

基于双向非线性学习的轨迹跟踪和识别

胡昭华^{1),2)} 樊 鑫²⁾ 梁德群²⁾ 宋耀良¹⁾

¹⁾(南京理工大学电子工程与光电技术学院 南京 210094)

²⁾(大连海事大学信息工程学院 辽宁 大连 116026)

摘 要 目标的运动轨迹是跟踪和识别目标行为的重要特征之一,在视觉跟踪等领域得到了广泛的应用.然而,由于轨迹数据具有高维和非线性等特点,因而直接建模目标的运动轨迹比较困难.为此,引入一种称为自编码(autoencoder)的双向深层神经网络,并结合粒子滤波提出一种轨迹跟踪识别算法.首先,自编码网络按照一定的学习规则将高维轨迹嵌入到二维平面上,通过该网络的逆向映射得到轨迹的生成模型,由轨迹生成模型可得到一系列可行性轨迹.跟踪过程中,每时刻粒子滤波器的粒子便从这些可行性轨迹中进行抽样,并利用颜色似然函数对抽取的粒子进行加权以及再抽样从而实现对目标状态的估计,最后在二维平面中利用“最小距离分类器”对跟踪轨迹进行识别.特别地,自编码网络提供了高维轨迹空间和低维嵌套结构的双向映射,有效解决了大多数非线性降维方法(例如局部线性嵌入算法(LLE)和等度规映射(ISOMAP))所不具备的逆向映射问题.跟踪和识别手写数字实验表明所提出的方法能在复杂背景下精确跟踪目标并正确识别目标轨迹.

关键词 自编码网络;轨迹生成模型;非线性降维;目标跟踪

中图法分类号 TP391

Trajectory Tracking and Recognition Using Bi-Directional Nonlinear Learning

HU Zhao-Hua^{1),2)} FAN Xin²⁾ LIANG De-Qun²⁾ SONG Yao-Liang¹⁾

¹⁾(School of Electronic Engineering and Optoelectronic Technology, Nanjing University of Science and Technology, Nanjing 210094)

²⁾(School of Information Engineering, Dalian Maritime University, Dalian, Liaoning 116026)

Abstract Object trajectory is one of the most important cues for tracking and behavior recognition and can be widely applied to numerous such as visual surveillance and guidance. However, it is a difficult problem to directly model spatio-temporal variations of trajectories due to their high dimensionality and nonlinearity. This paper proposes a novel trajectory tracking and recognition algorithm by combining a bi-directional deep neural network called "autoencoder" into a particle filter. First, the "autoencoder" network embeds the high-dimensional trajectories in a two-dimensional plane based on a peculiar training rule and learns a trajectory generative model by the inverse mapping. Then a series of plausible trajectories are generated by the trajectory generative model. In the tracking process, the generated samples from the plausible trajectory set are weighted by the color likelihood and are resampled so as to obtain target state estimation at each time step. Finally the tracking trajectory is recognized by min-distance classification method in the two-dimensional plane. In particular, the "autoencoder" provides such a bi-directional mapping between the high-dimensional trajectory space and the low-dimensional space and is therefore able to overcome the inherited deficiency of most nonlinear dimensionality reduction methods (e. g. LLE and ISOMAP) that do not have an inverse mapping. The experiments on tracking and recog-

收稿日期:2007-02-06;修改稿收到日期:2007-05-24. 本课题得到国家“十五”科技攻关计划项目基金(2004BA111B01)资助. 胡昭华,女,1981年生,博士研究生,主要研究方向为视觉跟踪、模式识别、粒子滤波及机器学习等. E-mail: zhaohua_hu@163.com. 樊 鑫,男,1977年生,博士,讲师,主要研究方向为目标跟踪、人脸图像处理及机器学习等. 梁德群,男,1940年生,教授,博士生导师,主要研究领域为图像处理、通信系统中的信号处理. 宋耀良,男,1960年生,教授,博士生导师,主要研究方向为自适应信号处理、雷达信号处理、混沌通信等.

nizing handwritten digits show that the proposed algorithm can robustly track and exactly recognize in background clutter.

Keywords autoencoder network; trajectory generative model; nonlinear dimensionality reduction; object tracking

1 引 言

运动目标跟踪是视觉领域基本且重要的研究内容之一,近十年来一直成为计算机视觉领域的活跃研究课题.目标跟踪可视为是一种连续状态的估计问题,其中未知状态(隐性状态)是目标的位置或其它运动参数,目标跟踪的任务就是依据随时间演变的一系列观测图像来推断目标在每一时刻的未知状态.它的数学描述是:在获取所有观测图像序列的情况下,对目标状态的后验概率进行概率推断.应用领域包括视频监控、人机接口、虚拟现实以及数字视频编辑等多个方面.

由于目标的突然运动,同一场景的相似目标间存在目标模糊现象.目标受背景遮挡以及目标与目标之间的相互遮挡等多种因素的影响,使得目标跟踪问题成为较难解决的视觉问题.目前,大多数跟踪方法按照一种递归的方式进行:基于前一时刻目标状态的估计值和当前时刻目标的观测值,来估计当前时刻目标的新状态.在贝叶斯框架下,上述跟踪问题一般可以表述成:在获取直到 t 时刻所有观测值 $y_{1:t}$ 的情况下,递归估计状态 x_t 随时间演变后验概率 $p(x_t | y_{1:t})$.粒子滤波器是目前常用的一种跟踪算法,它的主要思想是用一系列加权粒子集来近似后验概率分布,在具有杂波和遮挡的复杂环境下显示出很好的性能^[1].

跟踪器的性能在很大程度上依赖于感兴趣目标的状态描述.目标的运动轨迹可视为目标状态随时间的演变序列,它是用于跟踪最重要的信息之一,一个恰当的目标轨迹模型则能大大提高跟踪器的性能.Sun 等人^[2]提出了一种基于轨迹分割分析(trajjectory segment analysis)的双向跟踪算法,很好地解决了相似目标间的模糊现象以及目标处于长时间遮挡的问题,但事先必须确定初始帧和最后一帧中的目标模板,属于离线跟踪问题,因而很难在线跟踪目标.现有的一些动态轨迹模型要么建立在人们的先验知识之上,其中模型参数通过学习获得^[3],要么完全从一大类数据集中学习得到^[4].然而,由于轨迹

具有高维和非线性的特性,因而直接建模轨迹的空时演变模型比较困难.此外,跟踪器的性能与状态的维数密切相关,当状态维数超过 10 时,粒子滤波则很难给出较好的跟踪效果^[5].

处理高维数据的方式之一是对之进行降维,通过学习以获得低维的隐变量模型.传统的降维方法如主成分分析(principal component analysis)、独立分量分析和因子分析(factor analysis)能在高维数据集具有线性结构和高斯分布时取得好的效果.当数据集在高维空间呈现高度扭曲时,这些方法则难以发现嵌入在数据集中的非线性结构以及恢复内在的结构.2000 年 Science 上发表了两篇被称为等度规映射(ISOMAP)^[6]和局部线性嵌入算法(LLE)^[7]的有关非线性降维的开创性论文.从那以来,流形学习方法得到了广泛的应用.流形学习通常假定数据集具有内在维数,从而可以通过将数据集约简至低维空间来避免维数灾难问题,并发现隐含在数据中的内在物理意义.然而,现有大多流形学习方法的主要缺陷在于它们只能学习已知数据集的潜在低维结构,并不能给出高维空间中数据点到低维空间的确定性映射,同样也无法给出相应的逆映射.因此这些非双向映射方法只适用于原始的训练数据集而不能有效运用于新出现的数据集.那么,采用这类降维方法的跟踪算法必须运用其他复杂的技术来改进这种缺陷.

本文引入自编码深层神经网络(Autoencoder network)方法^[8],该方法通过训练具有多个中间层的神经网络将高维轨迹转换成低维嵌套并继而重构高维轨迹.自编码网络给出了输入的高维数据与低维嵌套之间的双向映射,从而克服了大多数非线性降维方法所不具备的逆映射问题.本文将由自编码网络学习得到的轨迹生成模型与一种概率跟踪器——粒子滤波相结合,建立了一种稳健的跟踪和识别系统.具体的处理过程如下:首先由训练轨迹序列经自编码网络学习获得轨迹的低维嵌套结构,然后在低维空间中抽取粒子.这些抽样粒子经学习得到的逆映射返回到高维轨迹空间,从而建立轨迹生成模型,根据轨迹生成模型便可得到目标的一系列

可行性轨迹,最后利用粒子滤波跟踪算法对特定轨迹目标进行跟踪和识别.

本文首先介绍一种新的非线性降维方法——自编码深层神经网络.接着第 3 节对高维轨迹空间和低维嵌套之间的双向映射进行描述;第 4 节重点介绍轨迹生成模型;概率跟踪算法将在第 5 节中进行介绍;第 6 节对轨迹识别方法进行简要叙述;第 7 节给出手写数字轨迹跟踪和识别的实验结果;最后总结全文并简单介绍一下将来的工作.

2 双向非线性流形学习

我们希望有一种能给出高维数据空间和低维嵌套之间双向映射的非线性学习方法.因此自编码网络能很好地满足这些要求,它采用自适应、多层编码(encoder)网络将高维原始数据转换成低维嵌套,并且利用类似的解码(decoder)网络从低维嵌套中重构高维数据^[8].

2.1 自编码网络

已知轨迹集 $R=[r_1 r_2 \cdots r_n] \in \mathfrak{R}^{D \times n}$ 中共 n 条轨迹,其中 r_i 是第 i 条 D 维轨迹,我们希望发现一种低维嵌套 $M=[m_1 m_2 \cdots m_n] \in \mathfrak{R}^{d \times n}$ 满足 $d \ll D$. 利用自编码网络则能找到高维轨迹数据的低维嵌套结构.

自编码网络系统结构如图 1 所示,整个系统由编码和解码两个网络构成.编码网络属于降维部分,作用是将高维原始数据降到具有一定维数的低维嵌套结构上;解码网络属于重构部分,可视为编码网络的逆过程,作用是将低维嵌套上的点还原成高维数据.编码网络与解码网络之间还存在一个交叉部分,称之为“码字层”(code layer),它是整个自编码网络的核心,能够反映具有嵌套结构的高维数据集的内在规律,并确定高维数据集的内在维数.

自编码网络的工作原理如下:首先初始化编码和解码两个网络的权值,然后按照原始训练数据与重构数据之间误差最小化的原则对自编码网络进行训练.先经过解码网络然后再经过编码网络采用向后传播误差导数的链式法则很容易得到所需的梯度值,进而将自编码网络的权值调协到最佳值.

如果自编码网络的初始权值接近最优解,运用梯度下降法则能达到很好的训练结果. Hinton 和 Salakhutdinov^[8]使用了一种称为限制玻耳兹曼机(Restricted Boltzmann Machine, RBM)^[9]的两层网络来求取自编码网络的适当初始权值.然而用 RBM 来建模连续数据并不太理想.因而本文我们引入了

限制玻耳兹曼机的连续形式(CRBM)^[10-11],它是一种连续的随机再生模型,能够用一种简单、可靠的训练算法来建模连续数据.我们将运用 CRBM 建模连续数据的过程称为“预训练”过程. CRBM 的结构如图 1 右框图所示.

在预训练多层 CRBM 后,编码和解码网络都将使用经 CRBM 训练得到的权值作为自编码网络的初始权值.对于高维数据集而言,这是一种逐渐揭示数据集内在低维结构的有效方式.接下来的全局调整过程则使用反向传播算法通过整个自编码网络对权值进行调整从而达到数据集的最佳重构.

2.2 连续 CRBM

CRBM 的结构包括一个可视层和一个隐层以及它们层间的连接.图 1 的右边框图显示了 CRBM 的结构框架,图中原始轨迹数据对应 CRBM 的可视单元,因为它们的状态是可观测的. CRBM 的输出对应的是隐单元.可视单元与隐单元之间由权值矩阵 w 连接.设 v_i 和 h_j 分别描述可视单元 i 和隐单元 j 的状态,且它们之间的双向权值相等,即 $w_{ij} = w_{ji}$.

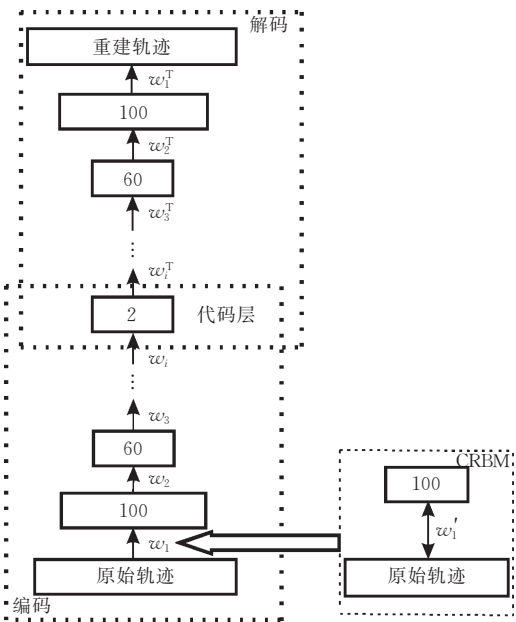


图 1 自编码网络系统结构图

为了建模连续数据,CRBM 通过在可视层添加一个零均值的高斯噪声从而引入一个连续随机单元.对于 CRBM 而言,隐层中每个单元的输入都来自于可视层中所有单元的状态,按照各个可视单元对每个隐单元的贡献大小,对它们之间的连接赋以相应的重要权值,即每个隐单元的状态值是所有可视单元的状态值按照对其贡献大小的加权和.为了表示方便,从现在起,以下都将采用同一个标识符 s

来表示可视单元和隐单元的状态, 设 s_j 表示输入来自于可视单元状态集 $\{s_i\}$ 的隐单元 j 的输出, 则有

$$s_j = \varphi_j \left(\sum_i w_{ij} s_i + \sigma \cdot N_j(0,1) \right) \tag{1}$$

式中函数 φ_j 的表达式如下

$$\varphi_j(x_j) = \theta_L + (\theta_H - \theta_L) \cdot \frac{1}{1 + \exp(-a_j x_j)} \tag{2}$$

其中 $N_j(0,1)$ 表示零均值、单位方差的高斯随机变量. 常数 σ 和 $N_j(0,1)$ 共同产生了一个噪声输入分量 $n_j = \sigma \cdot N_j(0,1)$, 其概率分布为

$$p(n_j) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-n_j^2}{2\sigma^2}\right) \tag{3}$$

由式(2)可知, $\varphi_j(x)$ 是渐近线在 θ_L 和 θ_H 处的 sigmoid 函数. 参数 a_j 控制着 sigmoid 曲线的斜率, 是噪声控制变量, 当 a_j 由小变大时, 可以完成从无噪声的确定性状态到二进制随机状态的平滑过渡.

CRBM 采用最小化对比散度 (Minimizing Contrastive Divergence, MCD) 训练准则替代了仅靠 Gibbs 抽样的玻耳兹曼机^[12] 的松弛搜索, 大大减少了计算量. MCD 训练准则用来更新 CRBM 的权值 $\{w_{ij}\}$ 以及“噪声控制”参数 $\{a_j\}$:

$$\Delta w_{ij} = \eta_w (\langle s_i s_j \rangle - \langle \hat{s}_i \hat{s}_j \rangle) \tag{4}$$

$$\Delta \hat{a}_j = \frac{\eta_a}{a_j^2} (\langle s_j^2 \rangle - \langle \hat{s}_j^2 \rangle) \tag{5}$$

其中 \hat{s}_j 表示单元 j 的一步重构状态, $\langle \cdot \rangle$ 表示训练数据的均值, η_w 是学习率.

式(4),(5)表明 CRBM 的训练准则只需进行简单的加法和乘法运算, 从而使得计算量不至于过大, 并且可以很容易完成权值的更新过程.

2.3 自编码网络的实现

为了表明自编码网络能够有效调整深层网络, 这一节我们利用“瑞士卷”数据集来训练一个深层自编码网络. “瑞士卷”数据集以及自编码网络结构如图 2 所示. 由图可见训练数据“瑞士卷”集是三维的, 但实际上它们却存在于一个嵌入三维空间的二维曲面上. 而三维数据集和二维曲面之间的关系是高度非线性的.

自编码由各层大小依次为 3-100-50-25-10-2 的编码网络和一个与之对称的解码网络构成. 为了提高网络的学习能力, 在输入层和中心隐层之间又加入了一些非线性中间层, 如图 2(b)所示. 自编码网络的训练数据是从“瑞士卷”上抽取的 12000 个三维点. 我们首先使用 CRBM 算法初始化网络的权值以至于这些初始权值能接近最优解, 然后再执行反

向传播算法以获得最优解. 自编码网络不仅将三维数据降到二维平面上, 而且还完成了二维点到三维数据的最佳重构. 图 3 分别显示了由 CRBM 对“瑞士卷”的重构图以及由自编码网络对“瑞士卷”的重构图. 显然, CRBM 的重构结果与真实数据之间存在一定的误差, 而自编码网络的重构效果几乎接近图 2(a)的真实数据.

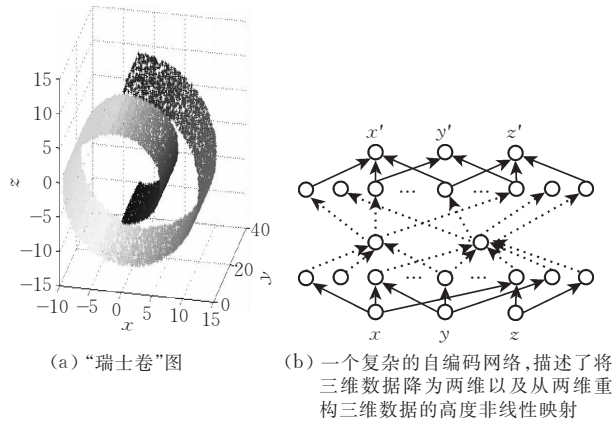


图 2

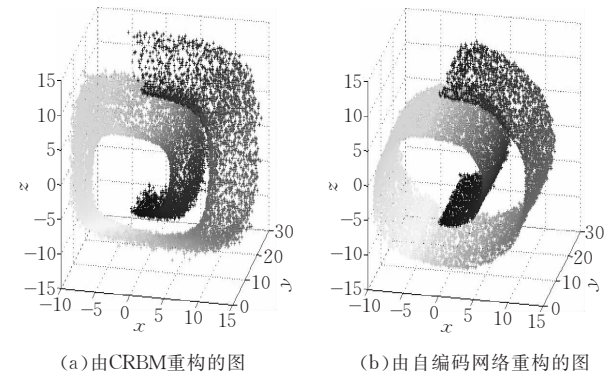


图 3 由 CRBM 和自编码网络重构的“瑞士卷”图

3 降维和轨迹重构

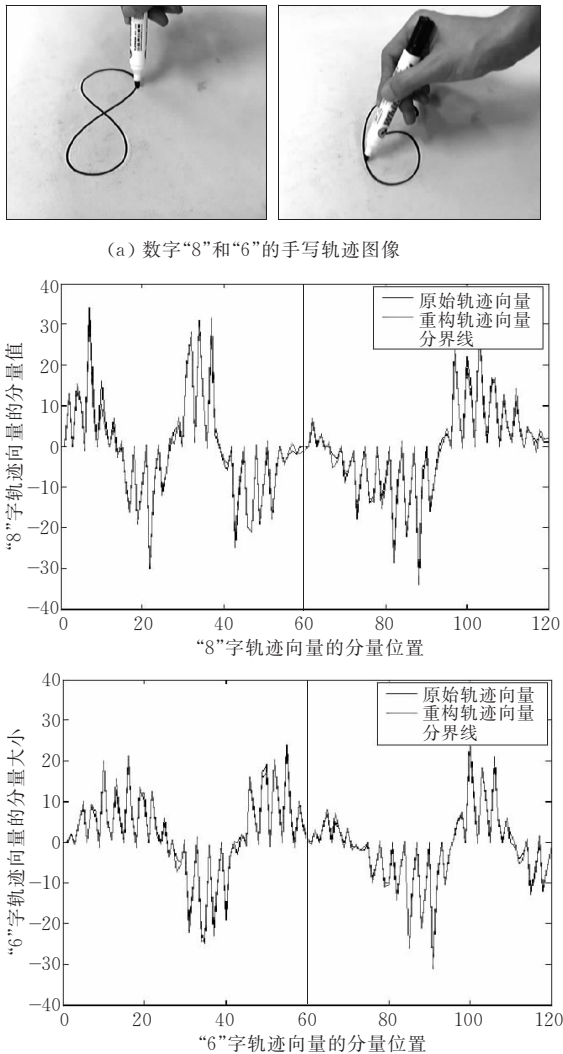
这一节我们采用真实图像序列中的数字手写轨迹作为训练数据集, 利用自编码网络对高维轨迹进行降维和重构.

在自编码网络框架下, 假设高维轨迹集 $R = [r_1 r_2 \cdots r_n]$, 其中 $r_i \in \mathbb{R}^D$ 落在 $d \ll D$ 维可能的非线性嵌套结构上, 我们的目标是希望找到这些高维数据集的低维嵌套结构: $M = [m_1 m_2 \cdots m_n]$, 其中 $m_i \in \mathbb{R}^d$. 因此该方法旨在发现轨迹数据集的内在规律以获得其紧凑的描述形式, 并且能从低维结构中有效地恢复原始轨迹数据.

为了获得比较精确的低维嵌套结构, 通常要求自

编码网络的训练数据应当足够多. 如果所有的轨迹样本都从真实视频中通过手动进行提取, 在实际应用中很难实施. 在应用中, 我们采取以下预处理方法: 首先从真实视频序列中手动提取少量(几十条)轨迹样本, 然后利用这些样本通过自编码网络得到相应的低维嵌套结构. 接下来便从该嵌套区域中任意抽取一定数量的点经反向映射后, 将得到的估计加上一定的随机噪声扰动从而生成一系列新的轨迹, 最后将这些新得到的轨迹作为自编码网络的训练样本.

该算法运用于手写数字“8”和手写数字“6”的轨迹空间. 图 4 显示了数字“8”和“6”的手写轨迹图像. 对于这个特殊的实例, 自编码网络发现了高维轨迹空间的二维嵌套结构.



(a) 数字“8”和“6”的手写轨迹图像

(b) “8”和“6”手写轨迹的原始训练数据和经自编码网络重构的数据(竖线是训练数据 x 坐标和 y 坐标的分隔线)

图 4 手写数字 8 和 6 的数据重构

实验中, 假设完成一次数字“8”或“6”的手写过程需要 60 帧图像的持续时间, 因此每条训练轨迹由

60 个位置点构成. 假设用 $[x_1, x_2, \dots, x_{60}, y_1, y_2, \dots, y_{60}]$ 描述一条轨迹, 其中 x_i, y_i 分别表示轨迹上第 i 个点的 x, y 坐标, 也即第 i 帧图像中轨迹的当前位置点. 接下来对每条训练轨迹 $[x_1, x_2, \dots, x_{60}, y_1, y_2, \dots, y_{60}]$ 进行归一化, 将轨迹上前一点位置坐标与当前点位置坐标的差值作为当前点归一化后的坐标值, 则归一化后的训练轨迹为 $[\Delta x_1, \Delta x_2, \dots, \Delta x_{60}, \Delta y_1, \Delta y_2, \dots, \Delta y_{60}]$.

设置自编码网络结构为 120-100-60-30-20-2, 并将该网络训练 1000 次, 其中的网络参数为 $\eta_w = 0.5$, $\eta_a = 0.9$, $\theta_H = 1$, $\theta_L = -1$ 和 $\sigma = 0.01$. 图 4 给出了数字“8”和“6”轨迹的原始训练数据以及经自编码网络重构后的数据. 由图可见, 自编码网络可在较小的误差范围内较好地恢复原数据.

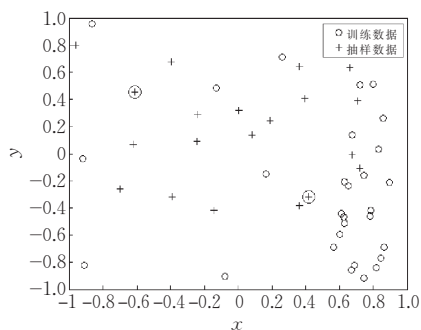
4 轨迹生成模型

由前所述, 从低维嵌套到高维轨迹空间的逆映射可由自编码网络通过训练数据学习得到, 一旦获得逆映射便可以很容易估计出每条由低维嵌套描述的高维运动轨迹. 自编码网络通过学习获得了低维嵌套与运动轨迹之间的双向映射, 则各种可能的轨迹便可由嵌套结构上的点经逆映射产生, 从而能够预测轨迹上的各点的状态.

我们仍然引用第 3 节中手写数字的例子, 在此仅用“6”字轨迹作为训练数据. 实验中采用结构为 120-100-60-30-20-2 的自编码网络对 30 条“6”字轨迹进行训练 1000 次. 训练后的结果, 高维轨迹全被映射到二维平面上, 且该二维平面的 x 坐标和 y 坐标范围都在 $[-1, 1]$ 之间, 也即高维训练轨迹全都映射于这个矩形中的二维点. 而且, 从该矩形平面内任意抽取一个点, 经自编码网络逆映射后得到的轨迹都是“6”字型轨迹. 因此, 各种大小、形状不同的“6”字轨迹都可由二维平面上的点经逆映射产生. 图 5 示出了经自编码网络训练后得到的二维嵌套平面以及由该平面上任意两点重构产生的“6”字轨迹.

在图 5(a) 所示的二维平面内, 圆圈表示训练数据经自编码网络降维后得到的二维点, “+”表示在二维平面中抽样得到的点. 任意选取其中两个抽样点(\oplus 表示的点)经逆映射得到“6”字型轨迹如图 5(b) 所示.

由此可见, 自编码网络能够有效描述具有嵌套结构的数据集的相关性和发现数据集的内在规律, 可以将数据集降维至低维空间并且从低维空间中重构高维数据集.



(a) “6”字轨迹的二维嵌套平面



(b) 由图(a)二维平面中两个黑色圆圈内的抽样点重构的“6”字型轨迹

图5 经自编码网络训练后得到的二维嵌套平面和由该平面上任意两点重构产生的“6”字轨迹

5 贝叶斯(Bayesian)跟踪

本节我们将由自编码网络得到的轨迹生成模型引入贝叶斯跟踪框架,实现目标轨迹的跟踪。

5.1 状态描述

我们将绘制数字的笔尖视作跟踪目标,用矩形框 $R = \{p, s \times w, s \times h\}$ 表示,其中 p 是矩形的中心, s 是尺度因子, w 和 h 分别是模板目标的宽度和高度。因此,目标的状态可表示成 $x = \{p, s\} \in \mathcal{X}$, 其中 \mathcal{X} 是状态空间。轨迹跟踪中,可设初始状态 x_1 是已知的。

5.2 观测模型

观测值是目标物体的颜色统计值。目标的颜色模型用直方图 $h^c = \{h_1^c h_2^c \cdots h_H^c\}$ 来表示,其中 $\sum_{i=1}^H h_i^c = 1$, H 是将颜色空间划分的段数,一般地取 $H = 10$, $c \in \{R, G, B\}$ 是颜色的通道类型。在此我们用 Bhattacharyya 距离^[13]来度量两个不同直方图之间的差异,则状态 x_0 的参考直方图 $h(x_0)$ 与状态 x_i 的候选直方图 $h(x_i)$ 之间的距离定义为

$$D(h(x_0), h(x_i)) = \left(1 - \sum_{j=1}^H \sqrt{h_j(x_0) h_j(x_i)}\right)^{1/2} \quad (6)$$

该模型只考虑了目标全局的颜色统计,如果目标具有一定的空间结构,则需要一个更为细致的多部分颜色模型^[1]进行描述。

基于上述直方图距离的定义,颜色似然函数可由以下公式表示

$$p(y_i^c | x_i, x_0) \propto \exp\left(-\sum_{c \in \{R, G, B\}} D^2(h^c(x_0), h^c(x_i))/2\sigma_c^2\right) \quad (7)$$

其中 σ_c^2 是颜色似然函数的方差。

已知初始状态 x_1 和一段视频序列观测值 $Y = \{y_1 y_2 \cdots y_T\}$, 在一阶 Markov 独立性假设条件下,整个状态序列 $X = \{x_1 x_2 \cdots x_T\}$ 的后验分布可以表述如下

$$p(X | Y) = \frac{1}{Z} \prod_{i=2}^T p(y_i | x_i, x_1) \prod_{i=1}^T p(x_i, x_{i+1}) \quad (8)$$

其中 $p(y_i | x_i, x_1)$ 由式(7)定义。两相邻状态之间的势函数 $p(x_i, x_{i+1})$ 定义为

$$p(x_i, x_{i+1}) \propto \exp(-D(x_i, x_{i+1})/2\sigma_p^2) \quad (9)$$

其中 $D(x_i, x_{i+1}) = \|p_i - p_{i+1}\|^2 + \beta \|s_i - s_{i+1}\|^2$ 是状态 x_i 和 x_{i+1} 之间的相似性度量, σ_p^2 是控制平滑程度的方差, β 是位置差和尺度差之间的权值。函数 $p(x_i, x_{i+1})$ 是对目标整体轨迹的一种平滑性限制。

5.3 粒子滤波跟踪算法

这一节将对概率跟踪算法进行简单描述。解决轨迹跟踪问题的数学描述就是需要求解式(8)的最大后验概率解(MAP)。为了有效估计每一时刻轨迹上的目标状态,我们提出了一种结合了轨迹生成模型的粒子滤波跟踪算法。

算法的基本思想如下:首先从二维嵌套平面中任意抽取 N 个样值点,然后经过自编码网络将这些抽样点逆映射到高维轨迹空间从而产生 N 条轨迹。每条生成轨迹中的第 t 个分量描述了 t 时刻目标的状态,也即 t 时刻所对应的粒子值。在粒子滤波算法的基础上,利用观测模型对每条轨迹中第 t 个分量(即 t 时刻)的粒子值进行加权。结果 t 时刻的后验概率密度 $p(X | Y)$ 可用元素个数为 N 的加权粒子集 $\{(x_t^{(i)}, \pi_t^{(i)}), i=1, 2, \dots, N\}$ 来近似,其中 $x_t^{(i)}$ 是离散随机样值,而 $\pi_t^{(i)}$ 是粒子 $x_t^{(i)}$ 的权值。

轨迹跟踪算法的流程总结如下。

1. 从训练得到的二维嵌套平面 $\{(x, y) | \|x\| \leq 1, \|y\| \leq 1\}$ 中均匀抽取 N 个粒子 $\{m_i\}_{i=1}^N$ 。
2. 利用自编码网络的逆映射将粒子集 $\{m_i\}_{i=1}^N$ 映射到高维轨迹空间得到 N 条可能的轨迹 $\{r_i\}_{i=1}^N$ 。
3. 用第一帧中目标的起始位置初始化每条抽样轨迹的初始状态 $\{(x_1^{(i)}, \pi_1^{(i)})\}_{i=1}^N$, 其中 $\pi_1^{(i)} = 1/N$, $x_1^{(i)} = (p_1^{(i)}, s_1^{(i)})$ 。
4. 假设 $t-1$ 时刻的粒子集为 $\{(x_{t-1}^{(i)}, \pi_{t-1}^{(i)})\}_{i=1}^N$, 则递归更新 t 时刻粒子集的过程如下:
 - 4.1. 从轨迹集中同时抽取每条轨迹上的第 t 个分量的值,即抽取 t 时刻目标在轨迹上的位置状态: $p_t^{(i)} \sim \{r_i\}_{i=1}^N$ 。
 - 4.2. 用式(7)描述的似然函数对抽取的粒子进行加权:

$\pi_t^{(i)} \propto p(y_t | p_t^{(i)}, s_{t-1}^{(i)})$, 其中 $\sum_{i=1}^N \pi_t^{(i)} = 1$.

4.3. 抽取 t 时刻目标尺度因子的值: $s_t^{(i)} \sim p(s_t | s_{t-1}^{(i)})$, 在此假设尺度因子的演变服从独立高斯随机游动模型, 即 $p(s_t | s_{t-1}^{(i)}) = N(s_t | s_{t-1}^{(i)}, \sigma_s)$, 其中 $N(\cdot | \mu, \sigma)$ 表示均值为 μ , 方差为 σ 的高斯分布, σ 是尺度因子随机游动的方差. 然后联合目标位置状态粒子 $p_t^{(i)}$ 和尺度因子粒子 $s_t^{(i)}$: $x_t^{(i)} \leftarrow (p_t^{(i)}, s_t^{(i)})$.

4.4. 权值更新: $\pi_t^{(i)} \leftarrow \pi_t^{(i)} p(p_t^{(i)}, s_t^{(i)} | p_{t-1}^{(i)}, s_{t-1}^{(i)})$. 同样 $p(x_t^{(i)} | x_{t-1}^{(i)})$ 也服从独立高斯随机游动模型, 其中 $\sum_{i=1}^N \pi_t^{(i)} = 1$.

如果需要再抽样的话, 则依概率 $\pi_t^{(k)}$ 对粒子集进行再抽样, 即 $a_i \sim \{\pi_t^{(k)}\}_{k=1}^N$, 从而得到 t 时刻的最终加权粒子集合 $\{x_t^{(i)}, \pi_t^{(i)}\} \leftarrow \{x_t^{(a_i)}, 1/N\}$.

6 识别算法

识别过程一般包括信息获取、预处理、特征提取和选择以及最终的分类器设计. 其中特征提取(选择)和分类器的设计可以说是比较重要的环节, 因为分类器的好坏和恰当与否直接关系到识别的精度, 而分类器一般是按照所提取的特征进行相应的设计.

手写数字轨迹的识别中, 我们提取的是数字轨迹的嵌套结构特征, 也即嵌套平面中的二维点. 这一步做起来比较容易, 因为前面我们已经利用自编码网络通过学习得到了高维数字轨迹的二维嵌套平面. 接下来我们只需利用最小距离分类器来确定某一特征点属于哪一类数字轨迹即可. 下面我们将简要介绍最小距离分类器.

假设有两类点集表示成 $C_1 = \{x_{1i}, i=1, 2, \dots, n_1\}$ 和 $C_2 = \{x_{2k}, k=1, 2, \dots, n_2\}$, 其中 n_1 和 n_2 分别是点集 C_1 和 C_2 的样本数目. 点 $p = (p_1, p_2, \dots, p_M)$

和点集 i 之间的距离定义为

$$d_i(p) = \sqrt{\sum_{j=1}^M (p_j - o_{ij})^2}, i = 1, 2 \quad (10)$$

其中 M 是点 p 的维数, $o_i = (o_{i1}, o_{i2}, \dots, o_{iM})$ 是点集 i 的中心, 则有

$$o_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik}, i = 1, 2; j = 1, 2, \dots, M \quad (11)$$

判决准则如下

$$d_i(p) < d_j(p), i, j = 1, 2 \Rightarrow p \in C_i \quad (12)^{\text{①}}$$

7 实验结果

这一节我们用数字手写视频序列来验证所提出跟踪算法的性能. 首先利用同一类手写数字轨迹单独作为训练数据来考察跟踪器的行为. 接着将两类手写数字混合输入自编码网络进行训练以显示跟踪和识别效果.

7.1 手写数字“6”的跟踪实验

实验中, 我们对数字“6”的手写轨迹进行训练以学习轨迹生成模型从而跟踪数字“6”的手写过程. 训练阶段, 采用 30 条“6”字轨迹作为训练数据输入自编码网络进行学习, 每条轨迹的维数是 120, 是从 60 帧连续视频序列中选取的. 相应的二维轨迹嵌套可通过结构为 120-100-60-30-20-2 的自编码网络训练得到.

由于一般粒子滤波器在预测当前时刻目标状态时仅利用到前一时刻目标的状态估计值, 因而常常是不稳定的. 跟踪器一旦丢失目标, 在此之后便很难再次锁定目标. 我们将轨迹生成模型结合进粒子滤波器, 因而所得的新的跟踪器可以有效预测目标全局的动态变化. 即便轨迹跟踪器偶尔丢失目标, 之后

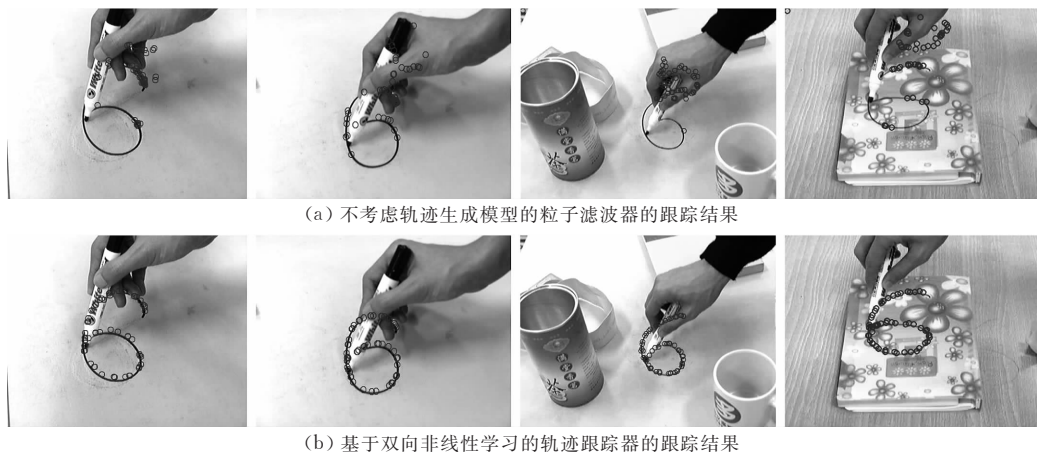
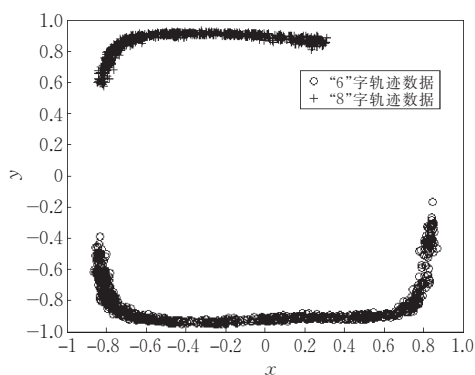


图 6 “6”字手写轨迹跟踪(两种跟踪器的实现场景既包括单纯白色背景也包括复杂背景)

① 此处 p 所代表的含义与第 5.3 节中 p 所指代的意义不一样.

也能很容易重新找回目标继而进行跟踪. 实验中为了将一般粒子滤波器和本文所提出算法的跟踪性能进行比较, 两种跟踪器抽取的粒子数都为 500, 并采用同样的观测模型.

图 6 示出了两种跟踪器对手写“6”字轨迹的跟踪结果. 一般粒子滤波器在跟踪开始几帧之后便丢失锁定, 而我们的跟踪器能够在手写过程中一直准确跟踪笔尖的位置. 实验结果表明新的跟踪器能够稳健跟踪手写数字轨迹并且不受手写数字大小和形



(a) 2000条“6”和“8”的手写轨迹经自编码网络训练得到的二维嵌套平面



(b) 轨迹跟踪器对数字“6”和“8”的跟踪结果

图 7 两种不同高维轨迹的降维结果

一般来说, 在线跟踪手写轨迹, 事先并不能确定手写的数字是“6”和“8”之中哪一个. 在这种情况下, 为了精确跟踪手写轨迹, 我们从二维平面中数字“6”的集中区域以及数字“8”的聚集区域同时抽取粒子. 接着将利用自编码网络的逆映射将抽样粒子返回高维空间得到许多可能的“6”字型 and “8”字型轨迹. 实验证明, 只要抽取的粒子数目足够多, “6”或“8”手写轨迹则能被精确跟踪. 图 7 则给出了数字“6”和“8”手写的跟踪结果.

实验中, 我们仅对“6”和“8”字这两种手写轨迹进行识别. 具体的轨迹识别过程如下: 将跟踪器估计得到的跟踪轨迹经自编码网络映射到二维嵌套平面从而得到一个二维点, 然后我们利用第 5.4 节中介绍的最小距离分类器来判断该点属于二维嵌套平面中哪个区域, 如果属于圆圈集中的区域, 则相应的轨迹属于数字“6”, 反之则属于数字“8”. 实验结果证明利用这种方法能够快速准确地识别出手写数字的类别. 虽然在此只对两种类型的手写数字进行识别, 该方法同样可以容易地推广到多类手写轨迹的识别.

8 结 论

本文, 我们介绍了一种用于视觉跟踪的轨迹生成模型的学习方法. 该框架是基于一种称为自编码

状以及背景杂波的影响.

7.2 两类不同手写数字的跟踪和识别实验

将“6”字手写轨迹和“8”字手写轨迹作为训练数据混合输入自编码网络以学习轨迹生成模型. 实验中采用 2000 个训练数据, 其中两种不同的轨迹数据各占一半. 图 7 描述了两两种不同高维轨迹的降维结果. 明显地, 训练数据集在二维嵌套平面上被分离成两部分, 其中每一部分分别对应数字“6”和数字“8”.

网络的双向非线性降维深层网络, 可以用来发掘高维轨迹所嵌入的低维嵌套. 自编码网络不仅能提供从高维轨迹空间到低维嵌套的映射, 而且也可以给出相反的逆映射. 基于这种优越的特性, 本文将此双向非线性学习方法结合进粒子滤波框架, 用来跟踪和识别手写数字轨迹. 实验结果显示, 在一般粒子滤波器跟踪失败的情况下, 这种新的跟踪器不仅能精确跟踪目标而且可以对手写数字轨迹进行有效识别.

将来的工作, 我们将建立一个更为有效的从低维嵌套空间的抽样方式和更为恰当的识别准则以改进跟踪性能和提高识别精度. 此外, 我们有意将双向非线性学习方法应用于更为复杂场景下的复杂运动(例如车辆的轨迹跟踪、人的外形跟踪和姿态识别等).

参 考 文 献

- [1] Perez P, Hue C, Vermaak J, Gangnet M. Color-based probabilistic tracking//Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark, 2002; 661-675
- [2] Sun J, Zhang W, Tang X, Shum H Y. Bi-directional tracking using trajectory segment analysis//Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China, 2005; 717-724
- [3] North B, Blake A, Isard M, Rittscher J. Learning and classification of complex dynamics. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(9): 1016-1034

- [4] Tay T, Sung K K. Probabilistic learning and modeling of object dynamics for tracking//Proceedings of the IEEE ICCV. Vancouver, Canada, 2001: 648-653
- [5] Forsyth D A, Ponce J. Computer Vision: A Modern Approach. New Jersey: Prentice Hall, 2002
- [6] Tenenbaum J B, Silva V de, Langford J C. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290: 2319-2323
- [7] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290: 2323-2326
- [8] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. Science, 2006, 313: 504-507
- [9] Hinton G E. Training products of experts by minimizing contrastive divergence. Neural Computation, 2000, 14(8): 1771-1800
- [10] Chen H, Murray A F. A continuous restricted Boltzmann machine with hardware-amenable learning algorithm//Proceedings of the 12th International Conference on Artificial Neural Networks (ICANN2002). Madrid, Spain, 2002: 358-363
- [11] Chen H, Murray A F. Continuous restricted Boltzmann machine with an implementable training algorithm. IEE Proceedings of Vision, Image and Signal Processing, 2003, 150(3):153-158
- [12] Hinton G E. Training products of experts by minimizing contrastive divergence. Gatsby Computational Neuroscience Unit, London; Technical Report GCNU TR 2000-004, 2000
- [13] Perez P, Vermaak J, Blake A. Data fusion for visual tracking with particles. Proceedings of the IEEE Publication Date, 2004, 92(3): 495-513



HU Zhao-Hua, born in 1981, Ph. D. candidate. Her research interests include visual tracking, pattern recognition, particle filter and machine learning.

FAN Xin, born in 1977, Ph. D., lecturer. His research

interests include object tracking, face image processing and machine learning.

LIANG De-Qun, born in 1940, professor, Ph. D. supervisor. His current research interests include image processing and signal processing in communication systems.

SONG Yao-Liang, born in 1960, professor, Ph. D. supervisor. His research interests include adaptive signal processing, radar signal processing and chaos-based communication.

Background

This work is supported by the Sub-Project of the National Key Technologies R&D Program of China on Image-based Intelligent Traffic Management under grant No. 2004BA111B01.

Object tracking and recognition is one of the key issues on intelligent video surveillance and management. Object trajectory provides an important cue for tracking and behavior recognition. This work addresses the fundamental issues to exploit the trajectory cue into tracking and recognition, i. e., how to represent the variations and uncertainties of trajectories and how to incorporate this model into the framework of tracking and recognition. The authors have developed the techniques to localize and recognize license plates from static images under the support of the grant. It is believed that the work provide a good starting point to immigrate the previous algorithms on static images to dynamic video sequence applications.

However, it is a difficult problem to directly model spatio-temporal variations of trajectories due to their high dimensionality and nonlinearity. To deal with large amounts of high-dimensional data, many researchers try to discover the latent structure of high-dimensional data. However, most existing methods, e. g., LLE and ISOMAP, merely find out a latent geometric structure of the data which preserves certain relationship between the data points of an available training set. They does not provide an explicit bi-directional mapping between the high-dimensional data space to the low-dimension-

al embedded space. To realize the mapping, these nonlinear dimensionality reduction methods have to resort to additional complicated techniques such as RBF (radial basis function) and GPLVM (Gaussian Process Latent Variable Models).

Considering above deficiencies, the authors introduce the "autoencoder" network which can convert high-dimensional data to low-dimensional codes by training a neural network with multiple hidden layers. The "autoencoder" provides such a bi-directional mapping between the high-dimensional trajectory space and low-dimensional latent space and is therefore able to overcome the inherited deficiency of most nonlinear dimensionality reduction methods. Moreover, the trained network implies a generative model for plausible trajectories which can be readily combined into a Bayesian framework for tracking and recognition. In this work, the authors develop the target tracking and recognition system by integrating the "autoencoder" and the particle filter. First, the "autoencoder" network embeds the high-dimensional trajectories in a two-dimensional plane based on a peculiar training rule and learns a trajectory generative model by the inverse mapping. Then a series of plausible trajectories are generated by the trajectory generative model. In the tracking process, the generated samples from the plausible trajectory set are weighted by the color likelihood and are resampled so as to obtain target state estimation at each time step. Finally the tracking trajectory is recognized by min-distance classification method in the two-dimensional plane.