

一个面向构思的手绘草图识别系统^{*}

邱庆华 张 斌 孙正兴

(南京大学软件新技术国家重点实验室 计算机科学与技术系 南京210093)

摘 要 手绘图形输入是用户设计意图的一种自然有效的表达方式。本文介绍了一个手绘图形识别系统—SketchEI。SketchEI由五个部分构成:用户交互、输入预处理、图形识别器、草图管理器以及用户适应性。SketchEI把它们有机地组合在一起并且给出了一个相对完整的手绘图形识别技术理论体系。

关键词 空间关系图,支撑向量机,用户适应性,增量学习

A Concept Oriented Sketchy Recognition System

QIU Qing-Hua ZHANG Bin SUN Zheng-Xing

(National Laboratory for Novel Software Technology, NanJing University, 210093)

Abstract Sketch-based graphics input is a natural and effective way to present designer's brainstorm or bursting ideas. We propose a sketch-based graphics recognition system—SketchEI in this paper, which can discovery latent geometric shapes from users' sketchy input and show the recognized regular shapes on the screen immediately. SketchEI organically integrates pen-based user interface, sketch preprocessor, sketch recognizer, historical sketch manager and user adaptor and produces good performance.

Keywords Sketch recognition, User adaptation, Sketch-based user interface, Conceptual design

1. 引言

手绘草图(Sketching)是人类的一种常用的思路外化和通信交流方式,它以视觉空间媒体表达和传递概念,既有抽象、隐喻等特点,还具有形象、直观等特征,易于理解和记忆,更适于推论和构思;与传统图形图像工具相比,它具有自然、简便,可表达不完备模糊概念^[1],便于用户进行创造性活动等特性^[2~4]。手绘草图输入的历史可以追溯到 Ivan Sutherland 的 SketchPad 系统,20世纪90年代中期以来,人们对草图识别技术及其应用进行了大量的研究,取得了一大批成果。归纳起来,他们的研究成果主要集中在这样几个方面:针对单一笔划的识别^[5~8],针对组合笔划的识别^[9~11]以及用户交互^[12~14],并在此基础上产生了一些系统^[10,15]。

尽管在草图识别领域取得了大量的成果,涌现了许多识别系统,但是这些系统都存在这样一些共同的不足:1. 只注重草图识别中的某一方面或某几方面的问题,缺乏相对完整的草图识别技术理论体系;2. 没有屏蔽用户绘制习惯的差异,不具有良好的多用户适应性;3. 没有考虑到草图设计过程中信息的管理和利用。为此,我们研究开发了一个面向构思的手绘草图识别系统 SketchEI,该系统的主要设计思想在于以下几个方面:

面向概念设计:系统是面向概念设计的,它支持的是用户的构思和创造性的工作而不是详细的设计或者是成熟的想法展现,系统识别的图形不集中在某个领域,因而识别算法本身具有一定的适应性。同时系统特别注重对用户笔划输入的随意性和模糊性的支持,保证用户设计工作时思维和表达上的流畅性。

完整的体系结构:系统不是针对草图识别的某个方面进行研究,它研究了草图识别相关的用户交互、输入预处理、图形识别、草图管理和多用户适应等多个方面,形成一个相对完整的草图识别技术理论体系。

人性化的用户交互:系统考虑了界面的设计风格、交互时机、Gesture 的设计以及可定制化的交互方式,使得用户可以按照自己习惯的方式和系统进行交互。

多用户适应性:SketchEI 充分考虑了对具有不同知识背景的用户的支持,通过提高算法的鲁棒性和对用户输入信息进行捕捉学习并建立用户模型的方式,最大程度地屏蔽了不同用户在输入习惯思维模式上的差异。

本文第2节对 SketchEI 的体系结构进行描述,简单概括了系统的五个组成部分:用户交互、多用户适应、预处理器、图形识别器以及草图管理器的主要功能和作用。第3节介绍了系统的三个关键技术:笔划分割器、组合图元识别和多用户适应性,最后简单小结并指出后续的工作。

2. SketchEI 体系结构

SketchEI 系统主要包括五个部分:用户交互、多用户适应、预处理器、图形识别器和草图管理器,其结构可由图1来表示,其中用户交互由绘制界面和用户反馈组成。

2.1 用户交互

用户交互包括用户工作界面、用户的工作方式即草图输入以及编辑修改方式、系统对输入草图处理结果的反馈以及用户对系统处理的干预等等。在 SketchEI 中我们重点考虑了用户设计时草图输入和编辑的自由和流畅性以及用户工作方式的定制。

^{*}) 本文得到国家自然科学基金项目(编号:69903006、60373065)资助。邱庆华 硕士生,研究方向为在线图形识别。张 斌 硕士生,研究方向为智能化人机界面;孙正兴 副教授,博士后,研究方向为多媒体挖掘和图形图像技术。

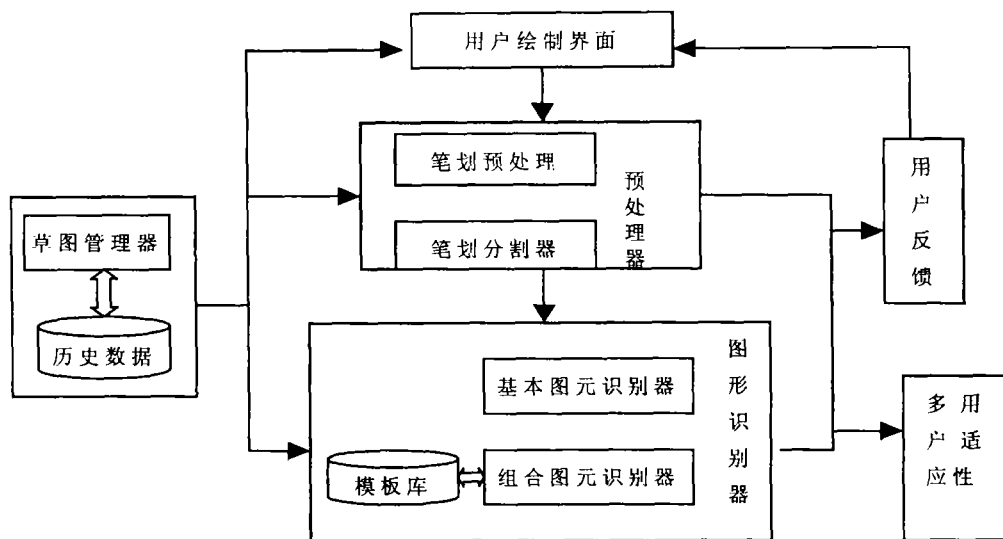


图1 SketchEI 结构示意图

为了实现方便易用的草图编辑和修改,我们设计了复制、删除、拖动、粘贴、撤销、恢复、清空面板、完成、圈定和进入共十种 gesture 供用户使用,在 gesture 设计过程中考虑了易画性和易记忆性等特点。在用户工作方式方面,SketchEI 提供多种方式让用户自由定制,用户可以根据自己的习惯和喜好来选择系统反馈的时机和方式。

2.2 预处理器

预处理的目的是消除由输入条件限制和输入习惯的差异而引起的噪音,完成笔划的分割,为图形识别做准备。预处理主要包括笔划预处理和笔划分割器。笔划预处理的目标是消除笔划冗余点,曲线闭合误差等噪音,它包括冗余点去除,聚点消除,端点校正和闭包计算。笔划分割器的作用是在笔划预处理的基础上,抽取笔划特征找出最佳分点,然后按照最佳分点对笔划进行分割,分割结果作为识别的前提。

2.3 图形识别器

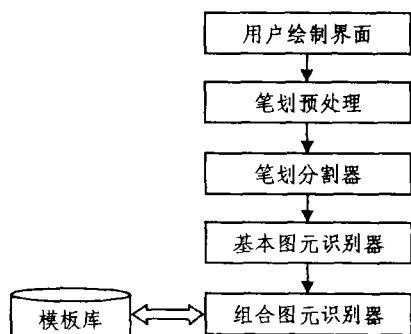


图2 SketchEI 识别过程

预处理器和图形识别器结合在一起就构成了 SketchEI 的主要识别过程(图2),图形识别器是整个草图识别系统的核心。图形识别器包含基本图元识别器和组合图元识别器。基本图元识别器是在笔划分割的基础上对笔划进行形状拟合,它根据笔划的拟合代价来确定最佳的拟合方案,最终将笔划识别成直线段、弧段和椭圆三种基本图元。组合图元识别器在图

元识别的基础上抽取图元之间的位置关系,获得组合图元的 SRG(Spatial Relation Graphs)表示,然后利用 SRG 进行模板匹配,以识别出用户输入的图形。

2.4 草图管理器

在用户的草图设计过程中,除了输入的笔划几何信息外还会产生许多其它信息如画图速率、笔划顺序等等,这些信息可以用来指导识别过程提高识别效果。这些信息需要专门的模块负责抽取管理和利用,这个专门的模块就是草图管理器。草图管理器是一个面向草图设计过程的信息收集管理工具,它负责收集用户设计草图过程中的用户原始输入信息和识别过程中产生的结果信息并加以管理利用。

2.5 多用户适应性

不同用户的思维和绘画习惯具有很大的差异,有些方面甚至是矛盾的,在进行草图绘制时,不同用户输入的不同图形尽管可能非常相似,但是他们的意图却是不同的,这就为正确的意图预测带来了困难。针对这个问题,SketchEI 建立了一种用户适应机制,这种机制能够通过用户的输入和反馈来收集用户相关的信息,然后通过学习的方法从用户数据中找到与其它用户不相冲突的特性,使得系统逐步适应用户特性或习惯,从而准确地预测用户的意图。

3 关键的技术

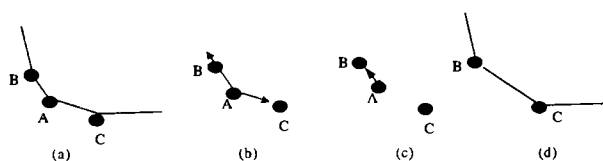
SketchEI 作为一个草图识别器,其核心在于识别,因此系统的关键技术也就是识别相关的技术。系统的关键技术体现在三个方面:输入笔划预处理、图形识别以及多用户适应性,这三者都是紧紧围绕着草图识别这个主题的,输入笔划预处理是对用户原始输入的处理技术,处理的结果提交给图形识别器来进行图形识别,而多用户适应性是为了屏蔽系统对于不同用户表现出来的性能差异,使系统的识别效果不会因人而异。

3.1 笔划分割器^[16]

笔划分割器是预处理器中的核心部分,它的任务是在预处理的基础上完成对用户输入笔划的分割,为图形识别做好准备。笔划分割器主要由基本图形分类和笔划分割两部分组成。

3.1.1 基本图形分类 SketchEI 利用“引力模型”来识

别基本图形。系统认为点与点之间有相互吸引的趋势,如果这种趋势大于某一个阈值的话,就认为这两个点之间的内部能量很大而受到外部的影响较小,可以将它们合为一个点。设曲线上任一点 A 的相邻点为 B 和 C, A 受到 B 和 C 的吸引。如果 B 对 A 的吸引力更强并且高于设置的阈值,则将 A 和 B 合并为一点。如图3所示。



(a) B, C 是 A 的相邻顶点; (b) A 受到 B 和 C 的吸引;
(c) B 对 A 的吸引力更强并且高于设定的阈值; (d) 将 A、B 合并

图3 顶点合并过程

给定 A, B 的转角分别为 α, β , $\text{Dis}(A, B)$ 是 A, B 两点的距离, $f(A, B)$ 是两顶点之间的引力。 $f(A, B)$ 应满足以下条件: ① α 越大, $f(A, B)$ 越大; ② β 越大, $f(A, B)$ 越小; ③ $\text{Dis}(A, B)$ 越大, $f(A, B)$ 越小。由此定义:

$$f(A, B) = \frac{\alpha}{\beta \text{Dis}^2(A, B)}$$

对于任意一顶点, 计算两个相邻点对它的吸引, 如果引力值超过了阈值, 将该点和相邻点合并。一般情况根据不同的用户输入进行调整。经过聚点消除后, 设剩下点的数目为 $V_{\min}$, 若 $V_{\min} \leq 3$, 则最终形状为三角形; 若 $V_{\min}$ 为 4(5, 6), 则分别对应四(五, 六)边形; 若 $V_{\min} > 6$, 则最终识别成椭圆。

3.1.2 笔划分割 系统采用搜索的方法寻找最优分割, 设曲线的节点数为 m , 拟合多边形的顶点数为 n 。设曲线的点序列为 S , 其中节点为 $P_i (i=0, 1, 2, \dots, m)$; 拟合后的多边形顶点序列为 T , 其中各个顶点为 $q_j (j=0, 1, 2, \dots, n)$ 。以下定义:

$D1 = \sum \text{Dis}(p_i, T) / m$ 为 S 到 T 的距离;

$D2 = \sum \text{Dis}(q_j, T) / n$ 为 T 到 S 的距离;

$D = D1 + D2$ 为曲线拟合成多边形的代价。 D 越大, 表明曲线与多边形的相似度越低。

搜索算法如下(最佳分点搜索算法):

1. 在 m 点中任取 n 点;
2. 根据上面所取的 n 个分点将曲线拟合成 n 边形;
3. 计算曲线与多边形之间的拟合代价 D ;
4. 如已列举所有情况, 跳至下一步, 否则跳至第一步;
5. 取拟合代价最小的情况作为最终的拟合结果。

由于第一步中是任意选取 n 个分点, 因此, 会有多种选择。如果在 S 中穷举所有可能的情况, 不仅增加了运算复杂度, 由于一些非关键点存在, 也会影响拟合精度。因此, 需要对算法优化。在聚点消除阶段, 我们通过阈值来控制最终点序列的规模。因此, 我们可以首先用较大的阈值获得候选点序列 S , 用较小的阈值在 S' 中确定曲线的最终形状 n 。在拟合阶段, 我们用 S' 代替 S 作为源点序列, 这样, 不仅降低了运算复杂度, 而且 n 是由 S' 确定的, 因此, 在 S' 中寻找分点进行拟合又可以保证最终的效果可以最大程度地逼近理想情况。

3.2 基于 SRG 表示的图形部分匹配^[17]

图形识别器是在笔划分割的基础上进行图形的识别, 它包括基本图元的识别和组合图元的识别。基本图元的识别是通过拟合的方法来实现的, 它最终将分割好的笔划拟合成线

段、弧段、椭圆三种基本图元。组合图形识别利用 SRG 部分匹配的方法, SRG 部分匹配可以很好地满足图形匹配中缩放不变性, 平移不变性和旋转不变性的要求, 具有较强的适应性。

3.2.1 组合图元的 SRG 定义 SRG 可以用一个 7 元组 $G = (V, E, T_v, R_e, E', G_e, P_e)$ 来描述, 这个 7 元组是在获得的基本图元信息的基础上做进一步的关系提取获得的。7 元组中 V 是组合图元顶点集, E 是边的集合, T_v 定义了从点集到边集的映射规则, 这三个元组可以直接由基本图元信息得到。 R_e 定义了从边集到空间邻接关系的映射, 空间邻接关系共分五种: 内部连接 (R_{ic}), T 型相交 (R_T), 十字相交 (R_{is}), 平行 (R_P), 包含 (R_c), 空间邻接关系可以用集合 $\sum_R \{R_{ic}, R_T, R_{is}, R_P, R_c\}$ 来描述, 空间邻接关系是旋转和镜像无关的。相对位置关系则主要用来描述平行关系的方向, 分成四类: 垂直, 平行, 左斜和右斜关系, 并用符号 g_v, g_h, g_L, g_R 分别来表示, 整个关系可以用集合 $\sum_r = \{g_v, g_h, g_L, g_R\}$ 来描述, 相对位置关系是旋转和镜像敏感的。 E' 是 E 中和其它边发生空间邻接位置关系的边的集合, P_e 则用来表示边的方向, R_e, E', G_e, P_e 这四个元组是通过基本图元做进一步的关系信息抽取获得的。

3.2.2. SRG 的部分匹配 有了图形的 SRG 描述之后, 我们就可以对图形进行部分匹配。对于两个 SRG, $G_1 = (V_1, E_1, T_{v1}, R_{e1}, E'_1, G_{e1}, P_{e1})$, $G_2 = (V_2, E_2, T_{v2}, R_{e2}, E'_2, G_{e2}, P_{e2})$, (其中 G_1 为源对象, G_2 为目标对象)。一个成功的部分匹配将满足以下四个条件:

- 1) G_1 中的每个顶点都能映射到 G_2 ;
- 2) G_1 中的每条边都能映射到 G_2 中;
- 3) G_1 中的每对相邻关系都能映射到 G_2 中;
- 4) G_1 中的相对位置关系映射到 G_2 后是不是一致。

条件(1)和(2)保证 G_1 中的每个顶点和每条边都能被正确地映射到 G_2 , 条件(3)检查了源对象中的每个空间邻接关系对于候选对象而言是不是都是可接受的。如果源对象中的每个边都能在候选对象中找到其对应的映射则这种映射被认为是成功的。条件(4)检查经过映射之后同一个组的相对位置关系有没有发生变化, 基本图元间联系方向变化不在检查之内。尽管相对位置关系是和基本图元间联系的方向是相关的, 但是相对位置关系之间的相似度比较却与之无关而且对镜像和旋转也是无关的(图4显示一个从 G_1 到 G_2 的成功部分匹配)。

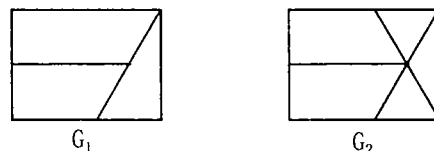


图4

3.3 基于 SVM 的增量学习算法^[18]

用户适应性的解决方法有两种: 基于规则的相关反馈方法和基于重复统计学习的方法。但这两种方法都存在缺陷, 基于规则的相关反馈方法在针对多用户的时候有过学习现象, 而经典 SVM 虽然没有过学习现象但是它采用的重复学习机制在样本膨胀情况下会导致效率急剧下降, 作为改进的方法, 基于 SVM 的增量学习综合了 SVM 和增量学习的优点, 有效地避免了过学习和效率下降的情况。本系统采用基于

SVM的增量学习算法来实现多用户适应,主要工作分为两个步骤:首先要抽取特征样本供学习使用,然后构造一个多值分类器来对样本进行增量学习。

3.3.1 特征抽取 在线图形识别系统中,处理的图元与它们的大小与位置是无关的,要对记录下的用户的图形进行规一化,即将图形的周长调整为1。然后使用转角函数对图形进行特征抽取,得到一个 d 维的特征向量。转角函数 $\Theta_A(s)$ 标识的是图形边界上由初始点到当前点(s)累积的转角。其中 $\Theta_A(O)$ 为选取的初始点 O 的切线方向与水平方向的夹角。

为了使转角函数更为简单,转角函数被加了一些限制,规定值在 $[0, 2\pi]$ 的区间内,并且初始点的值为0。为此,转角函数可定义如下:

$$V_i = \begin{cases} \varphi & \text{if } \varphi \geq 0 \\ \varphi + 2\pi & \text{if } \varphi < 0 \end{cases} \quad 0 \leq i \leq d$$

其中 $\varphi = \Theta_A(\frac{i \bmod d}{d}) - \Theta_A(\frac{i-1}{d})$, d 为需要抽取特征的维度,即将图形的边界 d 等分。

3.3.2 增量学习 用户在使用在线图形识别系统时,草图总是逐个输入、逐个识别,从机器学习效率的角度来考虑,用户每输入一个图形就对其进行学习,显然是低效的。可是等系统收集到足够的使系统稳定的数据时才开始学习,那必然要让系统长时间在较低精度下运行,显然也是不可行的。但又由于系统在开始学习时用户的数据总是比较少,因此在我们学习时用户的数据集的大小有一个下限和上限。当系统开始使用时,数据量积累到下限(即初始训练集—IS),系统就开始学习,然后等再次收集到一定量的数据(即增量训练集—INS)再开始学习,如此反复系统的适应性就越来越强。现在有许多种基于SVM的解决学习的策略,在这里我们使用文[7]的算法进行增量学习。

我们构造了一个多值分类器来实现增量学习,分类器总共处理了椭圆、三角形、四边形、五边形和六边形等5个类别。所以我们需要构造多值分类器,一般多值分类器的构造有两种模式:一对多模式(One Against All Model)和一对一模式(One Against One Model)。在这里我们采用一对一模式,表1是一对一模式构造的多值分类器的性能描述。

表1

模式	一对一模式
训练	低,平均为 $O(n^3/m)$
分类	需要 $m(m-1)/2$ 次分类
VC 维度	相对较低
增加一个新样本时的复杂度	低,重新训练 $m-1$ 个子分类器计算复杂度为 $O(n)$

因此,使用基于SVM增量学习的方法来实现用户适应,整个学习的过程可以表示为“识别(收集数据)—学习”循环反复的过程。系统收集用户的输入信息,并对其进行规整,然后使用转角函数将其抽取为特征向量。当系统收集的数据量达到一定量的时候,就开始进行学习。对每两个类别训练分类器,共需要训练10个分类器,并记录下它们的特征向量,即训练多值分类器。然后系统利用学习的结果来对用户输入的草图进行识别,用所有的分类器对用户的数据来进行分类,通过投票来决定具体的类别,即利用训练好的多值分类器来进行分类。用户通过界面来判断系统分类的正确与否。系统边识别边收集用户数据和分类结果,当收集的数据满足一定量的时候,将特征向量加入到新收集到的数据集,然后再训练多值分

类器。如此反复,系统逐渐趋于稳定,系统也就适应了多用户。

结束语 作为一个相对完整的体系,SketchEI系统研究了草图识别系统的多个角度方面的问题,包括从用户交互、用户手绘草图的识别、草图的管理以及多用户适应机制,完成了从用户输入笔划的处理到最终图形的识别,将其中有价值的部分加以存储和管理以指导识别过程,形成了一个相对完整的草图识别系统。但是仍然存在一些问题值得继续研究:如何提高较多组元组成的图形的识别效率;如何使图形识别不完全受到模板的制约使识别具有一定的扩展性和适应性;在用户适应性的问题上主动式的增量学习和用户建模是我们继续研究的方向。

参考文献

- 1 孙正兴,徐晓刚,孙建勇,金翔宇.支持方案设计的手绘图形输入工具.计算机辅助设计与图形学学报,2003,已录用
- 2 Maher M L, Boulanger S, et al. Assessing computational methods with a framework for creative design processes. Preprints Computational Models of Creative Design, Sydney: University of Sydney, 1995. 233~265
- 3 Gero J S. Recent advances in computational models of creative design. In: Pahl P, Werner H, eds. Computing in Civil and Building Engineering, Balkema, 1995. 21~30
- 4 van Elsland P A, Vergeest J S M. New functionality for computer-aided conceptual design: The displacement feature. Design Studies, 1998, 19(1): 81~102
- 5 Christine A, et al. A Framework for Multi-Domain Sketch Recognition. MIT Artificial Laboratory, Sep. 2001
- 6 Metin S, Thomas S, Randall D. Sketch Based Interfaces: Early Processing for Sketch Understanding. In: Proc. of PUI-2001, Nov. 2001
- 7 Hong J I, Landay J A. SATIN: A Toolkit for Informal Ink-based applications. In UIST 2000, ACM Symposium on User Interface Software and Technology, CHI Letters, 2(2): 63~72
- 8 Rubine D. Specifying gestures by example. Computer Graphics 25: 329~C337, 1991
- 9 Gross M D. The electronic cocktail napkin: a computational environment for working with design diagrams. Design Studies, 1996, 17: 53~69
- 10 Calhoun C, Stahovich T F, Kurtoglu T, Kara L B. Recognizing multi-stroke symbols. In: AAAI 2002 Spring Symposium Series, Sketch Understanding
- 11 Calhoun C, Stahovich T F, Kurtoglu T, Kara L B. Recognizing multi-stroke symbols. In: AAAI 2002 Spring Symposium Series, Sketch Understanding
- 12 Lin J, Newman M W, Hong J I, Landay J A. DENIM: An Informal Tool for Early Stage Web Site Design. Video poster in Extended Abstracts of Human Factors in Computing Systems, CHI 2001, Seattle, WA, March 2001. 205~206
- 13 Metin S, Thomas S, Randall D. Sketch Based Interfaces: Early Processing for Sketch Understanding. In: Proc. of PUI-2001, Nov. 2001
- 14 Gross M D. The electronic cocktail napkin - a computational environment for working with design diagrams. Design Studies, 1996, 17: 53~69
- 15 Hearst M A, Gross M D, Landay J A, Stahovich T F. Sketching Intelligent Systems. IEEE Intelligent Systems, 1998, 13 (3): 10~19
- 16 孙建勇,金翔宇,彭彬彬,孙正兴.一种快速在线图形识别与规整化方法.计算机科学,2003(2)
- 17 Xu X G, Sun Z X, Peng B B, Jin X Y. 2002b, 'ASRG-Based Online Composite Graphic Recognition Strategy for Sketch-Based User Interface'. In: Proc. of the 1st International Conf. On Machine Learning and Cybernetics, Beijing, pp. 723~728
- 18 彭彬彬,金翔宇,徐晓刚,孙正兴.基于SVM增量学习的用户适应性研究.计算机科学,2003(3)