

高武平, 刘芳, 姚新强, 杨绪连, 杨菲, 付占岭, 2018. 利用互联网数据建设城市房屋基础信息数据库的初步研究. 震灾防御技术, 13 (1): 237—244. doi: 10.11899/zzfy20180122

利用互联网数据建设城市房屋基础 信息数据库的初步研究¹

高武平 刘芳 姚新强 杨绪连 杨菲 付占岭

(天津市地震局, 天津 300201)

摘要 本文介绍了一种利用互联网数据建设更新城市房屋基础信息数据库的简便方法, 并通过实例对实现的基本步骤进行了介绍。其基本步骤为: 采集互联网城市房屋信息数据并与地图数据有效融合, 可快速地获得城市住宅类房屋的基础数据, 将互联网街景地图、遥感卫星图片与专家经验识别相结合, 可有效提取大多数公共事业类房屋基础数据, 再辅以必要的现场调查补充核实数据, 初步建立城市房屋基础信息数据库; 采集目标区不同期次的互联网地图进行叠加对比, 定位地图中的显著变化区, 开展针对性的补充调查, 及时更新数据库。通过该方法可高效快捷地建立城市房屋基础信息数据库, 并能够实现数据库的持续更新, 这为其他城市房屋基础信息数据库的建设提供了一种有益的参考。需要说明的是, 数据采集受到互联网数据本身准确度、可靠度及完整度的客观制约, 由此建立的数据库其精度受到一定限制。

关键词: 城市房屋信息数据库 专家经验识别 数据库更新

引言

城市房屋基础信息数据库是城市地震抗震能力评估、地震应急救援指挥决策、震害预测等工作的基础性数据库, 其可靠性、完整性对决策、评估的合理性与科学性具有重大影响, 获得准确的房屋数据, 对于震前御灾、震中救灾和震后重建都具有重要意义(韩贞辉, 2013)。建立符合实际情况的城市房屋基础信息数据库是防灾减灾工作的基础性工作。

一般地, 采用现场逐栋调查的方式可获得真实可靠、信息完整的房屋基础性数据库, 但在大城市、特大城市利用此类方法所耗费的人力、物力都十分巨大, 而且建设周期比较长; 同时也面临一个问题, 即数据库的更新如果还采取相同的方式, 重复投入对人力物力又是一种浪费。目前, 在城市基础信息数据库建设方面已经有所积累, 可通过部门合作或从测绘公司购置, 但同样存在这样一些不足: ①由于涉及信息的安全性与敏感性, 数据交换获得的数据库可能不完整; ②由于数据库建设主要是满足建设方的自身需要, 通过交换获得的数据库

1 基金项目 国家自然科学基金(项目批准号: 41772123)

[收稿日期] 2017-06-14

[作者简介] 高武平, 男, 生于1981年。高级工程师。主要研究方向: 城市活动断层探测、城市地震灾害风险评估、地震动数值模拟等。E-mail: gwpp123@126.com

可能难以满足实际研究需要；③通过合作或购置获得的数据库通常都会存在更新困难的问题（聂高众等，2002；韩贞辉，2013），根据报道（郑树平等，2015；段锋等，2010；王丽莉等，2008），一些已经建成的数据库其更新需要投入的时间与精力同样惊人，而如果不及时更新，又会制约实际工作中的应用。房屋基础性数据库的不完整、不能满足需要及更新不及时已经对防灾减灾工作产生了很大制约，建立相对完整并能够及时更新的城市房屋信息数据显得日益紧迫。

随着互联网的不断发展，遥感影像在城市建筑普查中得到了应用（哈丽娜等，2010；王斌等，2014），取得了很好的效果。随着大数据时代的迅速到来，百度、高德等互联网地图（包括遥感影像）更新越来越快、精度越来越高，街景地图也日益成熟，通过其可从多角度不同方位对建筑结构及其细部进行展示，而且互联网上还产生了越来越多的有关房屋基础性信息的数据，如果能从中提取有效信息，不仅能够为建立城市房屋基础信息数据库提供新数据源，节省大量经费与人力，而且也可据此对已有的数据库进行完善补充，甚至可以采用适当的方法实现随互联网地图更新同步更新数据库，故利用互联网数据建设城市房屋基础信息数据库不失为一种方便、快捷、经济的方法。

本文结合在天津市和平区开展房屋基础调查工作的实例，介绍从互联网提取基础数据建设城市房屋基础信息数据库的新思路及主要步骤，以期为其他地区开展类似工作提供参考。

1 基本技术思路与主要步骤

1.1 基本技术思路

本文采用了互联网结合专家调查的方式开展工作。其基本思想是：一方面充分利用互联网数据的便捷性与多元性，提取有效信息；另一方面充分利用专家丰富的知识、经验与综合分析判断能力对数据进行修正，有效提取数据并更新基础信息数据库。基本技术思路为：利用数据采集技术获取互联网上有关目标区的城市房屋基础信息数据，形成房屋信息数据表；采集互联网地图数据并数字化，进行图表信息融合，形成城市房屋基础信息 GIS 数据库；利用专家经验识别街景照片中房屋的信息，并开展必要的房屋现场调查，对初步形成的 GIS 数据库进行补充与修正，实现城市房屋基础信息 GIS 数据库的建设。通过定期采集互联网地图数据，比较不同期次地图差异，获取目标区地图与现有数据库对应地图的显著变化区，然后定位并据此进行针对性调查，以实现数据库的更新与维护。技术框图如图 1 所示。

1.2 主要步骤

(1) 互联网数据的采集

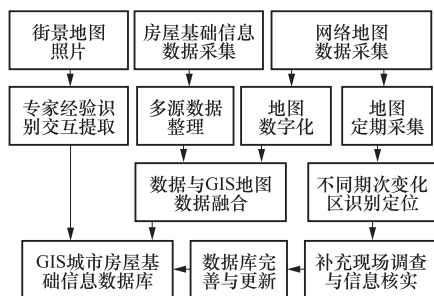


图 1 技术路线框架图

Fig. 1 Technical route frame

需采集的互联网数据可分为两部分，即房屋基础信息数据和互联网地图数据。对于房屋基础信息数据的采集，首先对数据源网站进行分析，这里主要的数据源网站是房屋管理部门门户网站以及大量的房屋中介网站。然后，利用网络爬虫技术从数据源网站搜索并下载与房屋有关的基础信息，比如小区、楼层、建筑年代、结构类型、建筑面积等。再次，对不同网站采集的数据进行整理，特别是对多源数据进行比较选择，提升数据可靠性。最后，融合形成统一格式的房屋信息数据表。对于互联网地图数据的采集，首先，

利用地图下载软件从目标地图网站下载分辨率较高的地图数据,一般为.jpg 格式的照片;然后,自行编程对地图照片进行拼接;最后,将拼接好的照片数字化,形成 GIS 基础地图图层。

(2) 图表数据的融合

将前述房屋信息数据表与 GIS 基础地图图层进行关联。鉴于数据量较大,宜通过编程实现自动关联,形成初步的城市房屋基础信息数据库。互联网数据虽然便捷廉价,但大量的公共建筑和事业单位办公用房屋信息缺失;同时数据采集过程中由于互联网数据本身不完整及受采集技术限制难免有所遗漏,故这个数据库是不完整的,还需采取其他方式对此进行补充。

(3) 街景地图与专家经验识别相结合提取房屋信息

目前在大中城市互联网街景地图非常普遍,通过街景地图不仅能看到房屋外观,而且一些房屋的细部结构也展示得十分清楚,有些还可进行拉伸放大显示,其效果几乎等同于一般的现场调查,使得通过街景地图与专家经验识别相结合提取房屋基础信息成为可能。此项工作涉及大量专业知识,而且专家应对不同历史时期城市房屋建筑发展特征有一定研究经验。一般,对于公共建筑及事业单位办公用房屋建筑基础信息,可将街景地图与专家经验识别相结合进行有效提取,其工作量主要取决于前期互联网数据的采集程度。

(4) 开展现场调查

对于一些断头路、死胡同地段的房屋信息“死角”,则需开展必要的现场调查以获取其房屋基础信息,进一步完善房屋基础信息数据库。一般而言,经过前几步的工作铺垫,现场调查工作在布置上会有明显的针对性,工作量也将十分有限。

(5) 利用地图对比识别技术进行数据库更新

由于近年来城市发展非常快,及时更新数据库显得十分关键。目前的互联网地图数据更新速度基本能够与城市建设速度相适应,如果定期从互联网地图采集目标区地图数据,通过对不同期次的目标区地图进行叠加对比,识别、定位显著变化区,即可确定数据库需要更新的区域,据此开展针对性的调查,即可实现城市房屋基础信息数据库的及时更新。这时的调查既可采用现场调查的方式,也可采用街景地图与专家经验识别相结合的方式。

2 天津和平区房屋信息基础数据库的建设

2.1 天津市和平区概况及房屋的基本特点

(1) 和平区概况

和平区位于天津市中心城区的核心区,面积 9.98 平方千米,常住人口 34.9 万,是全市金融、商贸、教育和医疗卫生中心,是天津繁荣繁华的标志区。和平区的中国历史文化名街——五大道,完整保存着 20 世纪初建造的 2000 多座各国风格的小洋楼,拥有全市 76% 的历史风貌建筑和名人故居。南京路沿途有大量现代化的高层、超高层建筑,形成了天津重要的城市发展集群;和平区还汇聚了天津大量优质教育、医疗资源,因此具有非常高的社会关注度。和平区很早就普及了互联网,随着经济的快速发展,房地产市场繁荣,与房屋有关的基础数据迅速积累,该区域的电子地图、街景地图日益精细,这些为利用互联网数据建设更新城市房屋基础信息数据库提供了有利条件。

(2) 房屋基本特点

天津市建有大量不同结构类型的建筑。20 世纪 50 至 80 年代经济发展初期,建设了许多

板楼；改革开放后，特别是近 20 年天津城市建筑结构变化很大，现代化高层建筑鳞次栉比，天津市具有国际现代大都市的建筑样式（王明浩等，2006）。天津城市的发展进程决定了天津城市房屋具有建筑风格多样、结构形式复杂、传统与现代并存等基本特点，和平区即是其中最典型的代表，这样的房屋特点对城市房屋基础信息数据库建设提出了更大的挑战。

2.2 和平区城市房屋基础数据库的建设

2016 年，在天津市和平区开展了城市房屋基础数据普查工作，采用了互联网数据与现场调查相结合的工作方式。

首先，从天津多个具有广泛影响的房地产中介网站与房屋管理门户网站采集了和平区房屋的基础信息，初步确定了和平区近 300 个小区的建筑面积、容积率、房屋结构类型、建筑年代等基础信息，获得了 3429 条房屋基本信息数据。第二，从百度地图下载了分辨率较高的和平区城市地图，并进行了数字化，确定了和平区有 5748 栋，这里栋指的是平面上各不相同的区域。第三，自行编程对多源数据进行融合整理，并实现与 GIS 数字化地图的自动关联，形成初步的 GIS 数据库，数据融合的可靠性分析采取地图抽样分析方式进行。通过这一步形成的和平区城市房屋数据库的数据分布见图 2，从中可以看到整个目标区中还有较多的连片空白区，这些空白区在地图上大多属于医院、学校、商场等各类商业、事业单位用地。这主要是由于获得的数据主要采集自房产中介，采集的房屋数据主要是住宅类房屋。



图 2 通过数据采集与融合形成的 GIS 数据库数据分布图

Fig. 2 Data distribution of GIS database after data collection and integration

由于大型商场、学校、医院等公共事业单位的房屋基础信息很少公开，很难从互联网上直接获取有效信息，但在中心城区，这类房屋的占比例又比较高，因此这类房屋的数据对数据库是不可或缺的。为此，我们采用了街景地图与专家经验识别的人工提取房屋信息方式进

行数据的核实与采集，虽然工作量比较大，但是通过反复尝试，发现仍然有章可循。首先，通过对和平区街景地图和高分辨率遥感卫星图片相结合的分析，发现在中心城区绝大多数建筑至少有一个立面临街，即房屋至少一面会被街景地图拍摄，高分辨率的遥感卫星图片可在平面上确定其基本布局；然后，由熟悉本地区房屋建筑结构特征、城市房屋建筑发展历史的专家对街景和遥感卫星图片进行解读，能够对房屋的基本信息进行有效识别。通过街景照片与遥感卫星图片不仅能够获得房屋的楼层、结构类型等基本信息，而且根据建筑外观特点、装修风格、开窗大小、通道设置等特征还能判断出大致的房屋建设年代（精度为 5 年以内）。在和平区房屋基础信息数据库建设中，通过这一步，数据库的完整性得到了很大提高，补充后其数据分布见图 3，基本填补了绝大多数的“空白”区。此项工作关键在于专家过硬的专业知识和现场经验，同时对目标区在城市发展不同阶段所形成的建筑风格、装修风格，结构样式等十分熟悉，以保证数据的可靠性。



图 3 通过街景地图的专家识别后 GIS 数据库的数据分布图

Fig. 3 Data distribution of GIS database after identification of experts through the street view

通过上述两步工作，数据库中还剩有少量“空白”，根据遥感照片的显示，主要是在老旧棚户区、断头路两侧、死胡同等地段，这部分工作只能通过实地调查考察加以补充。通过初步建成的数据库与遥感照片比较，调查目标能够非常直观地显示，使得调查目标更明确也更具针对性。由于和平区位于天津中心城区，经济发达，这部分工作的工作量比较小。图 4 显示了和平区信华南里小区的现场调查情况，从互联网地图上看，该小区内有较明显的“空白”（图 4 (a)），与遥感影像对比可以看到有大量低矮房屋（图 4 (b)），通过现场调查来看，的确有大量低矮二层、年代久远的房屋存在（图 4 (c)、(d)）。由于这一步的工作量较小，最终形成的 GIS 数据库与图 3 区别很小，这里不再展示。

数据库的及时更新十分关键。我们在和平区房屋基础信息数据库建设中，采用了对比不同期次地图数据的方式对数据库进行更新。首先，下载开始建设数据库时的地图数据，一般

而言这种地图数据都是.jpg 格式的图片, 将其拼接为覆盖整个和平区的图片; 第二, 每半年下载 1 次和平区的地图照片进行拼接; 第三, 叠加对比这两次下载的目标地图图片, 识别并定位影像变化显著区, 这里我们先将这两次的图片转为黑白图片, 然后再叠加对比; 第四, 对变化显著区进行补充调查, 实现数据库的更新。本项工作开展至今共下载和平区地图数据 3 次, 通过叠加对比, 未有明显变化区, 主要是由于目标区处于市中心, 城市建设基本完成, 故这 3 期地图数据几乎没有变化。



(a) 互联网地图显示的和平区信华南里小区



(b) 遥感影像显示的和平区信华南里小区



(c) 信华南里现场调查照片 1



(d) 信华南里现场调查照片 2

图 4 和平区信华南里小区现场调查

Fig. 4 Field investigation of Xinhuananli court in Heping district

3 结论与讨论

本文介绍了利用互联网数据建设与更新城市房屋基础信息数据库的技术思路, 并以天津

市和平区为例介绍了其主要步骤。该方法经济、便利、易实现,适用于大中城市的房屋基本信息普查工作。该方法有两个适用条件:一是目标区的互联网数据要丰富,能够提供必要的数据基础;二是调查专家对目标区房屋建筑特点、结构类型特征十分熟悉,可根据街景照片等有效提取房屋基本信息。上述两个条件直接决定了数据库建设的精度和可靠性,这也限制了该方法在广大农村地区房屋数据库建设中的推广。随著互联网日益发达,互联网数据对农村的覆盖将越来越深入,精度也将越来越高,数据提取技术更精准,其适用性将会得到不断提升。

在实际工作中有一些细节需要特别注意:①现场调查工作必不可少。互联网数据获取虽然方便,但不一定能满足实际工作需要,也可能夹杂虚假数据,故开展现场调查进行核实与补充十分必要;②数据库更新速度受制于地图供应商的数据更新速度,这样的更新速度是否适合研究需要还要具体问题具体分析;③目前的房产中介机构由于受商业竞争等影响,越来越重视对数据的保护,对大批量数据的采集逐渐采取了越来越多的技术封锁,同时个人也越来越注意隐私保护,这加大了数据采集难度。

本文提供了利用互联网数据建设城市房屋基础信息的基本技术思路,基本上能够解决城市房屋基础信息数据库的“有无”问题,但其结果还比较粗糙,还不能满足对数据精度要求高的工作需求。要建立精度高、更新快的城市房屋信息数据库,最好是部门间实现数据共享,具有统一的采集更新机制(李洺等,2008)。

参考文献

- 段锋,袁志祥,2010. 结合汶川 8.0 级地震实践探讨地震应急基础数据库更新途径. 四川地震, (1): 32—35.
- 哈丽娜,王万江,汤高举,2010. 电子影像图在既有建筑普查中的应用. 建筑节能, 38 (3): 71—73.
- 韩贞辉,2013. 基于震害分类的房屋数据格网化研究. 北京: 中国地震局地质研究所.
- 李洺,2008. 政府应急平台数据库的数据需求、实现路径与管理制. 电子政务, (5): 56—61.
- 聂高众,陈建英,李志强等,2002. 地震应急基础数据库建设. 地震, 22 (3): 105—112.
- 王斌,陈雪洋,林娜等,2014. 广域范围建筑物信息普查关键技术探讨. 地理空间信息, 12 (2): 32—34.
- 王丽莉,宋思然,姜德录等,2008. 辽宁省地震应急基础数据库建设. 东北地震研究, 24 (4): 28—36.
- 王明浩,李小羽,2006. 天津市建筑风貌与城市特色研究. 城市发展研究, 13 (3): 133—136.
- 郑树平,高翔,赵杰等,2015. 山西省地震应急基础数据库更新维护. 计算机时代, (1): 75—77.

A Preliminary Study on Building Basic Information Database of Urban Buildings by Using Internet Data

Gao Wuping, Liu Fang, Yao Xinqiang, Yang Xulian, Yang Fei and Fu Zhanling

(Tianjin Earthquake Agency, Tianjin 300201, China)

Abstract A new method for establishing and updating urban building database is proposed in the paper. First, basic information of residential building is directly extracted and effectively compromised with map from internet, and combined street view map and remote sensing photographs with expert experiences. Second, basic information of public and institution buildings were distinguished for those which could not be obtained from internet directly. Third, the data was supplemented and verified by on-site investigation, then a database would be established. Lastly, downloading maps of investigation area in different periods, and superimposing and comparing these maps, then the significant changes area is located where some supplementary investigations would do, a primary urban buildings database would be updated regularly. According to the case study of building database establishment in Heping district in Tianjin, we found that the method is convenient and very effective and can be updated periodically. It is a good reference for establishment of basic information database of other cities. It is very important to note that due to limitation of the integrity, reliability and details of data for internet, the accuracy of database is largely affected.

Key words: Urban housing information database; Expert experience identification; Database update