ISPRS Journal of Photogrammetry and Remote Sensing 141 (2018) 72–85

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Crowd-sourced pictures geo-localization method based on street view images and 3D reconstruction

Liang Cheng [a,b,c,d,*], Yi Yuan [b,d], Nan Xia [b,d], Song Chen [b,d], Yanming Chen [a,b,d], Kang Yang [a,b,d], Lei Ma [a,b,d], Manchun Li [a,b,d,e,*]

[a] Collaborative Innovation Center for the South Sea Studies, Nanjing University, Nanjing, China
[b] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, Nanjing, China
[c] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, Nanjing, China
[d] School of Geography and Ocean Science, Nanjing University, Nanjing, China
[e] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

ARTICLE INFO

Article history:
Received 19 April 2017
Received in revised form 2 January 2018
Accepted 16 April 2018

Keywords:
Street view images
Geo-localization
Image retrieval
3D reconstruction
Accuracy analysis

ABSTRACT

People are increasingly becoming accustomed to taking photos of everyday life in modern cities and uploading them on major photo-sharing social media sites. These sites contain numerous pictures, but some have incomplete or blurred location information. The geo-localization of crowd-sourced pictures enriches the information contained therein, and is applicable to activities such as urban construction, urban landscape analysis, and crime tracking. However, geo-localization faces huge technical challenges. This paper proposes a method for large-scale geo-localization of crowd-sourced pictures. Our approach uses structured, organized Street View images as a reference dataset and employs a three-step strategy of coarse geo-localization by image retrieval, selecting reliable matches by image registration, and fine geo-localization by 3D reconstruction to attach geographic tags to pictures from unidentified sources. In study area, 3D reconstruction based on close-range photogrammetry is used to restore the 3D geographical information of the crowd-sourced pictures, resulting in the proposed method improving the median error from 256.7 m to 69.0 m, and the percentage of the geo-localized query pictures under a 50 m error from 17.2% to 43.2% compared with the previous method. Another discovery using the proposed method is that, in respect of the causes of reconstruction error, closer distances from the cameras to the main objects in query pictures tend to produce lower errors and the component of error parallel to the road makes a more significant contribution to the Total Error. The proposed method is not limited to small areas, and could be expanded to cities and larger areas owing to its flexible parameters.

1. Introduction

With the popularity of electronic products such as digital cameras and smartphones and the ubiquity of the Internet, people are increasingly accustomed to taking photos of everyday life in modern cities and uploading them to the Internet. Photo-sharing websites (such as Flickr and Instagram) and social networks (such as Sina Weibo, QQ Qzone, and Baidu Tieba) contain massive amounts of crowd-sourced pictures from various sources and at various resolutions and magnifications (Li et al., 2016; Heipke, 2010).

Although nowadays, many mobile terminals including smart phones can take photos with inbuilt geo-tags, in many circumstances, lots of photos available to us have no location information. Many crowd-sourced photos from the Internet or from uncertain resources have unknown spatial positions (Ji et al., 2015). For instance, some pictures are taken by digital devices without GPS, such as professional camera equipment including digital SLR cameras, mirrorless cameras, and old digital devices. Some pictures lose their original location information after being corrupted, uploaded, processed, reposted, and downloaded. People also may conceal or disguise photo location information for security and privacy reasons.

If relatively accurate geo-tags were added to these crowd-sourced pictures, they could play a meaningful role in outdoor localization (Arth et al., 2015), pedestrian detection (Yin et al.,

* Corresponding authors at: School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China.
E-mail addresses: lcheng@nju.edu.cn (L. Cheng), limanchun_nju@126.com (M. Li).

https://doi.org/10.1016/j.isprsjprs.2018.04.006
0924-2716/

2015), and unmanned driving (Salmen et al., 2012; Tsai, 2015), which will be of great benefit to augmented reality views (Liu and Wang, 2009). In addition, the geo-tagged pictures could also be utilized for urban landscape analysis (Badland et al., 2010; Li et al., 2015; Yin and Wang, 2016), urban construction (Cheng et al., 2016a; Hara et al., 2015; Hara et al., 2013; Soheilian et al., 2013), and urban analysis (Cheng et al., 2018b; Cheng et al., 2018c).

Accurate geo-localization has been a difficult problem for a long time. It involves identifying, extracting, and indexing geo-informative features from a large variety of datasets, discovering subtle geo-location cues from the imagery, and searching massive reference databases (Zamir et al., 2016). To overcome these problems, this paper proposes a method in which Street View images are used as a reference dataset. The proposed method employs a three-step strategy—coarse geo-localization by image retrieval, selecting reliable matches by image registration, and fine geo-localization by 3D reconstruction—to attach geographical tags to pictures from unidentified sources so as to realize crowd-sourced picture geo-localization. In contrast to previous methods, the proposed approach employs photogrammetry techniques for positioning refinement by close-range 3D reconstruction (Fraser et al., 2002; Goncalves and Henriques, 2015; Yuan et al., 2009), which leads to more accurate geo-localization and provides flexible applicability parameters.

This paper is structured as follows. In Section 2, the literature related to the geo-localization of crowd-sourced pictures is discussed. The study area and datasets are introduced in Section 3, while the method proposed in this paper is described in detail in Section 4. The experimental results are presented and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Related work

Current methods for the accurate geo-localization of crowd-sourced pictures fall into three categories: image-based, point cloud-based, and semantic approaches.

(1) Image-based geo-localization

Image-based geo-localization employs similar techniques to image matching in the field of computer vision (Cheng et al., 2008; Lowe, 1999). Traditionally, large-scale image-based localization has been treated as an image retrieval problem (Sattler et al., 2011). In this method, it is assumed that a reference dataset consisting of geo-tagged images is available, and the problem is transformed to that of estimating the geo-location of a crowd-sourced picture by finding its matching reference images (Chen et al., 2011; Knopp et al., 2010; Nister and Stewenius, 2006; Zamir and Shah, 2010; Zamir and Shah, 2014).

Compared to LiDAR (Light Detection and Ranging) and remote sensing images, Street View images (first launched by Google in 2007) provide a distinctive "human perspective" map service in a manner that is consistent with crowd-sourced pictures. Street View images are often used as a reference dataset to realize picture geo-localization via methods such as vocabulary trees (Schindler et al., 2007), geo-information codebooks (Knopp et al., 2010), and data-driven scene matching (Hays and Efros, 2008), with local and global constraints subsequently added (Zamir and Shah, 2014). Additionally, feature descriptors such as speeded-up robust features (SURF; Yazawa et al., 2009) and PCA-SIFT (Ke and Sukthankar, 2004), which are variants of the scale-invariant feature transform (SIFT), also show improved effects in some cases (Park and Moon, 2015; Kamencay et al., 2012).

The use of deep learning algorithms has generated considerable interest in the fields of content retrieval (Tang et al., 2015) and object detection (Xiao et al., 2015) in remote sensing images (Blaschke, 2010), video content retrieval from UAV (unmanned aerial vehicle) videos (Se et al., 2011), and artificial intelligence systems such as PlaNet (developed by Google in 2016).

(2) Point cloud-based geo-localization

Point cloud-based geo-localization establishes a correspondence between 2D pictures and 3D points, and then aligns a query picture against a 3D point cloud or model (Lehtola et al., 2016; Mostafa and Schwarz, 2001). This kind of method can obtain more stereoscopic information than image-based geo-localization from a 3D reconstruction of the scene; as a result, higher localization accuracy can be achieved (Sattler et al., 2011). Li et al. (2012) proposed a method that scales to datasets with tens of millions of 3D points through the use of a co-occurrence prior for Random Sample Consensus (RANSAC) and the bidirectional matching of image features with 3D points. This approach brings together research on image localization, landmark recognition, and 3D pose estimation, and achieves excellent geo-localization results, but the geo-registered 3D point cloud used for evaluation is difficult to establish (Cheng et al., 2018a; Svarm et al., 2014; Gemhardt et al., 2015).

Recently, visual simultaneous localization and mapping (vSLAM; Karlsson et al., 2005), which builds 3D point clouds of the environment and tracks the location and direction of the camera simultaneously, has aroused the interest of researchers. The mass application of vSLAM in robotics promotes its potential integrations with deep learning algorithm (Poirson et al., 2016).

(3) Semantic geo-localization

Semantic geo-localization uses techniques based on high-level and semantic cues, ranging from human-related characteristics such as text, architectural style, vehicle types, or urban structures to natural properties such as foliage type or weather (Singh and Košecká, 2016). The three main challenges of semantic geo-localization can be summarized as: what features to employ, how to match them, and how to consolidate the diverse semantic cues in a unified system. Crandall et al. (2016) learnt landmark models with a multiclass SVM (Support Vector Machine) and incorporated semantic metadata available on the Internet. Their results show that textual tags and temporal constraints lead to significant improvement in geo-localization, especially for landmarks.

3. Study area and data

Fig. 1 shows the study area of Jianye District, which is one of the main urban areas in Nanjing, China. This area has developed into a modern city center, whose growth reflects Nanjing's development. The condition of the streets in Jianye District is broadly indicative of those across Nanjing. The study area covers 76.4 km^2, with an approximate length of 10 km and width of 8 km.

For the geo-localization of crowd-sourced pictures, we used 120 k Street View images as a *reference image dataset*. The Street View images covered 180 km of city roads at intervals of 12 m. At each sampling point, Street View images were crawled at 45° intervals from the initial azimuth to 360° (see Fig. 1c). As Google Street View is not available in China, Tencent Street View was selected as a suitable alternative. As the first HD Street View map provider in China, Tencent Street View covers virtually all prefecture-level cities in China.

For each *reference image* (Street View image), its exterior orientation in the geographic coordinate is known. The longitude and latitude of the location and the yaw, pitch, and roll of the pose are predetermined during the process of crawling Street View images. Altitude is obtained by DEM.