

基于参数调整的动态模糊神经网络算法

张德丰¹, 周 灵¹, 孙亚民², 马子龙³

(1. 佛山科学技术学院计算机系, 佛山 528000; 2. 南京理工大学计算机科学与技术学院, 南京 210094;

3. 哈尔滨工业大学电子工程系, 哈尔滨 150001)

摘要: 模糊逻辑与神经网络结合形成的模糊神经网络同时具有易于表达人类知识、存储与学习分布信息的优点, 基于此, 提出一种基于参数调整的动态模糊神经网络算法。采用扩展卡尔曼滤波器法将全局算法划分为线性和非线性部分, 线性参数由最小二乘法和滤波器法决定, 非线性参数由训练样本和启发式法直接决定, 线性和非线性参数可进行实时更新。仿真结果表明, 该算法能保证更简洁的结构和更短的学习时间。

关键词: 动态模糊神经网络; 模糊逻辑; 扩展卡尔曼滤波器

Dynamic Fuzzy Neural Network Algorithm Based on Parameter Adjusting

ZHANG De-feng¹, ZHOU Ling¹, SUN Ya-min², MA Zi-long³

(1. Department of Computer, Foshan University, Foshan 528000;

2. School of Computer and Technology, Nanjing University of Science and Technology, Nanjing 210094;

3. Department of Electrical Engineering, Harbin Institute of Technology, Harbin 150001)

【Abstract】 The union of fuzzy logic and the neural network forms the fuzzy neural network, it simultaneously has the advantages of expressing the human knowledge, storing and learning distribution information storage. This paper proposes a Dynamic Fuzzy-Neural Network(D-FNN) algorithm based on parameter adjusting. It uses Extended Kalman Filter(EKF) method to divide the overall algorithm into the linearity and the misalignment part. The linear parameter is decided by Least Squares(LS) method and the filter method, the misalignment parameter is directly decided by the training sample and heuristic method, the linearity and the misalignment parameter can carry on the real-time renewal. Simulation results indicate that this algorithm can guarantee more succinct structure and shorter learning time.

【Key words】 Dynamic Fuzzy-Neural Network(D-FNN); fuzzy logic; Extended Kalman Filter(EKF)

近年来, 由于对模糊系统和神经网络的结合的大量研究, 因此产生了模糊神经网络这个新的研究领域, 该网络致力于获得 2 种系统的优点、克服各自缺点。模糊神经网络在工程技术各领域已得到广泛应用。

1 基于参数调整的动态模糊神经网络算法

如果要辨识模糊规则数, 则不能预先选择动态模糊神经网络(Dynamic Fuzzy-Neural Network, D-FNN)的结构。因此, 必须对 D-FNN 提出一种新的学习算法, 该算法能自动地确定模糊规则并能达到系统的特定性能。

1.1 模糊规则产生

如果规则数太少, 则系统不能完全包含输入-输出状态空间, D-FNN 的性能将表现很差。如果规则数太多, 则不仅增加系统不必要的复杂性, 且将极大增加计算负担并导致 D-FNN 泛化能力变差。因此, 输出误差是确定新规则是否应该加入的重要因素。

误差判断描述如下: 对于第 i 个观测数据 (X_i, t_i) , 其中, X_i 是输入向量; t_i 是期望输出; 计算 D-FNN 现有结构的全部输出为 y_i 。

定义

$$\|e_i\| = \|t_i - y_i\| \quad (1)$$

如果

$$\|e_i\| > k_e \quad (2)$$

则考虑增加一条新规则。这里的 k_e 值是根据 D-FNN 期望精度预先选定的^[1]。

从高斯函数角度来看, 一个高斯函数具有良好的局部特性, 其输出随着与中心距离的增加呈单调递减。当输入变量的隶属函数采用高斯函数描述时, 即整个输入空间由一系列高斯函数划分。如果一个新的样本位于某个存在的高斯函数覆盖范围或称为可容纳边界内, 则该新样本可以用存在的高斯函数代表, 而无须 D-FNN 产生新的高斯单元。

可容纳边界判据的描述如下: 对于第 i 个观测数据 (X_i, t_i) , 计算输入值 X_i 和现有径向基函数(Radial Basis Function, RBF)单元的中心 C_j 之间的距离 $d_i(j)$, 即

$$d_i(j) = \|X_i - C_j\| \quad j = 1, 2, \dots, u \quad (3)$$

其中, u 是现有模糊规则或 RBF 单元数量。

找出

基金项目: 广东省自然科学基金资助项目(9151040701000002)

作者简介: 张德丰(1963 -), 男, 副教授, 主研方向: 模式识别, 智能系统, 小波分析及其应用; 周 灵, 副教授、博士; 孙亚民, 教授、博士生导师; 马子龙, 教授、博士生导师

收稿日期: 2009-12-21 **E-mail:** defengz@163.com

$$d_{\min} = \arg \min(d_i(j)) \quad (4)$$

如果

$$d_{\min} > k_d \quad (5)$$

则考虑增加一条新的模糊规则,否则,观测数据 X_i 可以由现有最近的 RBF 单元表示。其中 k_d 是可容纳边界的有效半径。

1.2 分级学习

分级学习的主要思想为:每个 RBF 单元的可容纳边界不是固定的,而是动态调节的,开始可容纳的边界设置较大,以实现全局学习,随着不断学习,边界逐渐减小,开始局部学习。根据该思想,本文提出一种简单的方法,即基于单调递减函数,逐渐减少每个 RBF 单元的有效半径和误差指数。即式(2)、式(5)中的 k_e 、 k_d 不是常数,由下式确定:

$$k_e = \max[e_{\max} \times \beta^i, e_{\min}] \quad (6)$$

$$k_d = \max[d_{\max} \times \gamma^i, d_{\min}] \quad (7)$$

其中, e_{\max} 是预先定义好的最大误差; e_{\min} 是期望的 D-FNN 精度; $\beta(0 < \beta < 1)$ 是收敛常数; d_{\max} 是输入空间的最大长度; d_{\min} 是本文关心的最小长度; $\gamma(0 < \gamma < 1)$ 是衰减常数^[2]。

分级学习的关键是确定产生输出误差大而没有被现有模糊规则覆盖的位置,这个阶段称为粗学习,当 k_e 、 k_d 分别达到 e_{\min} 和 d_{\min} 时,这个阶段称为细学习。该策略可缓解学习过程中的振荡问题,即模糊规则在开始时迅速增长,然后一些规则在学习过程中被剔除,因此,这种策略可以增强 D-FNN 鲁棒性并减少学习时间。如果把没有分级学习的 D-FNN 用于复杂系统,则很容易产生振荡并可能导致无止境的学习。

1.3 前提参数分配

当一条模糊规则产生后,要解决的首要问题是如何确定它的参数。仿真结果显示,RBF 单元的宽度对系统的泛化性能很重要。如果这个宽度小于邻近输入的距离(即 RBF 单元没有重叠),则模糊神经网络的泛化能力就会很差,D-FNN 将不能对未知的输入给出有用输出。但如果宽度太大,则 RBF 单元容易饱和,即无论输入是多少,它的输出都将会很大(接近 1.0)。因此,新产生规则的初始参数按照如下方式分配:

$$C_i = X_i \quad (8)$$

$$\sigma_i = k \times d_{\min} \quad (9)$$

其中, $k(k > 1)$ 是重叠因子^[3]。

当第 1 个观测数据 (X_1, t_1) 得到后,此时的 D-FNN 还没有建立,因此,这个数据将被选为第 1 条模糊规则: $C_1 = X_1$, $\sigma_1 = \sigma_0$, 其中, σ_0 是预先设定的常数。只有当 $\|e_i\| > k_e$, $d_{\min} > k_d$ 时,才要增加一条模糊规则,不满足该条件的其他 3 种情况的算法如下:

(1)第 1 种情况: $\|e_i\| < k_e$, $d_{\min} < k_d$

这种情况表明 D-FNN 可以完全容纳观测数据 (X_i, t_i) ,不需要再做任何处理或者仅须更新结果参数。

(2)第 2 种情况: $\|e_i\| < k_e$, $d_{\min} > k_d$

这种情况表明所建立的 D-FNN 具有较好的泛化能力,只有结果参数需要调整。

(3)第 3 种情况: $\|e_i\| > k_e$, $d_{\min} < k_d$

这种情况表示尽管 X_i 可以聚集在已有 RBF 单元附近,但 RBF 单元不是很重要,即覆盖 X_i 的 RBF 单元的泛化能力并不是很好。因此,该 RBF 节点以及结果参数会同被更新。

对于最接近 X_i 的第 k 个 RBF 单元按下式调整。

$$\sigma_k^i = k_w \times \sigma_k^{i-1} \quad (10)$$

其中, $k_w(k_w > 1)$ 是预定的常数。

1.4 D-FNN 参数调整方法

扩展卡尔曼滤波器(Extended Kalman Filter, EKF)是基于梯度的在线学习算法,该算法可用于平滑、滤波或者预测非线性动态系统的状态,同其他基于梯度的在线算法相比,EKF 可以加快收敛速度。

作为一种非线性更新算法,EKF 方法可以调节 D-FNN 的所有参数。由于全局方法将涉及大矩阵运算,将遇到巨大计算负担以及占用大量内存,因此这种全局方法可以被划分为一系列可处理的子问题。在实际应用中,可以用卡尔曼滤波(Kalman Filter, KF)方法来调节结果参数,同时 EKF 方法用于更新前提参数的中心和宽度,从而使所有参数都被修正。该思想等价于把全局算法分解为一系列解耦的算法。

实验结果表明,高斯中心对系统性能影响不大,因此,只须对高斯宽度进行更新。结果参数的更新呈线性,高斯宽度的更新呈非线性,可以用 EKF 做如下优化:

$$\begin{cases} K_i^\sigma = \frac{S_{i-1} F_i}{1 + F_i^T S_{i-1} F_i} \\ S_i = S_{i-1} - K_i^\sigma F_i^T S_{i-1} \\ \Sigma_i = \Sigma_{i-1} + K_i^\sigma (T_i^T - \Psi_i^T W_{i-1}) \end{cases} \quad i = 1, 2, \dots, n \quad (11)$$

初始条件为 $\Sigma_0 > 0$ 和 $S_0 = \rho I$, 其中, ρ 为一个小的正数; K_i^σ 是 i 第次观测的增益矩阵; S_i 是第 i 次观测误差协方差矩阵; $\Sigma_i = (\sigma_1, \sigma_2, \dots, \sigma_u)$ 表示经过 i 次迭代后的高斯宽度向量, $F_i = (\delta\sigma_1, \delta\sigma_2, \dots, \delta\sigma_j, \dots, \delta\sigma_u)$ 是第 i 次观测宽度的梯度向量^[4]。

$$F_i = \frac{\partial Y_i^T}{\partial \Sigma_{i-1}} \Big|_{\Sigma = \Sigma(i-1)} = -2 \Sigma_{i-1} \left\| X_i - C_j \right\| \psi_j \sum_{k=1}^u (w_j - w_k) \psi_k, \quad j \in \{1, 2, \dots, u\} \quad (12)$$

其中, C_j 和 w_j 分别是中心和第 j 个 RBF 单元的权值, ψ_j 是第 j 个归一化层的输出。

KF 算法或 EKF 算法用于 D-FNN,当一条规则产生/去除或者宽度有任何调整时,将产生额外计算负担,即在那些情况下,迭代必须从第 1 个样本开始。本文将分别使用线性最小二乘(Linear Least Squares, LLS)、KF 以及 EKF 法实现 D-FNN 学习算法,并比较它们在如下 4 个方面的性能:学习速度,模糊规则数,逼近精度,抗噪能力^[5]。本文使用如下 2 个输入 x_1 、 x_2 和单输出 y 的简单非线性函数:

$$y = \frac{\sin x_1}{x_1} \cdot \frac{\sin x_2}{x_2}, \quad -10 \leq x_1, x_2 \leq 10 \quad (13)$$

并用它来估计 3 种不同算法的性能。在该仿真中,使用 200 个训练数据训练 D-FNN,输入部分在输入范围内随机产生。学习时间、模糊规则数以及性能比较如表 1 所示。

表 1 不同参数调整方法之间的性能比较

性能	EKF	LLS	KF
规则数	7.000 0	5.000 0	9.000 0
学习时间/s	15.430 0	14.010 0	22.190 0
平均均方根误差	0.008 5	0.005 8	0.008 6
平均输出误差	0.022 5	0.036 8	0.015 7
最后的均方根误差	0.009 4	0.015 1	0.013 2
最大的输出误差	0.170 9	0.570 7	0.061 4
最后的输出误差	0.006 8	0.032 2	0.013 2

通过以上仿真结果可以看出：

- (1) LLS 方法是最快速的学习算法，需要最少的模糊规则数，在 3 种方法中具有最高的均方根误差。
- (2) 虽然 KF 方法最终需要最多的模糊规则数，但是它在整个学习过程中的实际误差一直较小且变化不大。
- (3) 从整体上讲，EKF 方法具有最好的性能，与 LLS 方法相比，模糊规则数略多，但当以大致相同的学习速度达到较低的均方根误差时，该缺点可以被弥补^[6]。

2 实验结果与分析

时间序列在解决许多实际问题中是很重要的。时间序列预测可应用于诸多领域，如经济和商业计划、库存和生产控制、天气预测、信号处理、控制和其他领域。本文用 D-FNN 进行 Mackey-Glass 混沌时间序列预测如图 1~图 4 所示。以下仿真在 D-FNN 中使用串-并行方法。

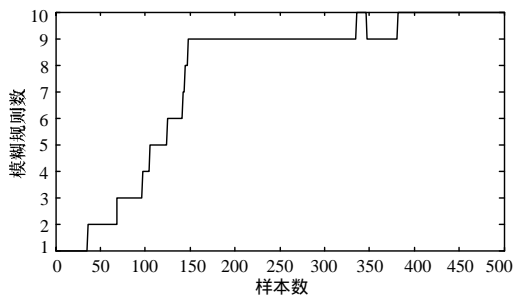


图 1 模糊规则的产生

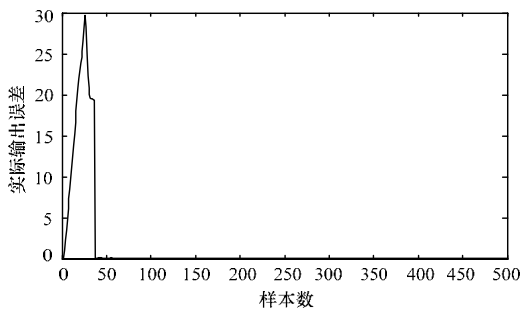


图 2 实际输出误差

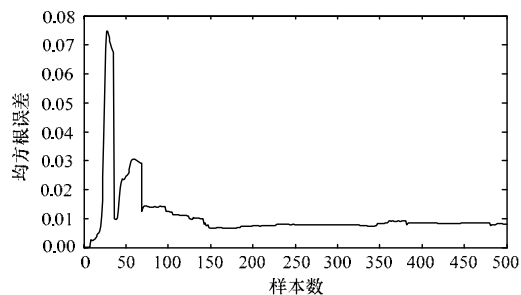


图 3 均方根误差

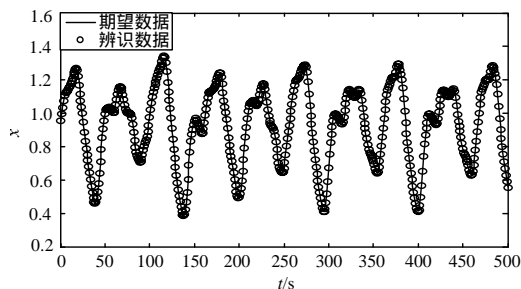


图 4 期望数据、辨识数据与时间的关系

Mackey-Glass 混沌时间序列是 Mackey MC 等在 1977 年提出的一个模型，用于描述白血病发病时，血液中白血球的数量变化，其方程如下：

$$x(t+1) = (1-a)x(t) + \frac{bx(t-\tau)}{1+x^{10}(t-\tau)} \quad (14)$$

由于白细胞新陈代谢生理机制方面的原因，因此式(14)表现了某种周期性与混沌特征。当 $\tau < 17$ 时，表现周期性，当 $\tau > 17$ 时，表现出混沌行为， τ 值越大混沌现象越严重。对这种时间序列的预测是一个典型问题。为了能在相同基础上进行比较，参数选择为 $a=0.1$ ， $b=0.2$ 和 $\tau=17$ ，预测模型如下：

$$x(t+p) = f[x(t), x(t-\Delta t), x(t-2\Delta t), x(t-3\Delta t)] \quad (15)$$

为了训练和测试，用式(14)在 $t=0\sim 6000$ 之间产生 6000 个数据。式(14)初始条件选为：当 $t < 0$ 时， $x(t)=0$ ， $x(0)=1.2$ 。这些数据将作为输入输出对，用来训练预测模型(式(15))，使之逼近式(14)，并测试其性能。

设 $p=6$ ， $\Delta t=6$ ，在 $118 \sim 617$ 之间选择 500 对样本，在 $118 \sim 1117$ 之间选择 1000 个样本作为训练数据。500 个~1000 个样本用来验证被辨识的模型。图 1~图 4 描绘了从 $t=118\sim 617$ 的训练结果和前 6 步预测结果。选择不同参数，将得到 2 种不同的模糊神经网络结构，其训练和测试仿真结果如表 2 所示。

表 2 训练与测试数据的仿真结果

结构	规则数	训练的均方根误差	测试的均方根误差
结构 1	5	0.013 0	0.013 3
结构 2	10	0.008 2	0.008 3

当 $p=6$ ，模糊规则数为 10 时，图 5、图 6 说明了 D-FNN 的泛化性的测试，其中，前 500 个点用于训练，后 500 个点用于预测。

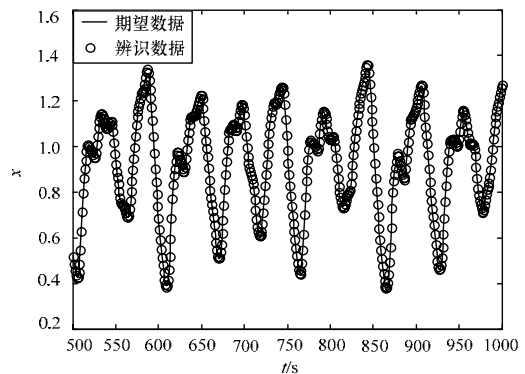


图 5 期望数据、辨识数据与时间的关系

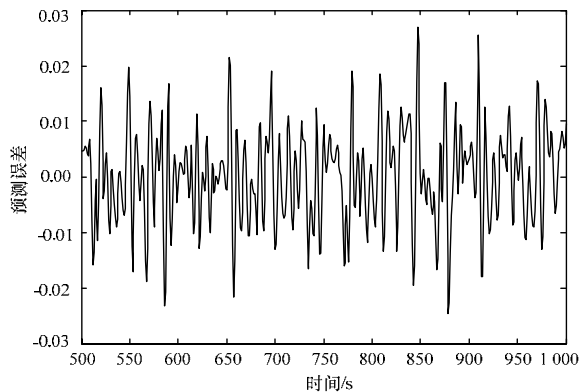


图 6 预测误差与时间的关系

由图 5、图 6 及表 2 可以看出,无论采样 5 条规则还是 10 条规则,构建的模糊神经网络都具有很强的泛化能力,其原因是测试误差几乎等同训练误差。

表 3 列出了 D-FNN 与基于自适应神经网络的模糊推理系统(Adaptive-based Neural-Fuzzy Inference System, ANFIS)、正交最小二乘网络(Orthogonal Least Square network, OLS)和径向基自适应模糊系统(Radial Basis Function-Adaptive Fuzzy System, RBF-AFS)性能的比较结果,训练条件为 $n=500, p=6$ 。

表 3 不同结构的性能比较

方法	规则数	参数数量	训练的均方根误差	测试的均方根误差
D-FNN	10	100	0.008 2	0.008 3
ANFIS	16	104	0.001 6	0.001 5
OLS	35	211	0.008 7	0.008 9
RBF-AFS	21	210	0.010 7	0.012 8

选择 $\Delta t = 6, p = 50, p = 80$, 为了方便比较,用归一化的均方根误差或无量纲指数评估泛化能力,其中,无量纲误差指数为目标序列标准差除均方根误差。当 $p=85$, D-FNN 与 ANFIS 的泛化能力比较如表 4 所示。

表 4 D-FNN 与 ANFIS 的泛化能力比较

方法	数据量	规则数	参数的数量	无量纲误差指数(NDEI)
D-FNN	500	14	140	0.337 5
ANFIS	500	16	104	0.036 0

当 $p=50$ 时, D-FNN 与资源分配网络(Resource Allocation Network, RAN)、扩展卡尔曼滤波的资源分配网络(Resource Allocation Network Extended Kalman Filter, RANEKF)以及最小资源分配网络(Minimal Resource Allocation Network, M-RAN)的泛化能力比较如表 5 所示。

表 5 D-FNN 与 RAN, RANEKF, M-RAN 的泛化能力比较

方法	数据量	规则数	参数的数量	无量纲误差指数
D-FNN	2 000	25	250	0.054 4
RAN	5 000	50	301	0.071 0
RANEKF	5 000	56	337	0.056 0
M-RAN	5 000	28	169	0.061 0

由此可知,即使 D-FNN 有更多的可调参数,它的性能也

没有 ANFIS 好,其原因是 ANFIS 通过迭代学习的方式训练,从而达到整体最优解,而 D-FNN 只能获得次优结果。与 RAN, RANEKF 及 M-RAN(它们也只能得到次优解)相比, D-FNN 结构小、泛化能力更强。

3 结束语

在 D-FNN 算法中,非线性参数是由训练样本和启发式方法直接决定,而没有用优化算法来确定,虽然高斯宽度在学习时可以自适应地调整,但学习规则却很简单。另一方面,线性参数每一步都由 LLS 或者 KF 方法决定,而不像 RAN, RANEKF 及 M-RAN 是由预测误差确定,这些最优参数可以在学习过程中得到。因此,用该方法得到的解是全局最优的,且只需一步训练而无需迭代学习。采用分级学习的思想能保证更简洁的结构和更短的学习时间。仿真结果表明, D-FNN 具有以下明显的优点:快速的学习速度,紧凑的系统结构,强大的泛化能力。

参考文献

- [1] Hong Ziguang. Algebraic Feature Extraction of Image for Recognition[J]. Pattern Recognition, 2007, 24(2): 211-219.
- [2] Turk M, Pentland A. Eigenfaces for Recognition[J]. Cognitive Neuroscience, 2006, 3(1): 71-86.
- [3] Valentin D, Abdi H, O'Toole A J, et al. Connectionist Models of Face Processing: A Survey[J]. Pattern Recognition, 2004, 27(4): 1209-1233.
- [4] 肖正, 张世永. 基于神经网络的 Agent 个性化行为选择[J]. 计算机工程, 2009, 35(24): 199-201.
- [5] Chen S, Cowan C F N, Grant P M. Orthogonal Least Squares Algorithm for Radial Basis Function Network[J]. IEEE Trans. on Neural Networks, 2001, 20(2): 302-310.
- [6] 吕志胜, 赖惠成. 基于 RBF 神经网络与 RLS 算法的均衡器[J]. 计算机工程, 2009, 35(22): 200-201.

编辑 陆燕菲

(上接第 191 页)

存在尺度,视角或是光照上的差异,但采用了模糊量化以后系统仍然具有较高的返回率,相比于采用严格量化,返回率有了比较明显的提高;第 11 行~第 14 行的图片来自于 Caltech256,而第 15 行~第 18 行的图片则来自于 Caltech101,由于这些图片当中包含的不是相同物体而只是同一类的物体,且场景差异非常大,返回率相比于前 10 行有比较明显的下降,特别是当系统采用严格量化时,系统的返回率呈急剧下降之势,可以看出,系统的返回率大部分只有 25%,即只能返回一张相关的图片,这就说明严格量化对相似图片的搜索做得不够好,而采用模糊量化之后则可以使系统的返回率维持在一个比较高的水平上,使得环境干扰或是拍摄条件所带来的负面影响尽量减小。另外,系统的响应时间主要由 SIFT 特征提取和特征量化 2 个部分来决定,而从表中数据可知, SIFT 特征提取过程还是非常迅速的,而受机器配置和实现方法(纯 Matlab 代码)的限制,特征量化的时间还是比较长,但是在同等条件下,比文献[2]中非层次化的搜索方法还快得多。

7 结束语

本文借鉴文本检索当中的 TF-IDF 模型,把图片转换成关于单词的权值向量表示,利用向量间的比较从而达到图片

搜索的目的。而针对搜索非层次化单词所带来的量化过程太慢的问题,使用层次化的词汇树模型来加快特征的匹配和量化,针对严格量化所带来的返回率偏低的缺陷,特别是对相似物体搜索时返回率非常低的情况,使用模糊量化来改进这一过程,增强了系统的鲁棒性和实用性。下一步研究方向是优化代码来缩短量化时间,调整某些参数选择,并增加更多图片进行测试。

参考文献

- [1] Lowe D G. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [2] Sivic J. Video Google: A Text Retrieval Approach to Object Matching in Videos[C]//Proc. of the International Conference on Computer Vision. [S. l.]: IEEE Press, 2003.
- [3] Baeza Y R. 现代信息检索[M]. 王知津,译. 北京:机械工业出版社, 2005.
- [4] Nister D. Scalable Recognition with a Vocabulary Tree[C]//Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2006.

编辑 陈文