学士学位, 现为华中科技大学图像所模式识别与智能系统硕士研究生。目前研究的主要内容包括模式识别、人工神经网络、图像处理、计算机视觉等。

(WENG Wen-Jie Currently a Master student in the Huazhong University of Science and Technology. He received his B.E. degree of control science and engineering from Huazhong University of Science and Technology in 2003. His research interests include pattern recognition, artificial neural networks, image processing and computer vision.)

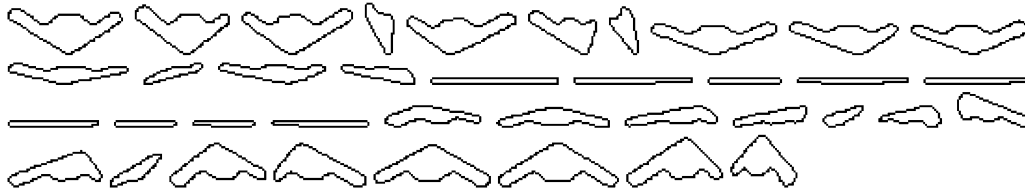
冯 军 2001 年于华中科技大学获得制冷与低温工程学士学位, 现为华中科技大学图像所模式识别与智能系统硕士研究生。目前研究的主要内容包括模式识别、计算机视觉、图像处理等。

(FENG Jun Currently a Master student in the Huazhong University of Science and Technology. He received his B.E. degree of freezing and low-temperature engineering from Huazhong University of Science and Technology in 2003. His research interests include pattern recognition, computer vision, and image processing.)

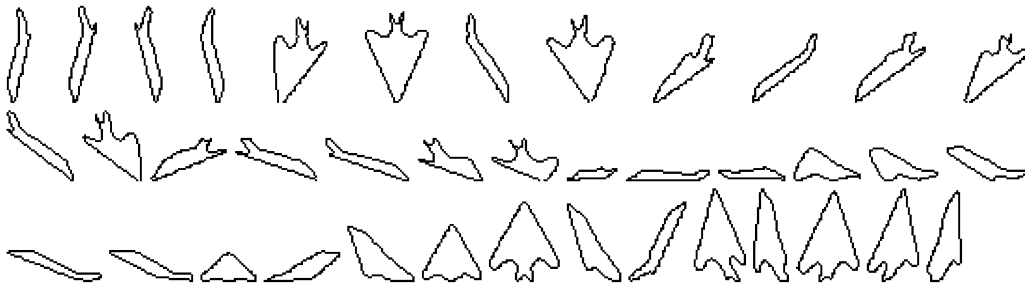
附录 A: 7 级特征模型库中 2 级特性视图模型例子 (对矩特征矢量聚类)

第 3 级尺度 3(96×96)

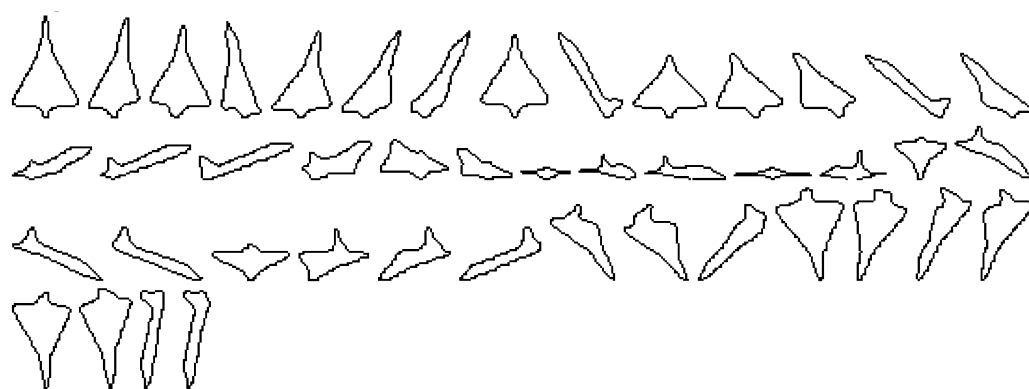
B2



F117



Mirage 2000



第 6 级尺度 $6(32 \times 32)$

B2



F117



Mirage 2000

