

基于笔序的手绘草图识别方法^{*}

尹建峰 孙正兴 邱庆华 蒋 维

(南京大学计算机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 本文提出了一种基于用户绘图过程笔序信息的手绘草图识别方法,该方法利用同一用户绘制同一草图的笔序趋向于一致这一特点,在收集用户绘图笔序信息的基础上建立用户模型,从而达到提高输入自由度和识别效率的目的,并用实验加以了证明。

关键词 手绘草图识别,笔序,用户建模,决策树

Freehand Sketch Recognition Based on Temporal Stroke Information

YIN Jian-Feng SUN Zheng-Xing QIU Qing-Hua JIANG Wei

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)

Abstract This paper proposes a method for online freehand sketch recognition base on the temporal stroke information during user sketching. It collects the information of stroke sequence, and then builds a user model which can be used for high freedom and efficient sketch recognition. The experiments has proved the efficiency.

Keywords Sketch recognition, Stroke sequence, User modeling, Decision tree

1 引言

手绘草图识别是根据人机交互过程中不断增长的草图信息,挖掘草图形状约束,领会用户输入意图,最终将自由的、不规则的草图识别成为规则几何形状。手绘草图识别的难易与用户草图绘制的自由度相关,对用户输入限制越多的草图越容易识别。因此,草图识别需要在用户绘制草图的自由度和识别难度之间进行折衷。如何在保证用户输入自由度的前提下提供高效的手绘草图识别方法一直是草图理解领域的一个研究重点。

已有草图识别方法主要包括四类:一是离线方法,如:Revankar 等^[1]将手绘图形分割成几个部分,通过一些规则建立决策树,每个叶节点输出一种形状,但在线草图识别受到输入的随意性以及视觉偏差的影响,不可能将图形分割成标准的笔划。二是基于特征相似度计算方法,如:我们项目组前期提出的边表方法^[2],以组成图形的图元类型和数目为特征计算相似度,这种方法速度极快,但是由于特征过于简单,检索结果不够精确;而 SRG 方法^[3]则以图元间的简单关系为特征计算相似度,具有精确的识别结果,但时间开销过大。三是基于过滤器的方法。如: Takayuki 等人提出用过滤器进行分类^[4],但除圆和直线外,其它图形必须与 X 轴或者 Y 轴平行输入,这个要求对于在线草图输入过于苛刻; Fonseca 和 Jorge 及其研究小组对这种方法进行了改进^[5],但改进的方法无法区分彼此诸如五边形和六边形等很接近的图形,而且,其扩展性较差。四是机器学习方法,神经网络、支撑向量机、隐马尔可夫模型等方法也已经应用于手绘草图识别,并具有识别率高的优点,但其对大样本数据的需求、训练时间长等缺点,限制了其推广范围,如: PDA 环境。据此,本文试图通过利用用户绘图笔序信息,建立手绘草图的用户模型,降低对用户输入的约

束,并提高手绘草图识别的效率。

2 基于笔序的草图识别方案

手绘草图识别本质上是鉴别其基本构成信息单元及其相互间的关系。由于用户输入同一个图形时可采用多种笔划组合来实现,这会使得图形识别复杂度大大增加。而一个图形的基本图元(如:线段、圆弧等)组成是唯一的。因此,以基本图元作为手绘草图的最小信息构成单元可降低手绘图形识别的复杂度并且提高识别方法的通用性。

手绘草图中基本图元间的相互关系包括两个部分:草图图形构成中两个基本图元之间空间位置关系;手绘草图中基本图元生成先后的时间关系(笔序)。我们假设:任一用户绘制同一个图形时具有固定的笔序,或是趋向于集中的几种笔序。利用这个假设,我们为用户建立一个可在线训练和动态调整的用户模型。在图形识别的过程中可以将某一图形的笔序信息保存在用户模型中,而在下一次同样的用户输入图形的识别中就可以利用用户模型进行快速的识别。

我们的方案如图1所示,主要包含三大部分:笔序信息抽取、用户模型和匹配机制。笔序信息抽取将用户输入的草图信息经消除噪声的预处理^[2]而将笔划分割识别成基本图元,进而将手绘草图表示为基本图元间的几何和时序关系。用户模型采用决策查找树来记录用户绘制习惯信息。匹配机制包括两个部分:主选匹配机制是基于用户模型决策树的图形匹配,辅助匹配机制是利用其它匹配方式来进行匹配,起到在线训练的作用。在本文中,我们采用的辅匹配机制是 SRG 匹配方法(这个是可替换的),在这种情况下,给出性能评价公式:设用 SRG 法识别的时间为 T_s ,用用户模型的识别时间为 T_u ,当命中率为 m 时,平均的识别时间为: $T = m * T_u + (1 - m) * T_s$ 。

^{*} 本文得到国家自然科学基金项目(编号:69903006、60373065)资助。尹建峰 硕士生,研究方向为智能用户接口;孙正兴 教授,博导,研究方向为:智能用户接口、多媒体挖掘和图像技术;邱庆华 硕士生,研究方向为在线图形识别;蒋 维 硕士生,研究方向为智能用户接口。

由于使用用户模型的识别速度非常快,当命中率较高时将明显地加快平均的识别速度。

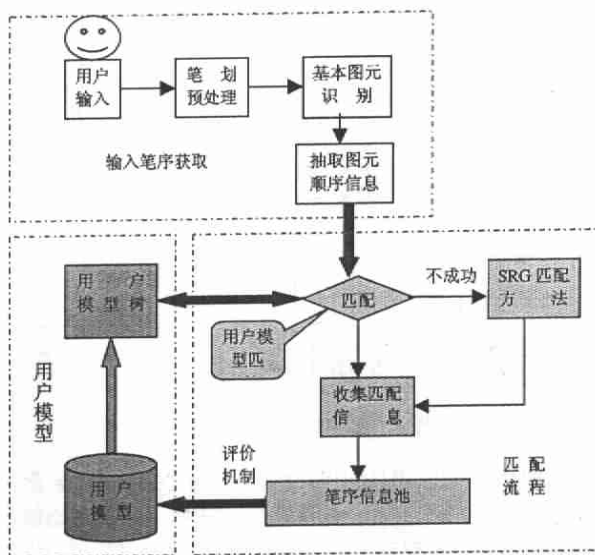


图1 基于笔序的手绘草图识别方案

3 基于用户模型的图形识别

基于用户模型的图形识别由三个部分:决策查找树、笔序信息池以及智能迭代器协作完成。笔序信息池忠实地记录了每一次用户绘制草图的笔序信息,决策查找树则是通过对笔序信息池中的数据进行统计、评价之后得到的,它为算法匹配直接提供信息支持。

3.1 用户模型

用户模型记录了单个用户在绘制某个模板库中图形时的有用信息,包括三个部分:笔序信息池、决策查找树以及智能体。

笔序信息池是保存用户绘图的笔序信息以及每种笔序的方法使用频率统计的模块,它以线性表的形式记录用户的每一次绘图的笔序信息。

决策查找树是用户绘图的笔序信息的树型组织形式。它由根节点、叶节点、中间节点和节点间的边组成。根节点是检索过程的入口。叶节点是图形的模板ID,也就是检索的结果。中间节点是图元的类型,每个中间节点表示一种图元。节点之间的边指相邻两种图元之间的位置关系,每一个非叶节点根据本节点表示的图元和下一个图元之间的空间关系进行分支。

决策查找树与笔序信息池是相同信息的不同形式的表示,这是由其不同的功能所决定的。决策查找树是用于检索的数据结构,它将笔序信息组织成决策树的形式,每一次的检索过程就是搜索决策树的过程。搜索的层数等于输入草图的图元个数,效率非常高。笔序信息池的功能主要是用户模型的管理。使用线性表的形式能够很容易地对使用频率等信息进行统计、或是选择机器学习的方法以确定每一种笔序画法的优先级。根据优先级对线性表进行增删,然后根据线性表对决策树进行剪枝等调整。

智能体是决策查找树和笔序信息池之间的粘合剂,目的是为了降低它们之间的耦合度。其具体功能是:(a)发起一个对决策树的查找;(b)在查找过程中帮助完成决策树的生长同时也在查找过程中收集信息;(c)查找完成后离开决策树,根据

自身信息主动去搜寻笔序信息池,修改相关信息。

3.2 智能体的创建

智能体是用于检索和信息收集的对象,它包括:用户输入草图预处理后的信息;决策查找树的指针(指向检索决策树当前搜索位置);检索进度指针(指向草图预处理后信息中当前搜索位);检索信息收集(将要加入到笔序信息池中的信息)。

系统接受了用户的输入笔划之后对其进行了预处理,得到一个有序的基本图元序列。然后抽取出所有相邻输入图元的空间关系,得到的序列不仅包括基本图元类型及其生成顺序,还包括图元间的相互关系信息。本文中给出了五种关系:相邻、T型相交、十字相交、平行和无关系。

3.3 智能体的搜寻

搜寻决策查找树的过程就是草图识别的过程,识别过程由智能体完成。智能体在创建时包含了用户输入草图的图元顺序及相邻图元间关系信息,智能体运行时根据这些信息在决策树中选择合适的分枝向下搜索,或者生成新的分枝。

在搜索过程中要完成三项工作:(1)识别草图;(2)扩充用户模型;(3)收集缓冲区数据信息。搜索算法如下:

①取得决策树的根节点,获取第一个图元的类型,根据该类型选择根节点的一个子节点作为当前节点。检索进度指针后移一位。

②判断检索进度指针是否已到头,没有则跳到④。否则说明已经搜索完最后一笔。

③查找本节点的model指针,model不为空,其指向的ID就是识别出的模板的ID;model为空说明这种笔序的画法在决策树中没有记录,则使用其他方法识别该草图,并将model指到识别出的ID;跳到⑤。

④获取下一个(关系,类型)二元组,并据此找到对应的子节点的指针。如果该指针不为空,则选择该子节点为当前节点;否则,建立一个新节点,让该指针指向新节点,设置新节点的相关参数,选新节点为当前节点。收集相应的缓冲区数据信息,转至②。

⑤将收集的缓冲区信息应用到数据链中。

一个完整的成功使用用户模型进行草图识别的过程如图2所示。

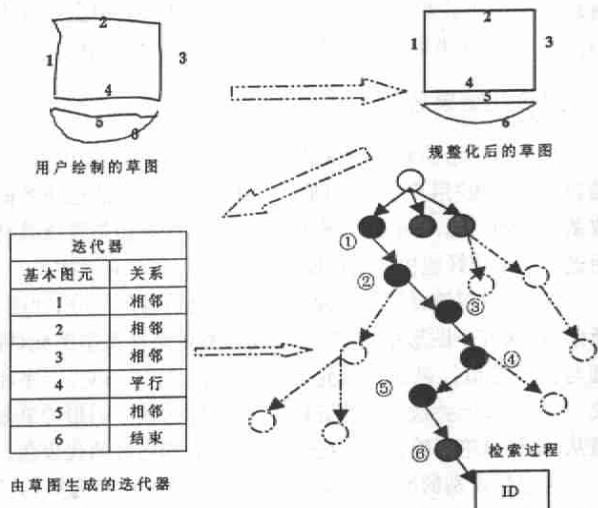


图2 基于用户模型的草图识别过程

3.4 智能体的消亡

完成搜寻的智能体现在拥有完整的信息,这时候它离开决策树,将自己的信息加入到笔序信息池中,然后自动销毁。

笔序信息池收到智能体的信息后分两种情况处理。如果信息表示这是一种新的笔序的画法,则建立一条新记录,包括模板 ID、笔序和使用次数(新建记录使用次数应为1)。如果该信息表示这种笔序的画法已经存在,则给相应的记录中使用次数加一。

当缓冲区满了之后,将其中的信息加入到用户模型中,同时根据需要,对搜索决策数进行管理,这部分内容在第四节有所论述。

4 用户信息模型管理和信息树管理

4.1 用户模型信息管理(评价机制)

每次检索结束后,智能体将收集的数据加入到用户信息模型中。随着使用次数的增加,用户模型中的数据必然越来越多;相应的,决策查找树也越来越复杂。而在这棵树中,可能有些分支被使用到的次数是很少的,这时候就需要对用户模型进行整理,在笔序信息池中删除那些使用很少的笔序画法,然后根据笔序信息池对树进行剪枝,使两者信息保持一致。

因此,决策树的调整实际上是根据线性表的调整进行的。这是一个对用户输入方法评价机制的问题。本文的评价机制是由线程调度机制中的彩票调度法改进而来。

考虑模板 A 的两种画法 a, b 。当对缓冲区进行处理时,如果有 n 个 a , m 个 b 则给 a $10 \times n$ 张彩票,给 b $10 \times m$ 张彩票。下一次对缓冲区进行处理时如果缓冲区中存在模板 A 的某种画法 a, b 或者其他,出现次数为 c ,则 a, b 分别消耗掉 n, m 张彩票,同时给相应的画法 $10 \times c$ 张彩票。

在用户信息库中,模板 A 的各种画法将按所拥有的彩票数排序。如果出现新的画法,并且彩票数大于信息库中的画法,则将用户库中彩票数最少的画法淘汰。实现方法是给每种画法一个记录出现次数的数组 $[a_0, a_1, \dots, a_9]$, 初始化为 0。第一次出现该画法是出现次数付给 a_0 ,再出现模板 A 时将 $a_1 \dots a_9$ 前移一位 ($a_i = a_{i+1}$), 令 a_0 为新的出现次数。该种画法的总彩票数可以表示为 $\sum a_i \times (i+1)$ 。

4.2 决策查找树的剪枝

在根据评价机制对线性表进行整理后,线性表中的用户信息已经得到了调整和精简。决策树的剪枝只要根据线性表的内容进行,使决策查找树的内容与线性表保持一致。

5 实验及结果分析

我们所提出的解决方案的成败取决于两个关键因素:一是我们的设想“用户多次绘制一个图形后绘制方法迅速趋向收敛”是否正确;二是在设想正确的情况下,我们是否能够利用这一点来很好地构建用户模型,获得预想的识别效率。

为此,我们选择了五个草图作为实验的输入部分,如图3所示。在这五个图形中,包括了直线和圆弧两种基本图元(椭圆与圆弧类似)。包含了图元之间的相邻, T 型相交, 十字相交, 平行以及无关这所有的五种关系。同时这五个图形的笔划数从6笔到14笔不等。所以这五个图形具有相当好的代表性。

考虑到识别的准确率和一定的识别速度,我们选择 SRG 方法作为当使用用户模型识别不成功时所用的识别方法。并且将它作为用户模型识别方法的一个横向比较。

根据收集的300个草图运行的结果显示(如图4所示):在用户信息初始为空的条件,有31次使用 SRG 算法匹配,269次使用笔序法匹配。特别是最后100条,只使用了1次 SRG 法,

命中率达到99%。这说明我们的假设是成立的,即一个图形绘制人员多次绘制同一个图形的笔序趋向于一定。

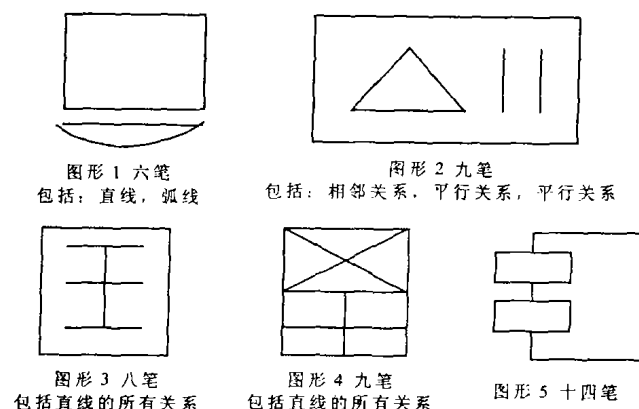


图3 实验所用图形示例

从图4中可以看出最初用户模型中信息为空的时候,命中率比较低,前50个只有30个使用笔序法识别,另有20种绘制方法还没有被用户模型所记录。但是随着用户的使用,用户模型中记录的笔序信息越来越多,命中率也随之提高。最后50个图形只有一个是使用 SRG 方法识别的。

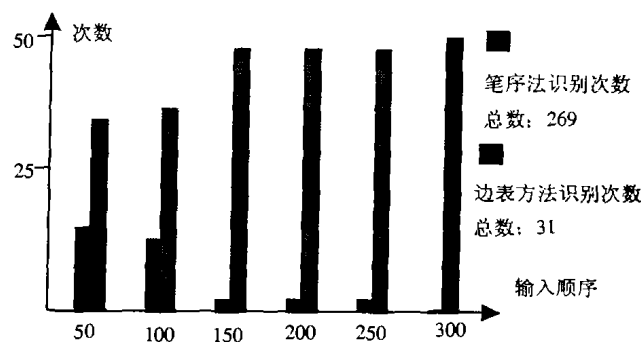


图4 两种方法使用次数比较

本方案的目的,就是通过对用户建模使得在草图识别中可以充分利用用户绘图时的笔序信息,从而加快草图识别的速度。实验结果如表1所示。从表中可以看出:单独一个图形的识别时间用笔序法仅为边表法的1/20。特别地,在边表方法中单个图形的识别时间随着图形笔划数的增加而显著地加长;而笔序法的识别时间几乎不受笔划数的影响。这主要是因为,用笔序法识别时在与模板进行匹配时只需要做相应基本图元类别的比较。匹配的时间相对于输入草图的关系抽取时间来说非常短,所以笔划数的增多并不能对识别时间产生大的影响。

表1 三种表示方法识别时间比较

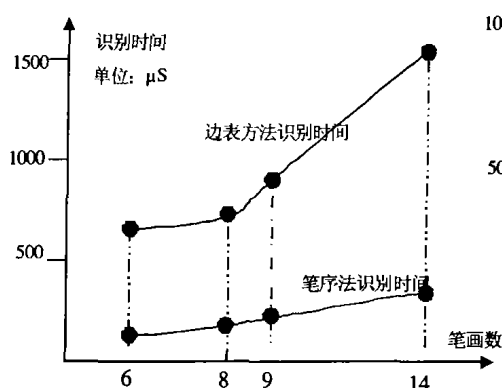
笔划数	笔序方法(ms)	边表方法(ms)	SRG 方法(ms)
6	0.02	0.58	111
8	0.03	0.59	105
9	0.03	0.73	102
14	0.04	1.55	很大

图5(a)是两种方法的识别时间比较。可以看出,相同笔画数图形的识别时间笔序法要比边表法低得多;并且,边表法的识别时间曲线随着笔画数的增加而急剧上升,而笔序法的识别时间曲线则比较平稳。

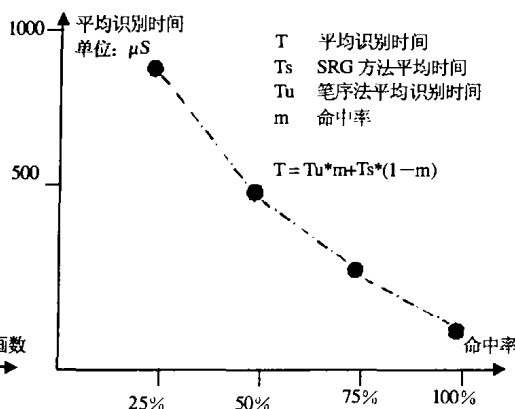
图5(b)是平均识别时间与命中率的关系图。该图清楚地

说明了平均识别时间随命中率的上升而迅速降低。当命中率达到一定的高度时,就能保证平均的识别时间远低于边表法

和 SRG 法的识别时间。



(a) 两种方法的识别时间比较



(b) 平均识别时间与命中率的关系

图5 识别时间比较

小结 本文通过捕捉用户的绘制习惯建立起用户模型,并且根据用户模型中的用户信息创建了决策查找树,实验证明,利用决策查找树进行图形识别大大提高了识别效率,达到了快速高效准确的草图识别的目的。本文方案的特点主要在于三个方面:

·在线训练。本方法虽然需收集用户信息,对用户建立模型,但是收集信息的过程是在用户的使用过程中完成的。即实现了训练的透明性。这一点,对于用户来说是极其方便自然的。

·旋转无关性。由于本方法将图形拆分成直线、弧线和椭圆三种基本图元,利用其绘制时间和空间关系来表示图形。这三种基本图元及其关系都是旋转无关的。从而实现了识别方法的旋转无关性。

·模板库的不敏感性。本方法以决策树的形式组织用户模型,不需要对模板库进行线性的搜索,所以识别所费时间不会随模板数的增加而增加,这大大提高了识别的效率。

在用户模型管理的评价机制中,本文仅采用了最简单的

基于频率统计的方法,这可能使模型不能准确地代表用户的习惯。在我们下一步的工作中,将采用机器学习方法的评价机制,以求达到更好的效果。

参考文献

- 1 Revankar S, Yegnanarayana B. Machine Recognition and Correction of Freehand Geometric Line Sketches. In: Proc. of IEEE Intl. Conf. on Systems, Man, and Cybernetics, 1991, 1: 87~92
- 2 孙正兴,徐晓刚,孙建勇,金翔宇. 支持方案设计的图形输入工具. 计算机辅助设计与图形学报, 2003, 15(9): 1145~1152
- 3 Xu X G, Sun Z X, Liu W Y. Matching Spatial Relation Graphs Using a Constrained Partial Permutation Strategy. Journal of Southeast University, 2003, 19(3): 236~239
- 4 Takayuki D K, Ajay A, Van V. Recognizing Multi-stroke Geometric Shapes: An Experimental Evaluation. In: Proc. of 6th Annual ACM Symposium on User Interface Software and Technology, 1993. 121~128
- 5 Fonseca M J, Jorge J A. Using Fuzzy Logic to Recognize Geometric Shapes Interactively. In: Proc. 9th IEEE Conf. on Fuzzy Systems, 2000, 1: 291~296
- 6 Liu W, Qian W, Xiao R, Jin X. Smart Sketchpad-An On-line Graphics Recognition System. In: Proc. of ICDAR2001, Seattle, Sep. 2001

(上接第197页)

已在医院系统投入应用,其效果良好。

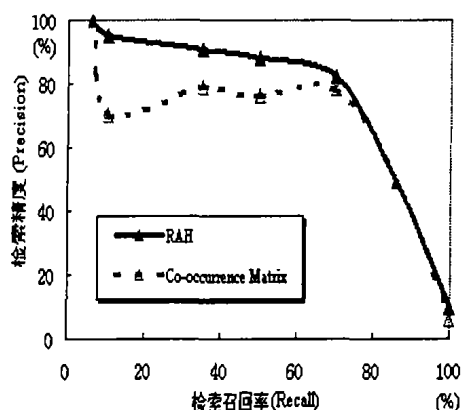


图4 二种纹理特征检索算法

参考文献

- 1 杨育彬. 基于内容的图像检索关键技术及其应用研究. 南京大学计

算机科学与技术系:[博士论文]. 2003(6)

- 2 Chang T, Kuo J. Texture analysis and classification with tree-structured wavelet transform. IEEE Transactions on Image Processing, 1993, 2(2): 429~441
- 3 Ma W Y, Manjunath B S. Texture features and learning similarity. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1996. 425~430
- 4 Jain A K, Farrokhnia F. Unsupervised texture segmentation using Gabor filters. Pattern Recognition, 1991, 24: 1167~1186
- 5 Wang J Z, Wiederhold G, Firschein O, Sha X W. Content-based image indexing and searching using Daubechies' wavelets. International Journal of Digital Libraries, 1998, 1(4): 311~328
- 6 Ma W Y, Zhang H J. Benchmarking of image features for content-based retrieval. In: Proc. 32nd Asilomar Conf. on Signals, Systems and Computers, California, US, Jan. 1998. 253~257
- 7 Augusteijn M, Clemens L E, Shaw K A. Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. IEEE Transactions on Geoscience and Remote Sensing, 1995, 33(5): 616~626