

第四章 历史图像与数字人文应用

——以“数字中国”数据库为例

张志云

本文将描述如何以两个学术机构的专项技术作为基础,建立以中国历史照片、图片与相关文本数据为主要内容的“数字化中国”(Digital China)平台,并阐述如何利用这一数据平台进行英国在华“非正式帝国”的研究。结合台北“中央研究院”人文社会研究中心地理资讯科学研究专题中心(以下简称GIS中心)的历史GIS技术与英国布里斯托尔大学历史系所搜集的大量中国老照片(由该大学“视觉中国”和“中国历史照片”两项计划提供),本文介绍的数据平台将建立全新的历史地理信息系统,并利用GIS技术将影像资料的时空分布精准化、实时化、全景化,进而令使用者更精准地搜寻影像资料,并使用老照片时空(spatio-temporal)属性及此维度下的宏观分布数据进行学术研究。除此之外,全新的数据平台结合了历史GIS技术、内容管理系统(Content Management System)与WebGIS技术,这将对图像史、环境史和历史地理等领域的研究大有裨益。

一、图像的的来源:19—20世纪的旅华英人

18世纪之前,坐落远东的中华大地对于西方人而言,是遥远的、令人迷惑的、忧虑与敬畏的东方国度。这些情绪是当时西方来客尝试前往中国、了解中国的感性动力。这种情况在鸦片战争后发生了本质改变,因为在华洋人的身份结构和来华目的发生均较大了变化。随着中国的门

户大开,外商与外交官(以及中国海关外籍职员)的人数与日俱增,传教士的活动也不再仅仅为英国的在华商业活动、文化建构提供支持,而是成为政治干预控制的传动轴。以上四类人共同构成了在华英国侨民的主体,他们于治外法权、在华贸易体系与对清政府政治干预等三个方面扮演各自的角色。在这个群体之中,形形色色的在华英人由于其身份、职业属性等层面的差异,活动轨迹也各相径庭。有的穿梭于各通商口岸之间,考察西方在华商业活动的运作情况及对中国社会人文的推动效应;有的则将足迹踏遍中国的大西南,寻觅藏传佛教在华印记与远古时期的“玉龙第三国”。

中国老照片

近代中国仍模糊地残存于集体历史记忆之中,国人的图像记忆大抵来自于当时拍摄并保存至今的老照片,其中相当部分为在华英人所贡献。他们光圈下的中国近代剪影,遍及中华大地无数角落与中国社会诸多层面。这些照片如今部分妥善惠藏于中、英两国乃至世界各地的博物馆、档案馆、图书馆与展览馆之中,另一些则由私人收藏至今。这些收藏者往往与中国历史有某种剪不断、理还乱的渊源。广览多若繁星的老照片,我们可以发现,不同身份摄影者的拍摄内容截然不同。如中国海关营造项(Works Department)外班工师(Clerk of Works)英人的安道立(Augustus Anthony)的摄影内容,主要反映了中国沿海的灯塔及周围设施;而思茅关护理税务司(Assistant-in-charge)三等后班帮办甘累(Frederic William Carey)的照片,则记录了云南地区的风景与人文风貌,因为他曾在1896年至1902年间于滇缅边境的蒙自关任职;彼得·克莱因(Peter Klein)是英国皇家海军军官,其照片多与在华英国军舰有关;上海公共租界巡捕沃尔特·帕莫(Walter Attwell Lynch Palmer)的摄影,则反映了20世纪20—30年代上海公共租界的面貌与亚细亚火油公司(Asiatic Petroleum Company)的企业设施等主题。

上述中国历史影像既保罗宏大叙事,又折射出彼时英国在华群体的

一个侧面。我们不仅可以从中看到当时中国社会的总体风貌,也可以感受以上海为代表的近代中国城市发展的微观印象,以及社会生活历史沿革的地方社群性记忆。例如,自1843年上海开埠并建立英租界以来,印度人开始出现在上海的街头。1883年起,工部局开始从印度招募巡捕与治安人员,其中多为信奉锡克教的锡克族人。杰逸(Isabella Jackson)在其文章《投射于南京路的英国对印殖民残片:上海通商口岸里的锡克族巡捕》(*The Raj on Nanjing Road: Sikh Policemen in Treaty-Port Shanghai*)中,阐述了锡克族警察在20世纪前半叶上海风貌中所留下的不可抹消的印记。^①在诸多关于公共租界巡捕房的上海老照片中,均能发现这些锡克族巡捕的身影。他们的存在及其在上海公共租界治安事业中扮演的代理人角色,可被视作英国在华影响之又一缩影。

相比19世纪后半叶而言,20世纪前半叶的影像资源更趋丰富,内容也更为多样。据毕可思(Robert Bickers)在《寓于近代中国的英国:社会、文化与殖民主义,1900—1949》(*Britain in China: Community, Culture and Colonialism 1900—1949*)的研究,自20世纪初,在华英国/外国侨民与英商/外商企业数量急剧攀升(表1)。在此期间,太古洋行创始人之子,华伦·施怀雅(Warren Swire)曾多次来到中国,其来华目的之一即增强太古洋行对中国社会的影响力,以及加深太古洋行在华商业贸易活动中的“英国属性”,及其给予生产、消费市场的英国印象。^②毕可思在其《消弭中国之界:太古洋行及其背后的商业帝国,1816—1980》(*China bound: John Swire & Sons and its world, 1816—1980*)一书中提到,“将这家公司及其相关企业、各方所提供的产品及服务的异邦属性,并借此昭示该问题之重要性,便成为了当时一项重要的政治性议题。”^③从华伦·施怀雅的照片集中,可见其摄影关注点主要集中在洋行货轮、货栈、靠泊码头、堤岸码

① Isabella Jackson. “The Raj on Nanjing Road: Sikh Policemen in Treaty-Port Shanghai”. *Modern Asian Studies*, 46(6), 1672.

② Bickers, Robert. *China Bound: John Swire & Sons and its world, 1816—1980*. Bloomsbury Publishing, 2020. 226.

③ Bickers, Robert. *China Bound*. 227.

头等商业货运设施,且数十载之间摄影内容未曾变化。由华伦·施怀雅的视角我们看到了进入20世纪后,“在华贸易体系”在“非正式帝国”构建中所发生的变化与起到的作用。

表1 (除香港外)在华英国/外国侨民与洋行总量,1910—1930

年份	英国洋行	外国洋行总数	英国在华侨民	外国在华侨民总数
1910	601	3,239	10,140	141,868
1915	599	4,735	8,641	182,404
1920	679	7,375	11,082	326,069
1925	718	7,743	15,247	336,841
1930	1,027	8,279	13,015	255,686

Robert Bickers. *Britain in China: Community, Culture and Colonialism 1900-1949*. Manchester: Manchester University Press, 1999. 13

未被发掘的新史料与史料的镜像保存

新史料的发掘、原始数据的数字化以及资料的系统整合,是整个“数字化中国”数据平台搭建的核心。全新的数据平台完成了对39个类别,共计愈22000张照片的数字化工作。此中,包括来自伯明翰大学的吉百利学术图书馆(Cadbury Research Library)的资料。私人收藏也是本数据平台的重要信息来源,例如巴斯皇家文理研究所(Bath Royal Literary and Scientific Institution)的瓦谢收藏品(Vacher Collection),其历史可追溯到1857年以前。此外,由费利斯·比托(Felice Beato)在1860年拍摄的有关华北战役的新相册,也被收入其中。以上数字化工作极大地丰富了本数据库,扩展了其时间跨度和地理范围,使之所反映的地点、时段与机构范畴得以空前深化。

档案保存技术与图书情报编目为传统史料的保存提供了极大便利,而近几年随着数字人文技术的发展,各种数据库也为史料的保存利用提供极大帮助。我们将各类历史图片上传至平台服务器,不仅极大帮助了这些珍贵史料的保存,同时也为世界各地的学者利用这些珍贵影像提供接口。然而,搭建全新的数据平台,需要进行服务器之间的镜像与转换。

为了将最初保存在法国里昂服务器上的元数据,迁移到布里斯托尔大学服务器,需要通过设计特定数据平台架构,以避免出现数据丢包与数据错误的现象。而在数据迁移之前需要开展大量整合与校正工作,在迁移完成后则需要清理老系统中的元数据。这些技术上的难题均可通过开源内容管理框架(CMF,即CMS+Framework)与全新的后端管理设施解决,而这两项数字人文技术也简化了图的批量上传与后台文件处理工作,从而释放了人力与算力资源。因此,这些技术都是数据平台的数字资产管理工具,同时扮演着史料档案管理员的角色。^①

同时,原有“视觉中国”平台(<http://visualisingchina.net/>)的内容也获得长足的提升,并且已经在基于“中国历史照片”网站的克隆站点中予以重建。“视觉中国”这一平台中的新史料来自哈佛燕京图书馆,包括杜德维(Edward Bangs Drew)收藏(约1870年代—1900年)的374张照片、赫达·莫里逊(Hedda Morrison)的相册(1933—1946年)的4683张照片以及毕敬士(Claude L. Pickens Jr.)传教士收藏的1080张反映中国西部回族人民生活状况的照片(20世纪30年代)。以上照片来源大大扩展了图片资料库的地理覆盖范围,并在传统中国近代史分期所覆盖的范围(1860年至1950年),丰富了早期和晚期两个区段的图片数据数量。

二、历史图像的整合:内容管理系统和地理信息系统的应用

数字人文技术为呈现并研究旅华英人的记忆提供了全新的工具,通过对这些照片时空信息的编码,图像史料间的比较研究,有助于增进学界对旅华英人这一群体的认知。在搭建名为“数化呈现中华大地的英伦踪迹”(Digitize the British Presence in China)的地理信息系统网站的过程中,内容管理系统(CMS)技术与开放资源和地理信息系统相结合,以清晰的时空视野展示收集自英国与中国台湾的历史照片。这些图像资源以

^① 该新数据平台将在今年上线,以纪念2006年“中国历史照片”项目启动。新平台将使用<http://www.historicalphotosofchina.net/>这一URI域名(该URI现已获取,只是平台系统及内容暂未上线)

动态形式予以呈现,并依据照片描述记录的空间属性进行分类。将照片以时间为脉络编列于地理空间参照图谱中,更加方便用户浏览某个特定时期、特定地点的全部影像资料。如何涉及网页布局以及实现在不同浏览模式下的数据适配,成为数据平台搭建过程中的重要问题。数据库提供物种不同的浏览搜索模式,为学者们应用历史图像进行研究提供不同的分析方向:

照片标注

照片时空信息的标注编码,有助于学者精确寻找其所需的历史图像。在主页上,图片按其内容所涉及的主题予以分类,例如服饰、建筑等类别(图1)。一旦使用者选定他们感兴趣的主题之一,系统则会自动呈现与该类别相关的所有图像(图2),以供使用者进一步甄选。与此同时,绝大多数数据库内的照片资源都根据其内容做了详细标注。每张照片都根据其拍摄内容,结合特定社会历史状况,赋予其简要、高辨识度的照片标题。该标题并不会完整出现在检索页面的缩略图下,取而代之的是照片所反映的历史事件名称或由城市、具体地标或路名等地理位置讯息组成的缩略版标题。这不仅有助于应用模糊检索的使用者以名称为基础,迅速定位符合自身需求的照片资源,也简化了应用分类标签进行检索的用户在浩如烟海的缩略图中寻找可用图片资源的过程。大多数情况下,学术工作者与其他领域用户手中最常见的信息即为历史事件与地理位置讯息,而以这两类信息为内容的缩略标题应能更有效地指引用户对图片内容的推断。

传统史学方法仍旧在搭建数据平台至关重要,照片信息的整合与勘误、标题中所涉及关键词信息的取舍与斟酌都需要学者逐一考证。正因如此,在传统数据体量庞大的图片数据库中,图片编码并不完善,导致数据库难于检索利用。学者们使用图片资源数据库时,难免淹没于缩略图所提供局部讯息的资讯洪流中。在较多情况下并无足够的文字资料佐证照片的时空信息,因此在数据平台的建设过程中仍需要依靠

传统历史考据方法,进而判断部分照片的来源,所拍摄的地理位置等信息。

在我们搭建的数据平台中,通过学者们大量考证,照片下都有详细的注释内容,其中介绍了该照片在布里斯托尔大学资料库的照片归档号、照片出自的摄影集或照片收藏集、拍摄者的个人信息、图像内容的细节描述、历史考据等。例如,图 2A 的注释介绍了达温特牧师(Rev. Charles Ewart Darwent),他不仅是上海佑宁堂的堂区牧师,也是一位摄影师,出版《上海:面向旅行者与居民了解外国人聚居区与本地人城市区域及周围景点的手册》(*Shanghai: a handbook for travellers and residents to the chief objects of interest in and around the foreign settlement and native city, 1904*)。图片右侧还给出了该照片的推定摄影年份(1902年)。此外,注释中还介绍了这张照片的内容、拍摄地点,即“位于苏州河沿岸的船坞”,并展示了照片所处的确切方位是南苏州路(当时的苏州路)与圆明园路的交汇点附近,而背景中的建筑从左到右依次是虹口医院、德意志福音教堂、德国领事馆。^①照片时空信息的标注与编码不仅依靠历史学者的考证,同时这些旅华英人所拍摄的照片也为中国近代史研究提供更多图片记忆。

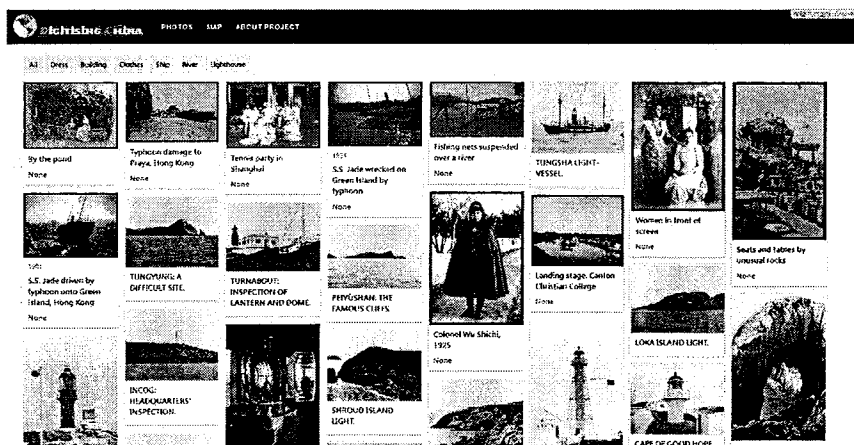


图1 主页上所呈现的照片分类情况

① <https://www.hpcbristol.sjtu.edu.cn/visual/da01-01>

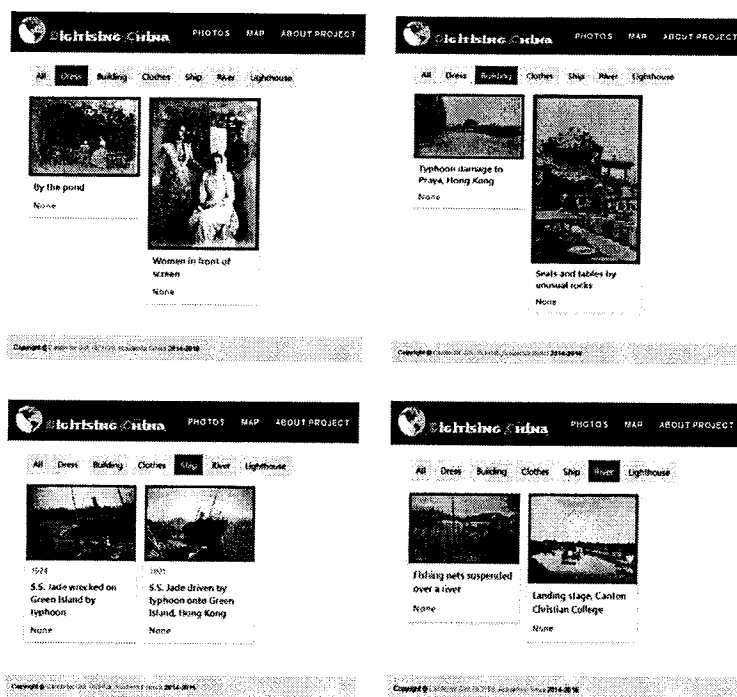


图2 基于不同分类标签筛选浏览照片



图2A 照片资源范例:自船库漫视苏州河、
(Soochow Creek from Boat House)

多种数字化平台上的适用与呈现

数据库的标注模式可以适应历史学者们的研究需求,除此之外为了满足多场景研究需求,数据库还需要考虑学者们在使用数据库时的不同终端。为了使数据库内容在台式电脑、笔记本电脑、智能手机和平板电脑等多种不同平台上的正常使用,该网站采用响应式网页设计(Responsive Web Design, RWD)进行编程。这使得网站的操作界面可以适应各平台的制式、大小与分辨率。基于液态网格(Fluid Grid)的RWD可使PC端网页平台的页面缩放百分比或字体大小缩放更具弹性,可让平台呈现的内容适应不同屏幕宽度。基于RWD的本平台拥有“网页要素宏观弹性化”、“图片响应式呈现”两方面特征。通过在恰当的地方使用合理的HTML标记,图片的尺寸可以随平台的变化、网页缩放的调节而自由调整,并可兼容部分移动设备所拥有的横竖屏切换功能;与此同时,由于图片资源的呈现应是本平台网页内容的核心功能簇,故同比例的图片缩放与分辨率的适应性调节对于用户搜索及使用该平台诸多资源至关重要。

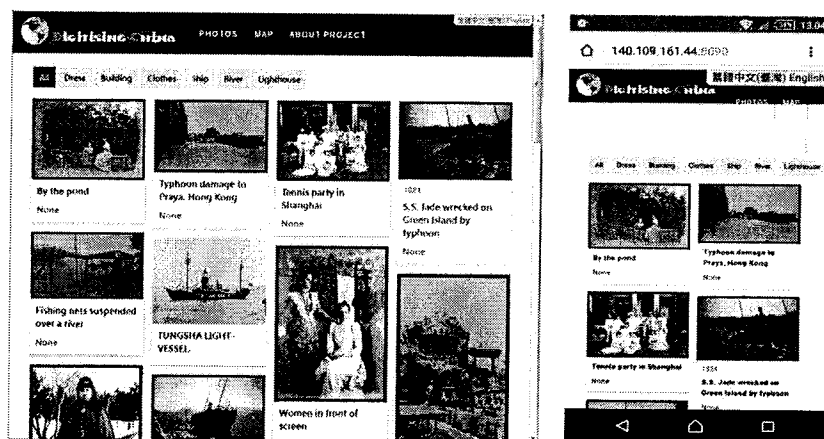


图3 基于RWD技术的多平台呈现效果

如图3示例,搜索界面下的缩略图在Windows系统的网页浏览器下以多图每行的方式呈现,在平板电脑的竖屏模式下,则为每行四张图片。与此同时,手机用户则会以每行两张图片的方式呈现。由于移动设备的网页呈现奉行紧凑度优先于整体结构的原则,图片的纵向间距取消,且网

页会根据图片本身的尺寸属性,智能选择图片的排布模式。例如图三左图中绝大多数图片都是横向图片,其中少量的纵向图片则会自动与其余横向图片等宽,并等比例呈现图片的纵向维度。

互动界面

除了首页的主题标签外,数据平台还设有互动搜索功能,供用户选择特定的拍摄地点或摄影师,以满足具体的史料追踪需求与学术研究需要。如前所述,根据照片上所呈现内容的不同,本平台对各图片资源进行了详细的标签设置。绝大多数图片的关键词标签都不少于8个,其中包含摄影师的姓名、图片上的人物身份、动物种类,主要景观(植被、路况、建筑、交通工具)、城市与道路名称等。同时,图片资源分门别类的标签设置也会体现在搜索功能的优化上。例如在高级搜索选项中,用户可根据摄影师、照片集、照片拍摄地以及其他标签名称等内容进行优先筛选。例如图4中,用户在高级搜索选项中限定了拍摄地“北京”,搜索结果则会根据各照片的照片信息与标签准确筛选呈现之。与此同时,通过数据库照片资源信息汇总,平台汇集了所有可能出现的拍摄地、拍摄者与照片集名称,在高级搜索选项界面下形成下拉菜单,供用户筛选。通过多个搜索条件的定位,用户便能如图5所示,准确定位所需搜寻的照片资源。



图4 根据拍摄地点进行高级筛选

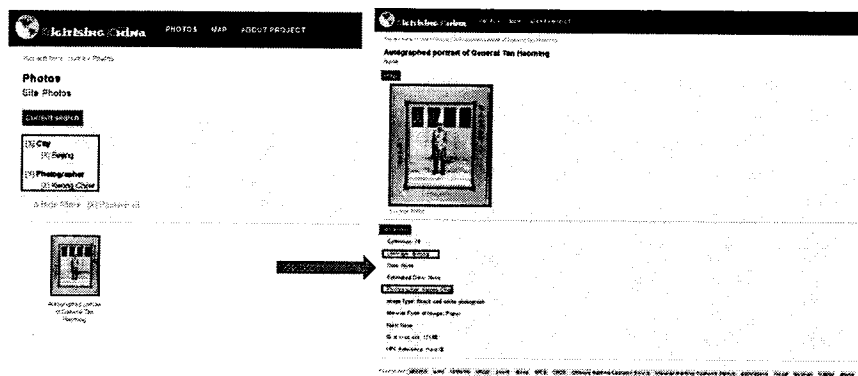


图5 叠加拍摄地点与摄影师两个筛选条件的高级筛选

地图模式下的搜索

在地图模式下的搜索,能使历史研究正直观地了解19世纪下半叶20世纪上半叶,旅华英人历史记忆的地理分布,并且能够探寻他们在中国各地的旅行踪迹。不仅如此,地理分布也使得比较不同英国旅行者的中国记忆成为可能。通过比较同一地理位置下不同时空环境的变迁,这些老照片能够为我们探究英帝国在华扩张的全新视角。

地图模式的操作界面会分成两个部分,其中上半部分呈现的是现实地图与历史地图相整合的综合地图,下半部分则以时间轴形式呈现相应照片的时间分布(图6)。现实地图与历史地图的整合并非简单的图片贴

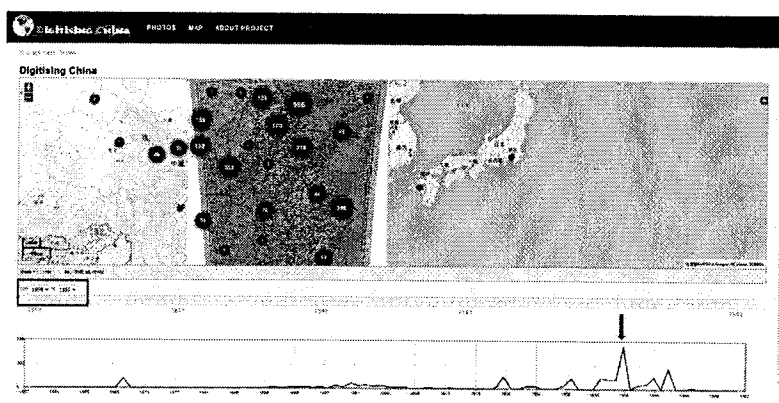


图6 地图模式

合,而是根据历史地图上的标志性地点与现实地图上相应地点之间的对应,调节历史地图的横、纵向梯度,并在特殊情况下采取特殊的缩放方式,直至地图上特定位置的点阵基本对应。而界面下部的时间轴展示了特定年份范围内各年可浏览照片的数量,形成较为直观的折线图。该折线图有助于用户了解其搜索条件下、不同年代图片资源之丰度,有助于其优化筛选条件,以便进一步利用上图的 geo-referencing 信息实现时、空两维度的交叉索引。

使用者亦可根据图片数量自行甄选起止年份,以便于 GIS 数据的呈现与图片的进一步筛选(图7)。在上半部分的地图界面中,用户在调整

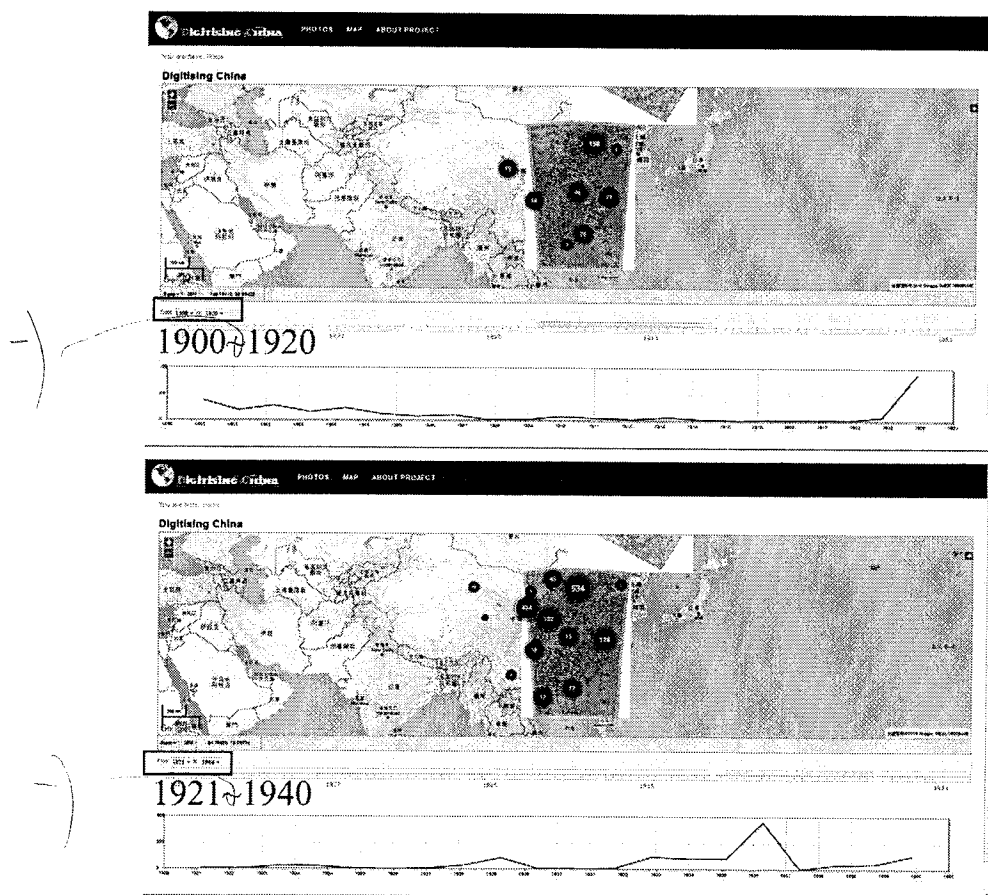


图7 时间线模式下的照片筛选地图视图

地图缩放程度的同时,图上表示照片拍摄位置及照片数量的红圈也会随之重新分离、整合、数据更新(图8),以此实现流畅而动态的数据呈现。这种基于GIS系统与数据呈现应用模型间集成的全新概念技术,使人机互动更为顺利,操作程序更为简便。研究者在阅读地理讯息时,可以获得更多的信息,获得更加动态的数据分布视野。



图8 通过各照片的地理属性畅享搜索过程

图片史料信息与GIS系统的整合则会进一步帮助历史地理研究、历史人物研究绘制人物与历史现象分布、迁移的宏观布局与循迹路径。例如根据照片的摄影者、拍摄地点标签可知摄影者的行动范围,而不同身份、不同文化背景摄影者所呈现的足迹范式应是截然不同的。达温特牧师因其堂区牧师的身份,与绝大多数拥有传教士身份的摄影者拥有不同的行动路径,前者拍摄地点集中在上海,后者则往往遍及各省(尤其沿海各省)的主要城市。华伦·施怀雅(G. Warren Swire)是太古洋行创始人约

翰·施怀雅(John Swire)的儿子。他曾经于1906—1907年、1911—1912年、1919—1920年、1923—1924年、1928—1929年、1933—1934年、1938年1940年多次造访中国各地,留下了大量的摄影资料,其摄影目的是构建当时太古洋行在中国形成的庞大商业、贸易架构,及其在各地的影响力与对中国社会产生的影响。因而由其照片拍摄时间与地点可见,此人足迹遍及上海、塘沽、天津、秦皇岛、牛庄、大连、南京、芜湖、芝罘等,与通商口岸高度重合,其拍摄内容也以航运、企业设施、运输与运营设施为主。因此,在本文搭建的GIS数据平台中,照片资源注释体系以及历史图像时空信息的编码有助于发现这些全新的历史现象,为历史研究者利用图像史料提供分析工具。

历史地图图层

在这一模式下,我们可以将历史地图叠加到地图模式下的搜索结果,通过GIS技术进行古今地理互考参照。图层的叠加不仅有助于认识历史地理变迁,同时,学者们也可以比较同一地点不同时间段的历史图像,探讨地理环境的历史变迁对旅英华人历史记忆的影响。例如,本网站提供1903年中国地图(比例尺1:3600,000)(图9)和1907年哈尔滨周边地区地图(比例尺1:1680,314)(图10),供使用者熟悉当时的地理信息,并将历史地图的影像数据整合于相应的现实地图之中。由于地图比例尺、绘制方式、各向缩放比例很难与现今地图制图法存在很大差别,故仅靠人工

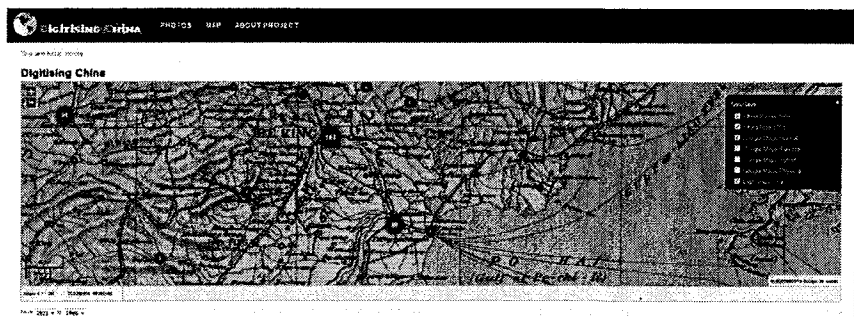


图9 1903年中国地图

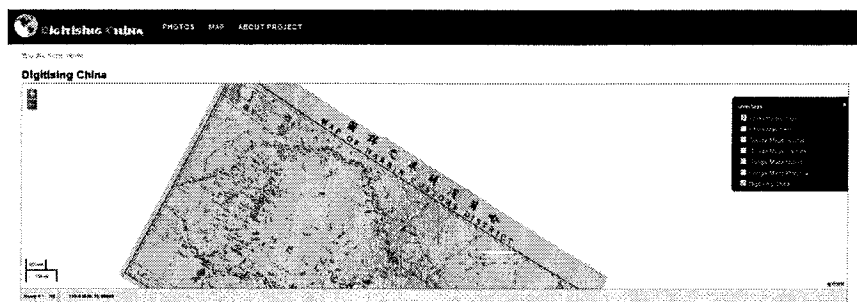


图10 哈尔滨海关关界详图(1907)

方式,欲使历史老地图与卫星图做到无缝重合,难度是非常高的。通过将各历史地图上可考地点的经纬度与现代地图上的相应点对应,并以此为依据在所需维度上缩放历史地图,新平台将实现古今地图信息有效整合。通过设置数个(理论上越多越好,最好不少于5个)位置可考的坐标配准点,然后运行坐标配准,即可将历史地图与地理信息系统上的位置相匹配。

结 论

本文介绍了历史图像与数字人文技术在历史研究中的各种面向。文章所描述的结合不同平台以及地理信息系统的研究方法,为中国近代史和中外关系史研究提供了新的面向。通过将地理信息系统与历史图像进行数位结合,该平台为历史研究者系统使用图像史料提供便利。尽管本文所阐述的数据库搭建合作过程往往侧重于技术开发层面,但从实际意义上讲,它为未来的跨学科合作和交流奠定了坚实基础。

本数据库收集的图像史料既是19世纪末20世纪初旅英华人的记忆,也是英国在华势力扩张的产物。因此,对这些图像史料的批判性使用是极为重要的,而数字人文技术为历史学者使用历史图像提供了新的分析工具,有助于史学家在图像史料中发现新的历史现象,增进学界对旅英华人群体的研究。而摄影照片的保存与分类一直是档案学与图书情报学的重要问题,而数据库的建设使得图像史料的保存归类更为合理,也为数据

库的内容编码提供了底层框架。从平台内容层面讲,影像资料的时间跨度、空间跨度与资料丰度都得以空前扩展。这对该平台的使用者而言裨益良多。平台将这些原本不为人知的独有史料越来越多地代入大众视野中并为之所用,并通过数字化手段将这些资料保存下来。不同领域技术的整合与新技术的应用,使得这一数据库成为一个学者社群中图片资源共享的平台,在相当程度上符合以学术工作者为核心的、各行各业人群的需求与使用习惯。本文中的数据平台不仅定义了图片史料资源共享、搜索的全新行动模式,在技术生态上也更具潜力与可持续性,有利于更长时段的史料资源扩充与功能优化。历史图片数据库的搭建也依赖对影像资料的搜集以及版权协商,这要求研究者进行更为广泛的学术机构跨国合作。

(作者单位:上海交通大学人文学院历史系)