Pattern Recognition 47 (2014) 2880–2893

Contents lists available at ScienceDirect

Pattern Recognition

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

Estimating the camera direction of a geotagged image using reference images

Minwoo Park [a,*], Jiebo Luo [b], Robert T. Collins [c], Yanxi Liu [c]

[a] Research and Engineering, ObjectVideo, United States
[b] Department of Computer Science, University of Rochester, United States
[c] Department of Computer Science and Engineering, The Pennsylvania State University, United States

ARTICLE INFO

Article history:
Received 11 July 2011
Received in revised form
29 September 2013
Accepted 3 March 2014
Available online 15 March 2014

Keywords:
GPS
Camera geometry
Geo-tagging
Viewing direction

ABSTRACT

Millions of smart phones and GPS-equipped digital cameras sold each year, as well as photo-sharing websites such as Picasa and Panoramio have enabled personal photos to be associated with geographic information. It has been shown by recent research results that the additional global positioning system (GPS) information helps visual recognition for geotagged photos by providing valuable location context. However, the current GPS data only identifies the camera location, leaving the camera viewing direction uncertain within the possible scope of 360°. To produce more precise photo location information, i.e. the viewing direction for geotagged photos, we utilize both Google Street View and Google Earth satellite images. Our proposed system is two-pronged: (1) visual matching between a user photo and any available street views in the vicinity can determine the viewing direction, and (2) near-orthogonal view matching between a user photo taken on the ground and the overhead satellite view at the user geo-location can compute the viewing direction when only the satellite view is available. Experimental results have shown the effectiveness of the proposed framework.

1. Introduction

With the explosion of photos and videos on the Internet, dealing with the large amount of unorganized visual data has become immensely challenging. To address this problem, one fast-emerging phenomenon in digital photography and community photo sharing is geo-tagging. The presence of geographically relevant metadata with photos and videos has opened up interesting research avenues in the multimedia research community for visual recognition of objects, scenes and events. For example, significant performance improvement in event recognition from photos can be achieved through the fusion of user photos and satellite images obtained using the global positioning system (GPS) information [1,2], while image annotation and image exploration can be enhanced using geotagged photos on the Internet [3,4].

However, the current GPS data only identifies the camera location while the interesting scene in the photo may not be at the specified geo-location; in fact it is often in the distance along an arbitrary viewing direction. Viewing direction data provided by a mobile device with a digital compass is typically unavailable, or otherwise error prone because the digital compass is sensitive to motion and magnetic disturbances. The importance of camera location and viewing direction has been recognized by many portable device manufacturers, such as Apple, Nikon, Nokia,[1] Ricoh, and Samsung, who have introduced (prototype) digital cameras and mobile phones that come with a GPS receiver and a digital compass.

GPS data associated with the photos taken by mobile devices are usually noisy also because GPS signals are weak in the proximity of tall buildings. These difficulties are further recognized by The 2009 and 2010 ACM Multimedia Grand Challenge [5] posed by Nokia where the primary goal is to derive the exact location and direction of a given photo with the aid of reference images. We note that there was no response to this particular challenge at the two past conferences.

In addition, the use of reference images has its own challenges because (1) reference images are not evenly distributed throughout the world, and (2) GPS data associated with the reference images found in the digital photo communities may be inaccurate and inconsistent (due to manual inputs). For example, some photos are associated with the GPS data at the interesting objects, some photos

* Correspondence to: Research and Engineering, ObjectVideo, 11600 Sunrise Valley Drive Suite 210 Reston, VA 20191 USA.
E-mail addresses: mpark@objectvideo.com (M. Park), jluo@cs.rochester.edu (J. Luo), rcollins@cse.psu.edu (R.T. Collins), yanxi@cse.psu.edu (Y. Liu).

[1] http://research.nokia.com/research/projects/mara/index.html.

http://dx.doi.org/10.1016/j.patcog.2014.03.002
0031-3203/

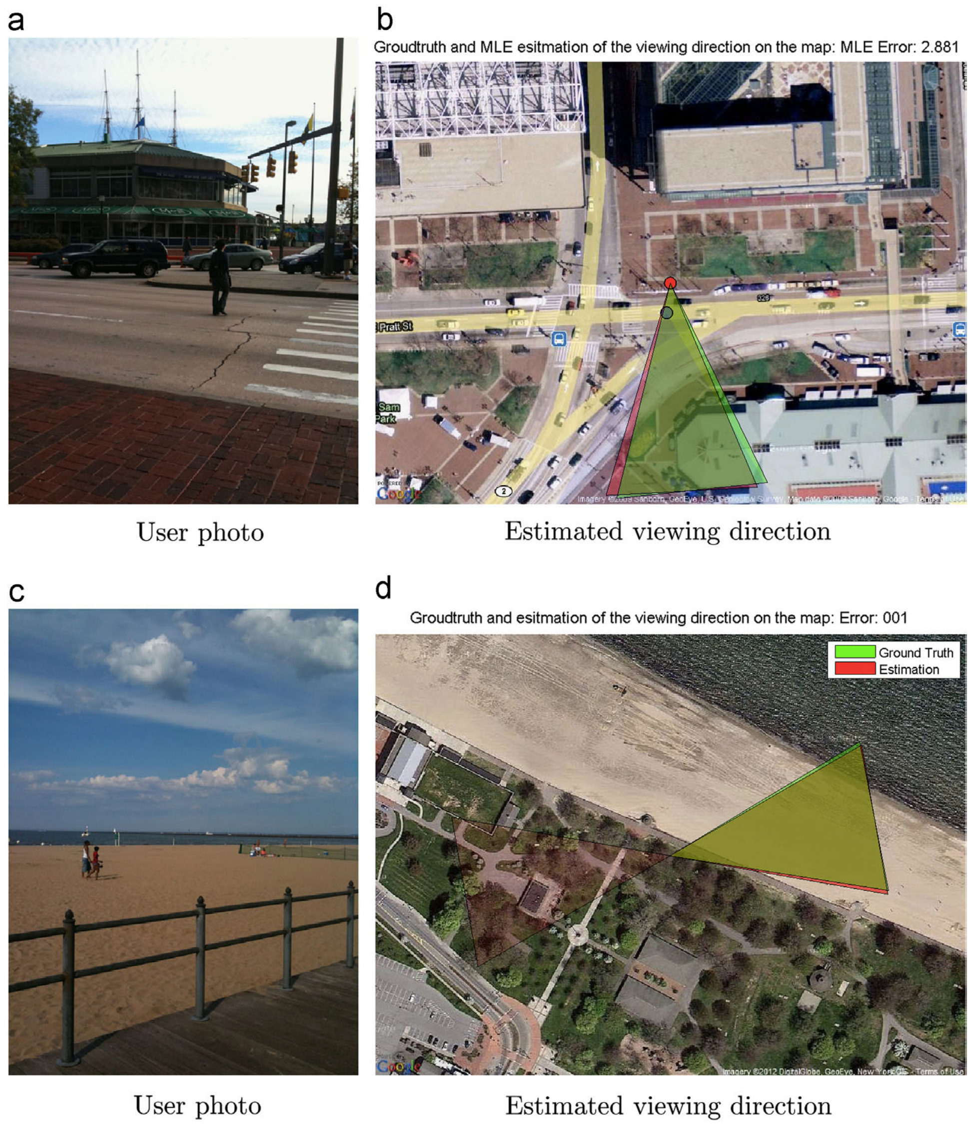

Fig. 1. The objective is to estimate the camera viewing directions. (a) and (c) Geotagged urban photo. (b) and (d) Geotagged suburban photo (green FOV triangle – ground truth, red FOV triangle – estimate). All figures are best viewed at 200% zoom on screen. Note that the red triangles for the estimated FOV can be covered by the green triangles for the ground truth FOV when the estimates are near perfect. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

are associated with GPS data at the user camera locations, and most tagged geo-locations are noisy. Indeed, estimating both geo-location and viewing direction simultaneously is an extreme problem.

In this study, the deficiency of the current GPS data and the scarcity of reference images are addressed by utilizing Google Street Views (covering major cities) when available and Google Earth satellite views (covering the entire globe) otherwise. Our goals are (1) to estimate the 2D viewing direction given GPS coordinates, and (2) to provide a general framework that can cover the entire world. Fig. 1 illustrates our goals with actual examples (with camera viewing directions estimated by the proposed algorithms) taken in both urban and suburban environments.

2. Related work

Snavely et al. [6,7] developed the Photo Tourism system for browsing large collections of photographs in 3D. Their system takes as input large collections of images from either personal photo collections or photo sharing web sites, and automatically computes each photo's viewpoint and a sparse 3D model of the scene. Their photo explorer interface then enables the viewer to interactively move about the 3D scene by seamlessly transitioning between photographs.

Later, Snavely et al. [8] also proposed a system where the goal is finding paths through the world's photos. When a scene is photographed many times by different people, the viewpoints often cluster along certain paths. These paths are largely specific to the scene being photographed, and traverse interesting regions and viewpoints. This work seeks to discover a range of such paths and turn them into control points for image-based rendering. Their approach again takes as input a large set of community or personal photos, reconstructs camera viewpoints, and automatically computes orbits, panoramas, canonical views, and optimal paths between views. The scene can then be interactively browsed in 3D using these controls or with five degree-of-freedom free-viewpoint control. However, the works introduced so far have not dealt with mapping of data back to actual maps. To address this problem automatically, Kaminsky et al. [9] proposed a method for aligning 3D point clouds with overhead images. They address the