草图搜索的魅力与挑战

王长虎 张 磊
微软亚洲研究院

关键词：草图搜索

什么是草图搜索

小时候，有一部叫《神笔马良》的儿童电影，让人印象深刻。电影中，马良有一只神笔，用这只笔在墙上画出的任何图画，不论是金元宝还是大船，都会变成真实的物体，令人羡慕不已。所画，即所得，便是我们这一代人儿时的梦想。

如今大数据时代的草图搜索就是通向所画即所得的一个尝试：通过手绘的线条图在海量图片中找到与之形状相似的图像。

在触摸屏设备（智能手机、平板电脑）逐渐普及的今天，可以预见，草图搜索技术具有广泛的应用前景。它可以帮助任何年龄段的人，可以帮助儿童认识世界，可以帮助少男少女方便地找到带有特定纹饰的体恤衫和花裙子，可以帮助设计师找到理想的图像素材。

草图搜索的魅力与难点

目前已商用的图像搜索方法有基于关键字的图像搜索和以图找图的搜索，而草图搜索技术仍然处于初步的研究阶段。草图搜索与基于关键字的图像搜索和以图找图搜索的最大区别在于，使用充满不确定性和创造力的线条作为检索入口，这既是研究草图搜索的最大难题，也是其魅力所在。

线条，自古以来便是人与自然沟通的一种重要方式。在旧石器晚期，人类便在法国肖维岩洞[1]中留下了逼真的犀牛的线条图。带着人类对自然的赞美与敬畏，逼真的线条渐渐发生了变化：从甲骨文到草书，从工笔画到写意画，从达芬奇的蒙娜丽莎到毕加索的格尔尼卡，丰富的创造力把线条推到了艺术的高度，也为草图搜索带来了巨大的挑战。科学，在很多时候是对事物和事实背后规律的揭示，因此某种程度上说是“异中求同”；艺术，作为一种传播和交流媒介，虽然在表达相同概念的时候要保留共性，但是作为一种创作，追求的往往是超越普通的一步，所以某种程度上说是“同中求异”。搜索属于科学范畴，然而线条画，却充满了创造力，属于艺术范畴。近几年，我们在国内外很多场合，向学术界、工业界的朋友和学生展示过我们的MindFinder草图搜索技术，并让大家试着自己画线条来寻找图片。每个人画的线条都各不相同，各自散发出独特的魅力和趣味。一百个人画飞机，将画出一百个不同的形状。

人类可以很容易地认出这一百个不同形状的飞机，并且只要有足够的时间和精力，便可以在数据库中找到飞机的图片。当然，在大数据时代，面对数以百万千万张的图片，人类无法逐

[1] 肖维岩洞（法语：Grotte Chauvet），位于法国南部阿尔代什省的一个洞穴，因洞壁上拥有丰富的史前绘画而闻名。1994年发现。部分历史学家认为洞内岩画可以追溯至32000年前。这些画由红赭石和黑色颜料绘制，绘画的主题有马、犀牛、狮子、水牛、猛犸象或是打猎归来的人类。

一进行比对。草图搜索就是让计算机在大规模数据库中自动、实时地找到形状近似的图片。然而，计算机对形状的认知和匹配与人脑存在巨大的差距。人脑，经过几十亿年生物进化，可以神奇地自然而然地认识物体；而计算机却只能是被动的，需要研究人员设定每一个步骤，来“教”计算机一步步地辨认目标。

计算机在草图搜索任务中主要面临特征表示、特征匹配和建立索引结构三个方面的难题。

特征表示 把手绘的线条图和数据库中的图像都转换为计算机能够“认识”的表示，即一组或若干组数字，这便是特征表示。我们需要找到有效的特征表示，使得同一类的物体尽可能有相似的特征，而不同类别物体的特征尽可能不同。

特征匹配 有了有效的特征表示还不够，我们需要根据特征表示方法，定义合理的度量来匹配所画线条图与数据库中图像的特征，从而计算二者的相似度。基于这个相似度，我们就可以把与手绘线条最相似的图像排在前面，并返回给用户。特征匹配方式是受特征表示方式制约的，不同的特征表示方式可能需要不同的特征匹配方式。同样，我们要找到有效的特征匹配方式，使得包含线条图所表示物体的图片尽可能排在前面。

建立索引结构 在数据库中的数据以千计算时，我们可以一张张地比较得到相似度。但是，当数据规模上升到百万千万时，这样做就比较慢了。怎样才能建立有效的索引结构，使系统在极短时间内返回搜索到的结果，便成了一个重要问题。

草图搜索，这个对于人类来说看似简单的事情，却成了研究人员几十年来孜孜不倦追求的目标。

早期的草图搜索

自20世纪70年代以来，性能的提高使计算机足以处理图像时（当时的大规模数据），从二维图像中提取一维线条的技术便快速发展起来。

从二维图像中提取线条是计算机视觉中的基本问题，也称为边缘检测。图像中线条的存在是由物体的轮廓或自身的纹理而导致的像素亮度差异引起的。这些亮度上的差异，可以通过分析每个像素周围的相邻像素，从信号检测的角度判断该像素是否是边缘像素。这一问题看似简单，实际上蕴含着非常大的困难。首先，自然图像中有很多噪声，简单的算法很难从局部的信号本身区分出哪些像素是真正的边缘，哪些像素是噪声；其次，边缘并不等同于真正物体的轮廓，仅仅依靠局部的信息来判断远远不够。

边缘检测几乎是计算机视觉领域中最早开始研究的问题。20世纪60年代末，埃文·索贝（Irwin Sobel）就提出了索贝尔算子（Sobel operator），从数字滤波的角度检测图像边缘；80年代初，戴维·马尔（David Marr）更为系统地从信号处理的角度分析和研究了这一问题，奠定了边缘检测的理论基础；约翰·凯尼

图1 快速多分辨率图像查询系统界面[1]

（John Canny）在1986年提出了凯尼边缘检测算法。凯尼算法利用戴维·马尔提出的边缘检测理论，着重解决了宽边缘定位的唯一性问题和边缘的连通性问题。因为该算法的效率高和检测结果的鲁棒性强，所以直到今天还被广为采用。尽管如此，凯尼算法本质上仍然是基于图像局部信息进行边缘检测，无法提取真正意义上的物体轮廓，因此，我们称之为边缘检测算法，而不是轮廓提取算法。

边缘检测问题的进展，使得研究人员得以考虑用线条画来检索图像。早在20世纪80年代初，美国普渡大学的张（Ning-San Chang）等人就开始了这方面的尝试，即query by pictorial example（通过图画检索）。后来美国匹兹堡大学的张系国（Shi-Kuo Chang）等也从不同的角度做了类似的尝试。早期的研究中，由于图像处理困难且自然图像稀缺，系统中多是类似于工程图纸、卫星拍摄的地面布局之类的图像。对于这类图像，边缘检测的问题还不算太大，而且，受早期输入设备的限制，查询图像也多是简单几何形状的组合，比如三角形、矩形等。这些因素都极大地限制了草图搜索的应用。

20世纪90年代初，随着数字图像数量的急剧增加，基于内容的图像检索应运而生。处理更大的图像库和支持更灵活的查询输入方式成为人们关注的研究热点。当时有两个值得一提的工作：一个是1995年美国华盛顿大学开发的快速多分辨率图像查询系统[1]，另一个是1996年美国哥伦比亚大学开发的VisualSEEk系统[2]。

文献[1]的系统的核心是在多尺度小波变换基础上，对图像的颜色布局进行有效表示，并将其用于高效率的检索。该系统的特点是可以让用户在一个画板上用彩色的“粗”线条来表达想要查找的图像，系统可以快速地从2万幅图像中返回匹配结果。即使在今天看来，这个系统也非常吸引人。然而，它不是基于线条图的查询系统，所支持的输入实际上是比较粗的彩色线条块，即用户输入的是一种颜色布局图，这种查询表示无法区分更精细的线条细节。

Visual SEEk系统综合了哥伦比亚大学张（Shih-Fu Chang）教授研究小组多年的成果。该系统允许用户在画板上用多个任意形状的彩色区域来表示查询意图，并主要考虑区域的颜色和区域之间的位置关系，在12000幅图像中快速查找相似的图像。尽管该系统的研究取得了很大进展，然而它仍然不是基于线条图的检索。它依赖的是区域的颜色和区域之间的相关位置关系，用户很难用这样的方式来表达用线条画出来的物体，如“汽车”、“长颈鹿”和“电脑”等。

草图搜索概念的提出始于20世纪80年代，但到了90年代进展仍然缓慢。

草图搜索的现状与挑战

2000年以后，伴随着互联网的发展，上传的图像呈现爆炸式的增长，人们对图像检索也有更大的需求。智能手机和平板电脑的普及，让人们更加关注草图搜索。

2009年，清华大学的研究人员在ACM SIGGRAPH Asia上发表了他们最新的研究成果Sketch-2Photo[3]，引起了学术界的关注。Sketch2Photo可以帮助用户通过勾画线条和添加关键字来合成图像。用户在一个画板上画出想要的任何图画，例如在夕阳西下的海边，有一对相拥的恋人，

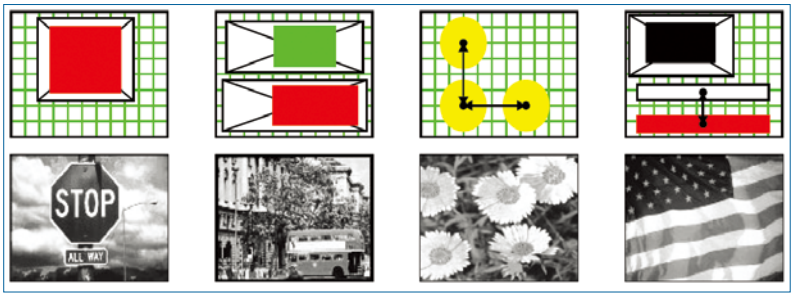

图2 Visual SEEk系统中用户所绘的查询图与目标图像举例[2]（第一行是4种不同的查询图，第二行是对应的想要查找的目标图像）

图3 Sketch2Photo图像合成系统示例[3]

远处的海面上有帆船几只、海鸥点点。系统会根据用户所画图画的各个元素，逐一在网上搜索图像并提取出真正匹配的物体区域，再把提取出来的区域，通过图像拼接的方法拼成一幅“天衣无缝”的新画面返回给用户，如图3所示。这个研究工作的应用前景非常吸引人，可是在现实中还有很多困难无法克服。例如，缺乏有效的基于线条图来搜索图像的方法，这个方法要求用户对画面中的每个元素提供一个关键词描述，系统先用关键词在网上搜索到相关图像，然后再把用户画的轮廓和返回的图像逐一比较，找到轮廓比较吻合的图像。这个过程非常耗时，往往需要数分钟的时间才能完成整个搜索。另外，要做到将图像真正“天衣无缝”地合成也并非易事。因为这些因素，这个工作离真正实用还需要更多的时间。尽管如此，Sketch2Photo的出现再次引起了研究人员对草图搜索的关注。

2009年，微软亚洲研究院也开始研究草图搜索技术，希望建立一个不依赖于关键字的、能在数百万张图片数据库上进行实时检索的草图搜索引擎。我们把这个项目以及相关的草图搜索引擎称为MindFinder[4,5]。

由于早期的草图搜索技术处理的图像数量和几何形状类型都有限，工作大多集中在特征表示与特征匹配上，我们开始尝试为大规模草图搜索构建一个有效的索引结构。

倒排索引是网页搜索、基于关键字图像搜索等大规模搜索技术中必不可少的。就像查字典一样，根据拼音或者偏旁部首的索引表，便可以快速地查询。数据规模的扩大不仅带来了建立索引结构的问题，还使得类别数大幅增加、每类数据的多样性急剧扩大、类间图像的相似性提高，这些都可能让小规模数据库的特征表示与特征匹配方法无法直接应用于海量数据。对于大规模草图搜索来说，特征表示、特征匹配与建立索引结构，是紧密结合在一起的。匹配算法取决于提取的特征类型，简单的匹配算法很难在大规模数据上找到合理的匹配，而复杂的匹配算法又会导致无法建立倒排索引。因此，需要通盘考虑特征表示、特征匹配与建立索引结构，以便提供一个完整的解决方案。

为了保证匹配的准确性，我们采取了最“笨”的方法——直

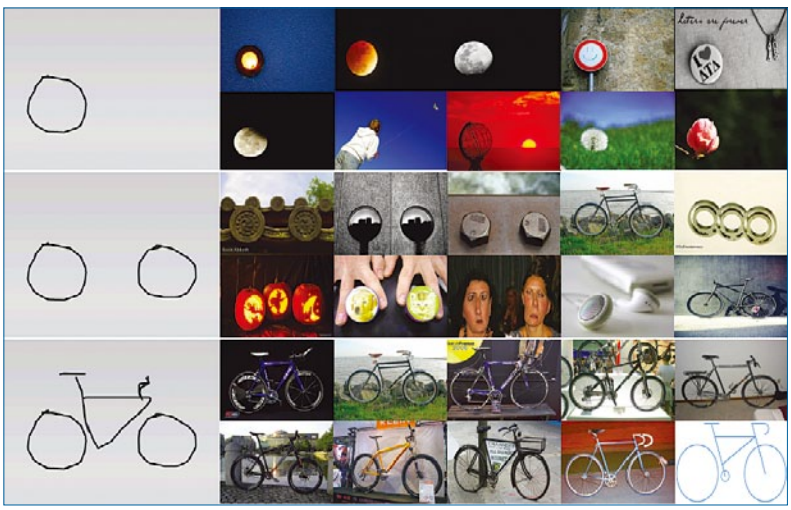

图4 MindFinder搜索引擎返回的前10个结果[5]

接用线条本身（即组成线条的每个像素）来表示线条，并且采用了匹配原始线条的有效方法——Chamfer匹配法来计算线条图的相似度。然而，采用这种方法从1000万幅图像中进行草图搜索，每检索一次，需要超过100台普通服务器的内存以及1个小时的时间。因此，我们提出了一种叫做Edgel Index的索引结构，并且改进了特征表示和Chamfer匹配法来适应这个索引结构。最终使得MindFinder系统索引1000万张图片仅需十几GB的内存，检索时间只需几百毫秒。此外，MindFinder检索效果也是当时最好的，如图4所示。近几年，我们在国际计算机视觉与多媒体会议上陆续发表了多篇论文，并展示了相关系统。在2010年ACM多媒体国际会议上我们的系统还获得了最佳演示奖。

然而，MindFinder系统依然面临很多困难，首先是对于仿射变换的鲁棒性。该系统虽然在局部范围内允许平移、放缩以及局部形变，但由于自然图像的复杂性，仍然无法找到物体位置和大小相差很大的图片。一种解决方法是对数据库中的每张图像都找到主要物体，然后针对物体而不是图像本身建立索引结构，这样便保证了检索的平移和缩放不变性。我们依此方法建立了一个卡通图像草图搜索引擎和一个产品图像草图搜索引擎。另外，如何设计有效的解决方案，使得系统能检索更多的图片，比如几十亿张图片，也是一个重要难题。

图5　基于人脸位置和属性的图像搜索[7]

草图搜索技术现在仍处在研究的初级阶段，要想达到商用，依然还有很长的路要走。然而，我们惊喜地发现，近几年，越来越多的大学和科研机构开始重视草图搜索的研究，包括与我们长期合作的上海交通大学的类人脑计算实验室、德国柏林技术大学（Technische Universität Berlin）以及中国台湾大学的研究人员。

2011年，德国柏林技术大学的研究人员尝试把以图找图中常用的词袋模型（bag of words）应用到草图搜索中[6]，即在图像的主要轮廓上均匀地采样，提取若干特征点，然后将每个特征点的邻域生成边缘方向直方图作为该点的边缘描述子。这样，一幅图像就可表示成一个词袋模型，从而利用基于倒排索引的方法进行索引。这种表示和索引方法可以在上百万的图库中进行快速查询。但是，简单的词袋模型完全丢掉了边缘像素在图像中的位置信息，而这一信息在准确的检索中是不可缺少的，因此这种方法很难获得准确的检索结果。

另一个值得一提的是中国台湾大学的研究人员于2011年在ACM多媒体会议（ACM Multimedia）上发表的论文[7]。他们采用了类似画草图的用户输入方式，目的是用来检索相册中的人脸图像。用户画的草图实际上是用一个圆形的图案表示一个人脸在图像中的位置，用户可以指定多个人脸，也可以指定每个人脸的一些属性，比如性别、是否戴帽子、肤色等，如图5所示。看似很简单的圆形图案，在这个工作中发挥出“神奇”的作用，尤其是在触摸屏设备上，给用户提供了方便的交互方式。基于草图的图像检索，无论是研究方面还是应用方面都有着很大的发展空间，十分值得我们进一步探索和研究。

相关的研究与应用

从线条到卡通画 孩子们与生俱来的乐趣和本领之一，就是在墙上和书本上涂涂抹抹。2011年微软亚洲研究院的家庭日，公司的儿童游乐区立起了两个大的触摸屏，让孩子们来体验一个叫做Sketch2Cartoon的系统，这是我们基于MindFinder系统研制的。这个应用是让儿童们通过画线条来创作卡通画，并分享给家长们。每画一个形象，计算机便自动推荐一些彩色的卡通形象，如果不喜欢还可以方便地更换卡通形象以及更改各种属性。孩子们玩得很开心。

草图搜索是基于线条（sketch-based）研究的基础。触摸屏时代的到来使得我们已经看到并将看到越来越多的基于线条的研究和应用。

从二维到三维 线条可以帮助我们与计算机世界沟通，也可以把我们带到计算机的虚拟世界中。在三维动画电影中，我们可以看到各式各样的三维玩具、汽车、动物以及人，十分逼真。虽然我们在现实世界中无法成为神笔马良，但在虚拟世界中可以画出心中所想。

随着互联网中人造三维物体数目的不断增加，画线条搜索三维物体的研究渐渐发展起来。三维物体搜索与图像搜索各具特点：人造的三维物体线条更干净，更清楚，而自然图像中的线条则可能淹没在复杂的背景中。从另一个角度来看，三维物体与一维线条进行匹配，比与图像上的线条匹配要难，毕竟物体可能有多个视角形成多个二维图像。

当前，三维物体的数量还比较少，最大的数据集规模也不过数以千计。因此，从千级规模跨越到百万千万数量级，还需要克服很多技术难关。

从检索到识别 2012年2月，你画我猜（Draw Something）游戏来袭，发售6周下载量突破了3500万次。这款游戏的内容就是你来画线条我来猜单词，可见线条的魔力。

随着触摸屏的普及，你画我猜的流行——线条，从肖维岩洞中复活了。“我，是谁？”

2011年，德国柏林技术大学马克·阿列克夏（Marc Alexa）教授的研究组和微软亚洲研究院的团队分别开始研究，并分别在2012年发表了相关论文[8,9]。文献[8]把草图识别问题转化成了分类问题，通过训练线条图分类器的方式来归类和识别线条图。文献[9]从互联网上收集了上百万张与线条图有着相似性的卡通图，基于MindFinder草图搜索技术，在搜索结果中建立概率图模型，从而识别所画的线条。

结语

绘画，属于艺术范畴；搜索，属于科学范畴。草图搜索，是在用有限的科学工具来探索无限的创造力空间。草图搜索面临的挑战来自创造力，而这也正是草图搜索的魅力所在。■

王长虎
微软亚洲研究院副研究员。主要研究方向为新一代多媒体搜索、草图搜索与理解、线条与形状特征的分析与理解等。
chw@microsoft.com

张 磊
微软亚洲研究院高级研究员。主要研究方向为多媒体检索、内容分析、计算机视觉和模式识别。
leizhang@microsoft.com

参考文献

[1] C. E. Jacobs, A. Finkelstein, and D. H. Salesin. Fast multiresolution image querying. In ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, Los Angeles, USA, 1995: 277~286

[2] J. R. Smith and S.-F. Chang. VisualSEEk: a fully automated content-based image query system. Proceedings of the Fourth ACM International Conference on Multimedia, 1996: 87~98, Boston, USA

[3] T. Chen, M. Cheng, P. Tan, A. Shamir, and S. Hu. Sketch2Photo: internet image montage. ACM Transactions on Graphics, 2009, Volume 28, Issue 5, Article No. 124: 1~10

[4] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang. MindFinder: interactive sketch-based image search on millions

of images. Proceedings of the 18th ACM International Conference on Multimedia, Florence, Italy, 2010: 1605~1608

[5] Y. Cao, C. Wang, L. Zhang, and L. Zhang. Edgel index for large-scale sketch-based image search. In Proc. The 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011: 761~768

[6] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa.Sketch-based image retrieval: Benchmark and bag-of-featuresdescriptors. IEEE Trans. Vis. Comput. Graph, 2011

[7] Y. Lei, Y. Chen, B. Chen, H. Su, L. Iida, W. Hsu. Photo Search by Face Positions and Facial Attributes on Touch Devices. ACM Multimedia, Scottsdale, October 2011

[8] M. Eitz, J. Hays, M. Alexa. How do humans sketch object?ACM Transactions on Graphics, Volume 31 Issue 4, Article No. 44, July 2012

[9] Z. Sun, C. Wang, L. Zhang, and L. Zhang.Query-adaptive shape topic mining for hand-drawn sketch recognition. Proceedings of the 20th ACM International Conference on Multimedia, Nara, Japan, 2012