Image and Vision Computing 52 (2016) 218–229

Contents lists available at ScienceDirect

Image and Vision Computing

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

Aerial image sequence geolocalization with road traffic as invariant feature☆

Gellért Máttyus[a,*], Friedrich Fraundorfer[a,b]

[a]German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany
[b]Graz University of Technology, Institute for Computer Graphics and Vision, Graz, Austria

ARTICLE INFO

Article history:
Received 2 December 2014
Received in revised form 25 February 2016
Accepted 27 May 2016
Available online 20 June 2016

Keywords:
Geolocalization
Georeferencing
Geotagging
Geometric hashing

ABSTRACT

The geolocalization of aerial images is important for extracting geospatial information (e.g. the position of buildings, streets, and cars) and for creating maps. The standard is to use an expensive aerial imaging system equipped with an accurate GPS and IMU and/or do laborious ground control point measurements. In this paper we present a novel method to recognize the geolocation of aerial images automatically without any GPS or IMU. We extract road segments in the image sequence by detecting and tracking cars. We search in a database created from a road network map for the best matches between the road database and the extracted road segments. Geometric hashing is used to retrieve a shortlist of matches. The matches in the shortlist are ranked by a verification process. The highest scoring match gives the location and orientation of the images. We show in the experiments that our method can correctly geolocalize the aerial images in various scenes: e.g. urban, suburban, and rural with motorway. Besides the current images only the road map is needed over the search area. We can search an area of 22,500 km^2 containing 32,000 km of streets within minutes on a single cpu.

1. Introduction

It is required for numerous applications that images and videos are tagged with their location on the earth's surface (geolocation). The geolocalization of aerial images is particularly essential for extracting geospatial information (e.g. the position of buildings, streets, and cars) and for creating maps. The aerial imaging systems typically include an accurate (and expensive) GPS and IMU. By avoiding these instruments a simple consumer camera could also be used to create georeferenced aerial images. People would be able to use their own camera during leisure and touristic activities (e.g. hang gliding, air balloon flight, sightseeing flight, and glider flight) to create georeferenced and orthorectified images. A crowd-sourced database of orthorectified aerial images (similar as OpenStreetMap for maps) could be used for mapping applications.

Although nowadays even consumer grade cameras can provide geotagging, it might be needed to find the geolocation of the camera, image and the objects in the scene based only on the visual information (see the IARPA Finder program[1]). Even with a GPS position tag the orientation of image is usually still unknown. Aerial image camera systems equipped with a very precise GPS and IMU also might lack the geolocation in case of an outage. The automatic geolocalization of these images can reduce expenses by avoiding a new flight or laborous manual work. It is also getting common to acquire images from a unmanned aerial vehicle (UAV). These can only carry light payloads, limiting the accuracy of the onboard GPS and IMU. Retrieving a more accurate image position and orientation during post processing could be needed.

UAVs also need a backup localization system in case of an outage of the GPS (GNSS). A GPS outage can happen either due to a problem in the device or because of external jamming of the GPS signal. The jamming of the GPS signal is a realistic threat, not only in military environments. The problem of GPS jamming is addressed in research, e.g. in [11].

Our goal is the geolocalization of aerial photo sequences on a large scale by utilizing only the image information and the road network. We propose to match the information acquired from the current

☆ This paper has been recommended for acceptance by Enrique Dunn.

* Corresponding author.
E-mail addresses: gellert.mattyus@dlr.de (G. Máttyus), fraundorfer@icg.tugraz.at (F. Fraundorfer).

1 http://www.iarpa.gov/Programs/ia/Finder/finder.html

http://dx.doi.org/10.1016/j.imavis.2016.05.014
0262-8856/

aerial image sequence to an object database invariant to the lighting and weather conditions, the road network map. Since the maintenance of an accurate, up-to-date road map is needed for many other applications (e.g. navigation and administration), it is more easily accessible than a large image database and requires significantly less storage[2].

We detect the road traffic in the image scene by tracking cars over the frames. By assuming the cars drive on the roads the vehicle trajectories can be interpreted as subsets of the roads, and be matched to the road map.

We propose a method for the fast retrieval of a shortlist containing the possible correct location. This retrieval is based on Geometric Hashing. We call it polyline based geometric hashing (PLBGHashing), and it can search rapidly over a larger search area. The more complex verification matching needs to be done only on the retrieved shortlist. The car tracks can be considered as a road detector with low completeness but high correctness. We combine this with a simple pixel color based road detector (with high completeness but low correctness) for the verification.

We analyzed the proposed PLBGHashing on synthetic data generated from the road network of two large cities. This confirmed that the pattern of the road network is discriminative on a larger scale. An evaluation was done on 20 image sequences captured over urban, suburban, rural with motorway, and industrial scenes by a consumer-grade camera mounted on an airplane, Fig. 1 shows the location of the scenes. In this test the PLBGHashing provided the shortlist, while the verification ranked the correct geolocation as the best match in most of the cases.

Our main contributions are as follows: (1) We use car tracks to detect parts of the road network. This does not need a known ground sampling distance (GSD[3]) as standard road detectors. (2) We utilize road networks as appearance invariant features to localize an aerial image sequence over a large area. This also avoids the need for an image database over the search area. (3) We present a geometric hashing method to match partial line segments to line structures. This does not need a complete road detector which detects relations between roads as a graph, e.g. intersections.

2. Related work

The standard for creating georeferenced aerial images is using an accurate GPS and IMU and/or measuring ground control points manually [15]. Other approaches utilize the idea to localize an image from only the image information also on a large scale.

The queried image can be matched to a huge pool of landmark photos [32], while the 3D building information might also be utilized [3]. Li et al. [17] determine the absolute (world) camera pose using 3D point clouds. Müller et al. [21] improve the absolute geometric accuracy of satellite images by automatically extracting ground control points from existing orthorectified reference images. In Wu et al. [30] satellite images are localized with feature-based indexing, but not on a real large scale (maximal 16 km×12 km). In Lin and Medioni [19] UAV images are matched to georeferenced ortho images (map). But the search area is limited by manually labeling several correspondences between the first frame of the UAV sequences and their corresponding satellite images.

All these methods require an already existing image database, and capturing and maintaining this is laborious. Using crowd-sourced image databases (e.g. Flickr, Google Picasa) this difficulty can be overcome. However, the landmark images are unevenly distributed and biased to touristic highlights. The work Lin et al. [18] addresses this problem by matching terrestrial image to remote sensing data, which also provide coverage over locations where no crowd-sourced data are available.

A new direction is the usage of non-image databases. Baatz et al. [4] employ the terrain map to geolocalize images from the mountain silhouette in the Alps.

Brubaker et al. [8] utilize the road network and visual odometry to locate a driving vehicle without any other information. They show that a long track on the road network is characteristic enough to obtain the geolocation of the car. Our method shares the idea of exploiting the pattern of the roads in the form of car tracks, but instead of a single long vehicle trajectory we use many short ones.

In Konzempel and Reulke [14] aerial images are orthorectified (projected on the Earth's surface) without an IMU. The orientation is initialized by an accurate GPS measurement and optimized by matching the detected streets in the image to the road network. In comparison to this method, our approach does not require an accurate position information, but only a search area, which might be as large as an entire metropolitan area.

In [25], road networks are represented as graphs, with road intersections as graph vertices. It is assumed that the intersections are detected correctly and the georegistration problem can be solved as a graph matching problem. In comparison to this work, in our approach the road intersections does not need to be detected in the image. We assume that there is no appropriate intersection detector or there might be no intersections in the image (just non intersecting roads).

In Wilson and Hancock [27] the graph structure of the road network (junctions and roads between them) in aerial image is extracted using a relaxational line-finder. Then the graph junctions are matched to junctions in the map using probabilistic relaxation. This method assumes that road junctions are present and they can be detected. In contrast, our method works without junctions, it needs only segments of the road network and we consider a much larger search area.

Li et al. [16] extract and match line segments instead of junctions. They define rotation and translation invariant features between the segments and use these pairwise features to calculate the cost of an assignment between aerial image and map. This combinatorial optimization problem is solved via continuous relaxation labeling. Solving this optimization problem is non trivial, specially for a large number of variables. In [16], the orientation of the image is used to limit the possible matches. In our problem we do not have access to this orientation information. We provide a more detailed analysis on this method in Section 4.2.3.

Gros et al. [10] use local, geometric invariant features to match images related by similarity or affine tranformations. For similarity transformations they compute geometric invariants from pairs of line segments having an endpoint in common. The invariants are the angle between the two segments and the ratio of the length of the segments. These invariants are matched between the images, similarity transformations are computed from the matches and they are aggregated in a Hough-transform manner, i.e. clusters are searched in the parameter space. In contrast to this method, in our problem the segment length ratio invariant can not be applied. A track segment can lay anywhere within the map segment. Additionally, our tracks are mostly straight (or with a small curvature), the angle between the consecutive line segments is usually around zero and thus the angle invariant is not discriminative enough.

The COCOA system [1] addressed the tracking of moving objects in aerial videos. Xiao et al. [31] track cars in wide field of view aerial videos, however they need already georeferenced images to leverage traffic flow priors from road maps.

[2] The OpenStreetMap is an open, crowd-sourced map with good coverage and fast updates.

[3] Ground sample distance, the distance between pixel centers measured on the ground.