Towards land cover map updating with ground-level panoramic photos

Nehla Ghouaiel, Nailya Bogrova, Sébastien Lefèvre

TOWARDS LAND COVER MAP UPDATING WITH GROUND-LEVEL PANORAMIC PHOTOS

Nehla Ghouaiel Nailja Bogrova Sébastien Lefèvre

Univ. Bretagne-Sud, UMR 6074 IRISA
Campus de Tohannic, 56000 Vannes, France
{nehla.ghouaiel,sebastien.lefevre}@univ-ubs.fr

KEY WORDS: Image Analysis, Remote Sensing, Change Detection, Crowdsourcing, Mapping, Mobile Computing

ABSTRACT:

Geographic landscapes in all over the world may be subject to rapid changes induced, for instance, by urban, forest and agricultural evolutions. Monitoring such kind of changes is usually achieved through remote sensing. However, obtaining regular and up-to-date aerial or satellite images is found to be a high costly process, thus preventing regular updating of land cover maps. Alternatively, in this paper, we propose a low-cost solution based on the use of ground-level geo-located landscape panoramic photos providing high spatial resolution information of the scene. Such photos can be acquired from various sources: digital cameras, smartphone, or even web repositories. Furthermore, since the acquisition is performed at the ground level, the users immediate surroundings, as sensed by a devices camera, can provide information at a very high level of precision, enabling to update the land cover type of the geographic area. In the described herein method, we propose to use inverse perspective mapping (inverse warping) to transform the geo-tagged ground-level 360°photo onto a top-down view as if it had been acquired from a nadiral aerial view. Once re-projected, the warped photo is compared to a previously acquired remotely sensed image using standard techniques such as correlation. Wide differences in orientation, resolution and geographical extent between the top-down view and the aerial image are addressed through specific processing steps (e.g., registration). Experiments on publicly available datasets made of both ground-level photos and aerial images show promising results for updating land cover maps with mobile technologies. The proposed approach contributes to the crowdsourcing efforts in geo-information processing and mapping, providing hints on the evolution of a landscape.

1. INTRODUCTION

The concept of Volunteered Geographic Information (VGI) refers to involving human volunteers in gathering photo collections that can be further used to feed geographical information systems. In fact, every human is able to act as an intelligent sensor, equipped with such simple aids as GPS and camera or even the means of taking measurements of environmental variables. As state by Goodchild (2007), "the notion that citizens might be useful and effective sources of scientifically rigorous observations has a long history, and it is only recently that the scientific community has come to dismiss amateur observation as a legitimate source" .

Over the past few years, VGI has become more available, for instance through web services. A range of new applications are being enabled by the georeferenced information contained in "repositories" such as blogs, wikis, social networking portals (e.g., Facebook or MySpace), and, more relevant to the presented work, community contributed photo collections (e.g. Flickr [1] or Panoramio [2]). The advantages of VGI are its temporal coverage, which is often better both in terms of frequency and latency than traditional sources. However, they come with a loss in data quality since user inputs are usually made available without review and without metadata (e.g. data source and characteristics). Georeferenced photo collections are enabling a new form of observational inquiry, which is termed "proximate sensing" by Leung and Newsam (2010). This concept depicts the act of using ground-level images of close-by objects and scenes rather than images acquired from airborne or satellite sensors.

While large collections of georeferenced photo collections have recently been available through the emergence of photo sharing websites, researchers have already investigated how these collections can help a number of applications. Research works in this context can be classified into two main categories according to Leung and Newsam (2010): i) using location to infer information about image and, ii) using images to infer information about a geographical location. In the first category, methods for clustering and annotating photos have been proposed Quack et al. (2008); Moxley et al. (2008). Images are labeled based on their visual content as depicting events or objects (landmarks). Other approaches (e.g., Hays and Efros (2008)) attempted to estimate the unconstrained location of an image based solely on its visual characteristics and on a reference dataset. In the second category, some researchers tried to address the problem of describing features of the surface of the earth. Examples of works in this area include: using large collections of georeferenced images to discover interesting properties about popular cities and landmarks such as the most photographed locations Crandall et al. (2009); creating maps of developed and undeveloped regions Leung and Newsam (2010), where the problem faced is then related to spatial coverage non-uniformity of images collections; computing country-scale maps of scenicness based on the visual characteristics and geographic locations of ground-level images Xie and Newsam (2011). Although Xie and Newsam (2011) demonstrated the feasibility of geographic discovery from georeferenced social media, they also reported the noisiness of obtained results.

The work presented in this paper is closely related to Leung and Newsam (2010); Xie and Newsam (2011) since we are here exploring content of georeferenced photos to infer geographic information about the locations at which they were taken. Conversely to existing work, we are not excluding available aerial or satellite images but we propose to use them in conjunction with recently available ground level images. The purpose of this work is therefore to update and checkup existing maps (built from

[1] https://www.flickr.com/
[2] http://www.panoramio.com/

standard remote sensing techniques) based on change detection performed with available ground level images. We thus investigate the application of proximate sensing to the problem of land cover classification. Rather than using only airborne/satellite imagery to determine the distribution of land cover classes for a given geographical area, we explore here whether ground level images can be used as a complementary data source. To do so, we present some preliminary work aiming to compare recently acquired ground level images to a previously acquired remotely sensed image using standard techniques related to computer vision and image analysis.

The remainder of this paper is organized as follows; Sec. 2. described the study area and the data set considered in the experiments. The technical approach is presented in Sec. 3.. We detail carried out experiments and discuss obtained results in Sec. 4. before providing some conclusions and directions for future research.

2. STUDY AREA AND DATA SET

The study area focuses on Vannes city in France, more precisely on the surroundings of the Tohannic Campus which hosts Université Bretagne Sud and IRISA research institute where are affiliated the authors. This choice is motivated by: i) the availability of ground truth that can be assessed by the in situ, and ii) the appearance of many new buildings over the last few years (with availability of data acquired both before and after these changes). It covers a 1 km^2 area. The geographical extent is provided in Fig. 1.

Figure 1: Aerial map from Bing Maps© with blue rectangles highlighting zones that have been recently transformed.

Ground level images were grabbed from Google Street View [3] or taken in-situ from people involved in this work equipped with mobile camera. Both kinds of images consist in panoramic views covering 360°(resp. 180°) field of view horizontally (resp. vertically). We assume here the following scenario: given the acquired image is georeferenced, it is possible to download an associated map from existing sources (Bing Maps, Google Maps, OpenStreetMap). We consider here maps of 150×150 m^2 downloaded through a Bing Maps [4] request according to measured GPS position.

For the sake of clarity, we denote the images with the following terms in the sequel:

- A: Aerial image, or high flying UAV image (dimensions $m \times n$).
- P: Panoramic image, or wide field-of-view image from user mobile device or Google Street View (dimensions $p \times q$).
- T: Top-down image, or bird's eye view of the ground (dimensions $r \times r$).

[3] https://www.google.com/maps/streetview/
[4] https://www.bing.com/maps/

Let us note that the proposed method is assessed here only on a single area, for which ground truth, aerial images, and in situ observations are available so to ease experimental validation. Experiments at a wider scale will be considered in future work.

3. PROPOSED METHOD

Since the images were taken from 3 different kinds of sensors (Google Street View's car, user camera and aerial vehicle) several image pre-processing steps are required before change detection can be performed. The flowchart of the proposed method including these different pre-processing steps is given in Fig. 2.

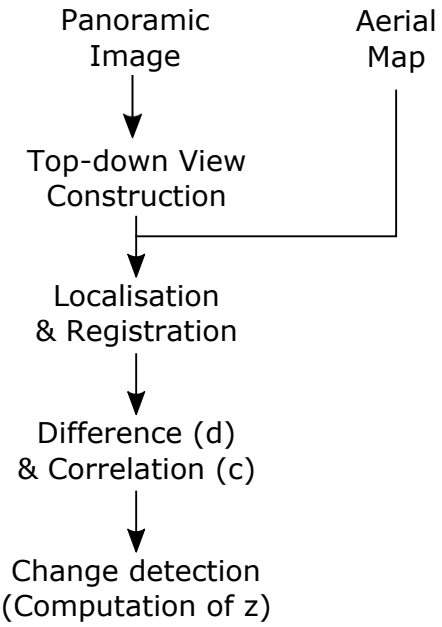

Figure 2: Flowchart of the proposed method.

3.1 Top-down View Construction

The panorama images used in this work cover $(360°, 180°)$ field of view on (horizontal,vertical) dimensions. For a given scene, the panorama image P is warped to obtain a bird's eye view T (as shown in Fig. 3). To do so, the 3D model of spherical image is first reconstructed using ground line following the method proposed by Xiao et al. (2012). Next, for each pixel in the top-down view, the position of the corresponding pixel in the panorama is computed using the inverse warping. The color for the ground location is then obtained using bi-linear interpolation from the panorama pixels.

Figure 3: Example of a top-down view T (right) constructed from a panorama image P (left).

3.2 Ground-level Image to Aerial View Registration

The next step aims to detect the area occupied by the top-down view in the aerial image. It is considered as a fine localization problem that can be formulated as matching image descriptors of the warped ground level image T with descriptors computed over the aerial map A. The proposed solution (see Fig. 4) is inspired from the work from Augereau et al. (2013). Various image descriptors are available to perform this matching. A recent study Viswanathan et al. (2014) comparing the performance of SIFT, SURF, FREAK, and PHOW in matching ground images onto a satellite map has shown that SIFT obtains the overall best performance, even with increasing complexity of the satellite map. We thus rely here on the SIFT descriptor Lowe (2004) in the matching process.

First, SIFT keypoints are detected and relative descriptors (feature vectors) are extracted for both aerial map A and top-down view T. Then, the similarity between the ground sample T descriptor vectors and each descriptor from A is computed. Each match is considered as correct or incorrect based on the Euclidean distance. To select the best match among candidate ones, we adopt the common approach relying on k-NN (nearest neighbor) classifier. Its complexity is however quadratic as a function of the number of keypoints. Using kd-tree for vector comparison improves search time in low dimensions. Using multiple randomized kd-trees has the advantage of speeding up k-NN search. FLANN Muja and Lowe (2009) provides an implementation of this algorithm where multiple trees are built in 5 random dimensions. Hence, FLANN library was used to achieve the process of keypoints matching.

In order to find the geometric transformation between matched keypoints, homography matrix H is computed Agarwal et al. (2005). At this level, RANSAC algorithm Fischler and Bolles (1981) is used in order to discard outliers. In fact, the aim of geometric transformation estimation step is to split the set of matches between good matchings (inliers) and mismatches (outliers) by using RANSAC algorithm. In order to estimate the 9-parameter transformation matrix H between key points of T denoted P_2 and their correspondences in H denoted P_2, the most representative transformation among all matches is sought. The matrix H has the following shape:

$$P_2 = HP_1 \tag{1}$$

or equivalently

$$\begin{bmatrix} x_2 \\ y_2 \\ w_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \times \begin{bmatrix} x_1 \\ y_1 \\ w_1 \end{bmatrix} \tag{2}$$

where locations of P_1 and P_2 are represented by homogeneous coordinates.

Finally, if at least t inliers are validated, T is considered to be situated in the aerial image A. We chose here $t = 4$, which is the minimum number of points necessary for homography computing. An illustration of this process is given in Fig. 5. We can observe that the technique is robust to a certain level of changes between the content visible in T and A.

3.3 Ground-level Image and Aerial View Comparison

Several change indices have already been proposed for estimating the change of appearance at two identical locations, from simple image difference or ratio to more elaborated statistics such as the Generalized-Likelihood Ratio Test (GLRT) Shirvany et al. (2010) or the local Kullback-Leibler divergence Xu and Karam (2013).

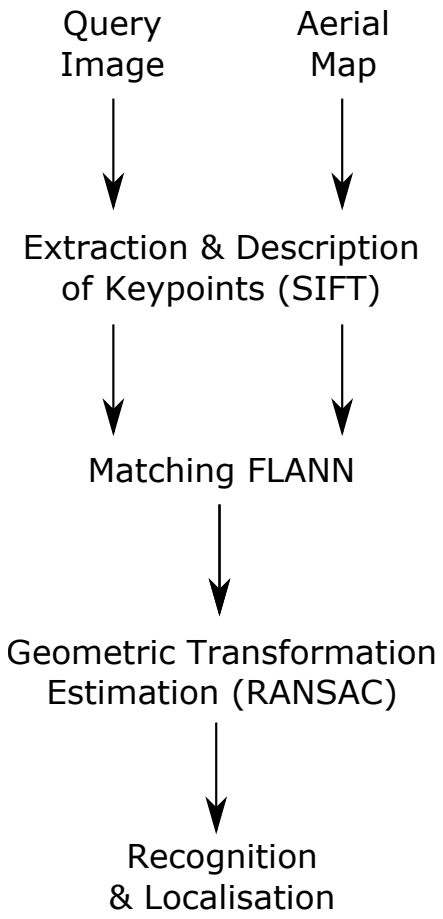

Figure 4: Standard object recognition and localization process.

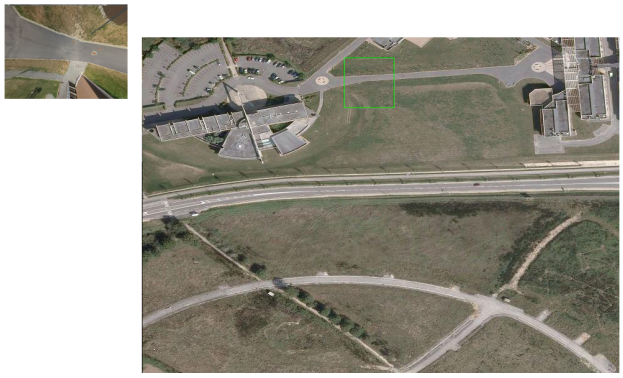

Figure 5: Example of top-down view localization in the aerial image. The top-down view T is the small image on the top left of the Figure, while the green area in the large aerial image A denotes the found localization of T.

For the sake of illustration, we have chosen here to rely on the well-known correlation index between the top-down view and the portion of the aerial map corresponding to it (see Sec. 3.2). The correlation coefficient r between two images a and b of size N is computed as follows:

$$r = \frac{N \sum_{i=1}^{N} I_{ai} I_{bi} - \sum_{i=1}^{N} I_{ai} \sum_{i=1}^{N} I_{bi}}{\sqrt{\left(N \sum_{i=1}^{N} I_{ai}^2 - (\sum_{i=1}^{N} I_{ai})^2 \right) \cdot \left(N \sum_{i=1}^{N} I_{bi}^2 - (\sum_{i=1}^{N} I_{bi})^2 \right)}} \quad (3)$$

where i is the pixel index, I_{ai} and I_{bi} are the intensities of the two images for pixel position i. In the correlation image, a low correlation value means a change. However, Liu and Yamazaki (2011) pointed that even if there was no change, some areas might be characterized by a very low correlation value. In this respect, they propose a new factor z used to represent changes, which combines the correlation coefficient r with the image difference d. The latter is defined by:

$$d = \overline{I}_{ai} - \overline{I}_{bi} \quad (4)$$

where $\overline{I}_{ai}$ and $\overline{I}_{bi}$ are the corresponding averaged values over a $M = k \times k$ pixels window surrounding the i-th pixel. We follow here a standard setting, where the window size is set as 9×9 pixels.

The factor z is then expressed by:

$$z = \frac{|d|}{\max_i(|d|)} - c \cdot r \quad (5)$$

where $\max_i(|d|)$ is the maximum absolute value of difference d among all pixel coordinates i, and c is the weight between the difference and the correlation coefficient. Following Liu and Yamazaki (2011), we weight the difference as 4 times the correlation in order to omit subtle changes, which means that c is set to 0.25.

A high value of z means high possibility of change. We adopt here the threshold value used by Liu and Yamazaki (2011) and consider the areas with $z > 0.2$ as changed areas.

4. EXPERIMENTS AND RESULTS

We recall that our method was evaluated with preliminary experiments on Vannes, France (see Sec. 2.). Aerial maps have been extracted from Bing Maps. Ground-based imagery have been either downloaded from Google Street View or captured in situ by some volunteers involved in these experiments. Aerial data date from 2011 while ground-level data were taken either in 2013 (Google Street View) or in 2015 (Google Street View or in situ observations). 22 significant locations were selected in the study site and therefore related ground-level P and aerial A images were included in the experiments. Figure 6 shows the 22 panoramic images used in our study.

To evaluate our approach, we distinguish between three different zones categories: unchanged unstructured (i.e. without building but possibly including roads), unchanged urban and new constructed areas. Each ground-level image is manually assigned to one of these categories based on its visual content and the information brought by the older aerial data. Obtained z values for these 22 images yield a variation between 0.10 and 0.34, with a change threshold set equals to 0.20 as in Liu and Yamazaki (2011).

Experimental results were analyzed through standard statistical measures and the confusion matrix is provided in Tab. 1. We can observe that the proposed method detect changed areas with a recall and precision scores of 72% and 45%, respectively. The F1-measure for this class is 55% (75% for unchanged unstructured areas and 46% for unchanged urban areas). Let us underline that recall is here more important than precision, since it is always possible to proceed with further manual inspection of potential changes. This emphasizes the feasibility of change detection by comparing ground level to aerial views. Most omission errors happen chiefly when only a few parts of buildings are appearing in the top-down view. To deal with this issue, another comparison step based on image features could be added. False positives are either still unbuilt areas or still built areas. Errors actually belonging to this first category are caused by the presence of cars or panels in the ground based image. Since buildings are seen from their roofs in the aerial view and from their sides or facades in the ground-level images, a lot of unchanged built areas are classified as changed by the proposed method. In the future work, this kind of errors would be removed by considering methods for aerial to ground building matching Bansal et al. (2011).

5. CONCLUSION

In the herein presented work, land changes are detected from comparing new acquired ground level images to less recent aerial images. To do so, we propose to transform the geo-tagged panoramic photo onto a top-down view as if it had been acquired from a nadiral aerial view. Once re-projected, the warped photo is compared to a previously acquired remotely sensed image using a technique combining correlation coefficient and image difference. The obtained results confirm the efficiency of the described method in addressing the presented issue, with a specific use case being the detection of new built areas.

In the aim of enhancing current results, we will consider more advanced images comparison methods and will complete our preprocessing pipeline by other steps such as photometric correction. Other future works include enlarging geographic extent of the study area and increasing the volume of test data and metrics. The final goal would be to perform land cover updating with our method, to illustrate the strength of crowdsourcing as an ancillary but important information source for geo-information management.

References

Agarwal, A., Jawahar, C. V. and Narayanan, P. J., 2005. A survey of planar homography estimation techniques. Technical Report IIIT/TR/2005/12.2005, Centre for Visual Information Technology, International Institute of Information Technology.

Augereau, O., Journet, N. and Domenger, J.-P., 2013. Semi-structured document image matching and recognition. In: SPIE Conference on Document Recognition and Retrieval, Vol. 8658, pp. 1–12.

Bansal, M., Sawhney, H. S., Cheng, H. and Daniilidis, K., 2011. Geo-localization of street views with aerial image databases. In: ACM International Conference on Multimedia, pp. 1125–1128.

Crandall, D. J., Backstrom, L., Huttenlocher, D. and Kleinberg, J., 2009. Mapping the world's photos. In: International Conference on World Wide Web, pp. 761–770.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image

Figure 6: Panorama images used in experimental evaluation. Images from the 6 top rows were grabbed from Google Street View while the 2 bottom rows correspond to images acquired through crowdsourcing with mobile camera.

	unchanged unstructured	unchanged urban	new constructed areas	total labeled
unchanged unstructured	6	0	2	8
unchanged urban	0	3	4	7
new constructed areas	2	0	5	7
total classified	8	3	11	22

Table 1: Confusion matrix

analysis and automated cartography. Communications of the ACM 24(6), pp. 381–395.

Goodchild, M. F., 2007. Citizens as sensors: web 2.0 and the volunteering of geographic information. International Review of Geographical Information Science and Technology 7, pp. 8–10.

Hays, J. and Efros, A. A., 2008. Im2gps: estimating geographic information from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8.

Leung, D. and Newsam, S., 2010. Proximate sensing: Inferring what-is-where from georeferenced photo collections. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2955–2962.

Liu, W. and Yamazaki, F., 2011. Urban monitoring and change detection of central Tokyo using high-resolution X-band SAR images. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 2133–2136.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, pp. 91–110.

Moxley, E., Kleban, J. and Manjunath, B. S., 2008. Spirittagger: a geo-aware tag suggestion tool mined from flickr. In: ACM International Conference on Multimedia Information Retrieval, pp. 23–30.

Muja, M. and Lowe, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: International Conference on Computer Vision Theory and Applications, pp. 331–340.

Quack, T., Leibe, B. and Van Gool, L., 2008. World-scale mining of objects and events from community photo collections. In: International Conference on Content-based Image and Video Retrieval, pp. 47–56.

Shirvany, R., Chabert, M., Chatelain, F. and Tourneret, J.-Y., 2010. Maximum likelihood estimation of the polarization degree from two multi-look intensity images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1198–1201.

Viswanathan, A., Pires, B. R. and Huber, D., 2014. Vision based robot localization by ground to satellite matching in gps-denied situations. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 192–198.

Xiao, J., Ehinger, K., Oliva, A. and Torralba, A., 2012. Recognizing scene viewpoint using panoramic place representation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2695–2702.

Xie, L. and Newsam, S., 2011. Im2map: Deriving maps from georeferenced community contributed photo collections. In: ACM SIGMM International Workshop on Social Media, pp. 29–34.

Xu, Q. and Karam, L. J., 2013. Change detection on SAR images by a parametric estimation of the KL-divergence between gaussian mixture models. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1109–1113.