



柳井 啓司
Keiji YANAI

研究课题

从Web自动获取视觉概念以及Web图像挖掘、普通图像·映像识别与语言概念之间的相互对应、位置信息和图像的语义关系的分析及利用

关键词

图像识别, 图像理解, 物体识别, 机器学习, 信息检索

所属专业	研究生院信息理工学研究科 综合信息学专业
研究成员	柳井 啓司 副教授
所属学会	信息处理学会, 电子信息通信学会, 人工智能学会, IEEE Computer society, ACM
研究设备	计算专用的计算机 80 台, 数据存储约 100TB

研究概要

与语言相关的普通图像·映像的识别

由于互联网在全世界得以普及,因而出现了图像及映像共享的服务,这项服务被众多用户所采用。该研究室开展各种研究,比如使用上传到网站的图像或映像,挖掘不限定对象的普通图像·映像,使其能够应用到各种用途中。

首先,在图像识别方面,该研究室研究语言与图像的关联性。这项技术是推断语言所对应的图像以及图像所对应的语言。当然用GOOGLE也能检索图像,但这到底不过只是检索图像附带的标签(文本)信息而已。

带名词、形容词的图像特点

该研究室从事的研究是理解基于语言的图像特点,观看图像并判断其特征。为了实现这个目标,一开始就要从网页大量自动收集名词所对应的图像,并分析其特征。具体来说,就是从图像搜索引擎中检索「拉面」这个名词,会收集到大量的图像。再挖掘这些图像,就能知道它们共同的特点是是大碗沿和叉烧肉。像这样他们对大量名词进行相同的操作,收集名词对应的图像特点。

接着再研究黄色、黑色、坚硬、柔软这类形容词。使用他们规定的熵,对6000种名词进行排序,将各自的特征收集到了对应的形容词里。

食物识别系统

作为使用这两个识别方法的系统,该研究室和企业共同研究开发出了「食物识别系统」。这个识别系统是通过挖掘从食物名称收集到的图像来获得具有特征的局部图样以及采用独特的比率混合来实现形容词的色彩识别。如果使用这个方法,就能从图像中识别85种食物,识别率达到60%以上。这个识别率在国内可以说是达到最高水平。使用这项技术有各种应用方法,比如用户仅用手机连续拍摄食物图片,就能显示1天平均摄入的热量以及1周之内摄入食物的趋势等。

添加了位置信息的图像

该研究室在最近开始研究的课题中,还有在单词里添加了位置信息的图像识别。在Flickr等可上传图像的网站,除了图像,还能获取图像所在位置的信息。在自动收集网页图像时,还会收集图像所在位置的信息进行比较。比如输入「noodle」进行检索,就可以知道这个单词在日本大都指拉面,而在欧洲多是指意大利面的这种世界各地不同的倾向及差异。他们今后还会尝试检索各种不同的单词,以此来推断不同国家的特点。

映像及视频

该研究室还从事映像相关的各种研究。使用地面波数字广播的附加信息(文本)只提取所需部分

或者从Youtube等视频网站收集大量的图像,研究如何用映像也能进行上述「带名词、形容词的图像特点」所阐述的作业。

他们的映像分析成果在TRECVID(国际映像处理研讨会)竞赛上,比赛从指定的映像中自动提取目标映像的算法在日本国内获得了第二名。

优势

研究更接近于人类思考的图像·映像识别

该研究室研究的图像识别始终主要着眼于如何接近人眼观看的感觉。如今数码相机里大量搭载的面部识别功能是把识别对象限定在面部上,因此实现起来比较简单。但是他们研究的是以照片里的所有图像内容为对象。比如普通的肖像画中,除了人的面部以外,还会同时拍入背景中的风景、建筑物、桌子上的菜肴、脚边的小狗等各种景象。可以说像这样不限定对象就能进行识别是他们的一大优势。

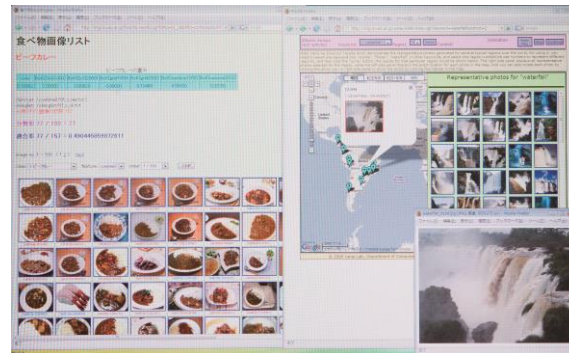
Web图像挖掘专用机

还有为了自动收集Web图像,挖掘收集到的大量Web图像,他们准备了60台计算专用的计算机(320核心)和约100TB的存储器,采用大规模并行分散计算进行图像·映像处理。

未来展望

加强普通单词所对应的图像识别

有关该研究室目前正在研究的语言和图像的识别,他们希望能够对应更多的单词。因为语言非常广泛,如果能够识别普通单词所对应的图像,那么就能扩大它的应用范围。另外食物识别所采用的方法由于图像识别的精度很高,因此使用其算法,就能得到有效的应用。还有除了单词和图像的关系外,有关加入了位置信息的物体,要进一步收集大量的样品,希望把它变成更有用之物。比如取「城」castle中的一个单词,日本人和美国人能够想到的图像完全不一样。像这样根据语言的不同,分析、理解被识别图像的差异,就能推断出不同国家或地区的特点,因此如果有效利用这项技术,就能应用到各种领域中。



(左) 食物识别系统, (右) 附带位置信息的图像挖掘



计算机群系统



大容量存储器