



HAL
open science

Coupling ground-level panoramas and aerial imagery for change detection

Nehla Ghouaiel, Sébastien Lefèvre

► To cite this version:

Nehla Ghouaiel, Sébastien Lefèvre. Coupling ground-level panoramas and aerial imagery for change detection. *Geo-spatial Information Science*, 2016, Special Issue: Mobile Mapping with Ubiquitous Point Clouds, 19 (3), pp.222-232. 10.1080/10095020.2016.1244998 . hal-01391926

HAL Id: hal-01391926

<https://hal.science/hal-01391926v1>

Submitted on 13 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coupling ground-level panoramas and aerial imagery for change detection

Nehla Ghouaiel  and Sébastien Lefèvre 

Institute of Research in Computer Science and Random Systems, Université de Bretagne-Sud, UMR 6074 IRISA, Vannes, France

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 camera device, 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 nadir 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 data-sets made of both ground-level photos and aerial images show promising results for updating land cover maps with mobile technologies. Finally, the proposed approach contributes to the crowdsourcing efforts in geo-information processing and mapping, providing hints on the evolution of a landscape.

ARTICLE HISTORY

Received 29 March 2016
Accepted 30 July 2016

KEYWORDS

Image analysis; remote sensing; change detection; crowdsourcing; mobile mapping; panoramic photos

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 stated 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 (<https://www.flickr.com/>) or Panoramio (<http://www.panoramio.com/>)). The advantages of VGI are its temporal coverage, which is often better both in terms of frequency and latency than traditional sources. How-

ever, 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 properties). 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 photos 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 (Moxley, Kleban, and Manjunath 2008; Quack, Leibe, and Van Gool 2008). Images are labeled based on their visual content as depicting events or objects (landmarks). Other approaches such as Hays and Efros (2008) attempted to estimate the

CONTACT Sébastien Lefèvre  sebastien.lefevre@irisa.fr

© 2016 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

unconstrained location of an image based solely on its visual characteristics and on a reference data-set. 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); or more recently recognizing geo-informative attributes of a photo, e.g. the elevation gradient, population density, demographics using deep learning (Lee, Zhang, and Crandall 2015; Workman, Souvenir, and Jacobs 2015). 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. But here we are not excluding available aerial or satellite images. Instead, we propose to use them in conjunction with recently available ground-level images. The purpose of this work is therefore to update and check up existing maps (built from 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 first 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. In this context, our work share some similarity with Murdock, Jacobs, and Pless (2013; 2015), where it is proposed to use webcam videos together with satellite to estimate cloud maps. However, conversely to these previous studies, we do not use aerial imagery solely for training the ground-based image recognition process. Our goal is rather to perform comparison between both available data sources.

The remainder of this paper is organized as follows: Section 2 describes the study area and the data-set considered in the experiments. The technical approach is presented in Section 3. We detailedly carried out experiments and discuss obtained results in Section 4

before providing some conclusions and directions for future research.

2. Study area and data set

The study area focuses on several cities in France: Vannes, Rennes, and Nantes in Brittany, and Dijon in Burgundy. For the sake of conciseness, we provide visual illustrations for the Vannes city only, but reported accuracy includes the whole data-set. Experiments on Vannes city focused on the surroundings of the Tohannic Campus which hosts Université Bretagne Sud and IRISA research institute where the authors are affiliated. This choice is motivated by: (i) the availability of ground truth that can be assessed by in situ observations 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-km² area. The geographical extent is provided in Figure 1.

Ground-level images were grabbed from Google Street View (<https://www.google.com/maps/streetview/>) 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, and OpenStreetMap). We consider here maps of 150 × 150 m² downloaded through a Bing Maps (<https://www.bing.com/maps/>) 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 $r \times r$).

Beyond Vannes city for which it was both possible to build ground truth from in situ observation and acquire panoramic photos with crowdsourcing activities, we also consider three other data-sets to evaluate the robustness of our methods. On these other cities, panoramic photos are grabbed from Google Street View and were subsequently used to build ground truth through visual interpretation.

3. Proposed method

Since the images were taken from up to three different kinds of sensors (Google Street View's car, user's camera and aerial vehicle), several image preprocessing



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

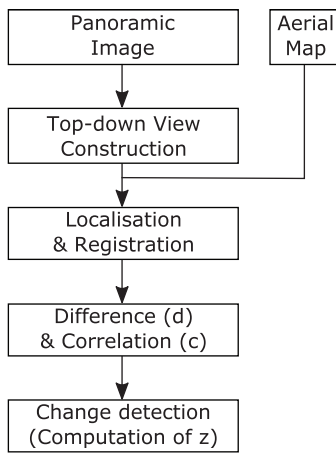


Figure 2. Flowchart of the proposed method.

steps are required before change detection can be performed. The flowchart of the proposed method including these different preprocessing steps is given in Figure 2.

3.1. Top-down view construction

The panorama images used in this work cover $(360^\circ, 180^\circ)$ 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 Figure 3) following the method proposed by Xiao et al. (2012). First, the world coordinates of the panoramic image P are computed using the inverse perspective mapping technique. To do so, the 3D model of the image P is generated using the following equations (Xiao 2012):

$$X = r \cos\left(-\theta_x + \frac{\pi}{2}\right) \quad (1)$$

$$Z = r \sin\left(-\theta_x + \frac{\pi}{2}\right) \quad (2)$$

$$Y = \begin{cases} 0, & \text{if points are lying on the ground} \\ c^h + r \tan \theta_y, & \text{otherwise} \end{cases} \quad (3)$$

where $r = c^h |\cot \theta_y|$, $\theta_y = \pi y/H$ is the angle between the optical axis and the horizon and $\theta_x = 2\pi x/W$ represents the angle between the projection of the optical axis on the flat plane ($y = 0$) and z axis. Finally, c^h denotes the camera height (in meters) from the ground plane.

In the next step, the obtained 3D image is projected onto the plane $Y = 0$ to get a new image T representing the top-down view of the original image. This remapping process produces image $T(u, v)$ by recovering the texture of the ground plane as shown by the following equations (Muad et al. 2004):

$$u(X, 0, Z) = \frac{\gamma(X, 0, Z) - (\theta_x - \alpha)}{\frac{2\alpha}{W-1}} \quad (4)$$

$$v(X, 0, Z) = \frac{\theta(X, 0, Z) - (\theta_y - \alpha)}{\frac{2\alpha}{H-1}} \quad (5)$$

where α is the camera angular aperture, $W \times H$ are the image dimensions, $\theta = \tan^{-1}(c^h/(x^2 + Z^2)^{0.5})$ and $\gamma = \tan^{-1}(Z/X)$. The color for the ground location is obtained using bi linear interpolation from the panorama pixels.

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 Figure 4) is inspired from the work from Augereau, Journet, and Domenger (2013). Various image descriptors are available to perform this matching. A recent study (Viswanathan, Pires, and Huber 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

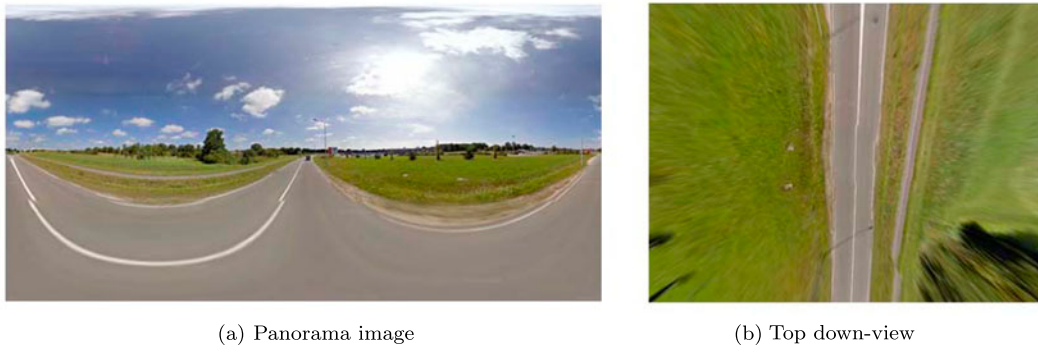


Figure 3. Example of a top down-view T (b) constructed from a panorama image P (a).

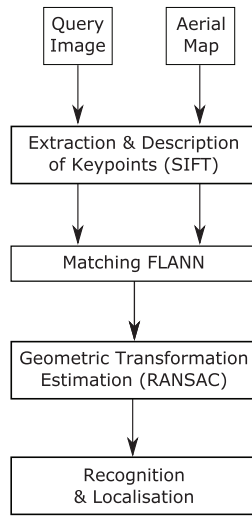


Figure 4. Standard object recognition and localization process.

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 q and each descriptor p from A is computed. Each match $m(q_i, p_j)$ is considered as correct or incorrect based on the Euclidean distance $|\cdot, \cdot|_2$.

$$m(q_i, p_j) = \arg \min_j |q_i, p_j|_2 \quad (6)$$

In order 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. The multiple randomized kd-trees algorithm (Silpa-Anan and Hartley 2008) has the advantage of speeding up k-NN search. Thus, we used FLANN (Muja and Lowe 2009) library that provides an implementation of this algorithm where multiple kd-trees are searched in parallel. We note that for the randomized k-d trees, the split dimension is chosen randomly from the top 5 dimensions with the highest variance.

In order to find the geometric transformation between matched keypoints, homography matrix H is computed by (Agarwal, Jawahar, and Narayanan 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) using RANSAC algorithm. In order to estimate the 9-parameter transformation matrix H between key points of T denoted P_1 and their correspondences in A denoted P_2 , the most representative transformation among all matches is sought. The matrix H has the following shape:

$$P_2 = H P_1 \quad (7)$$

or equivalently

$$\begin{bmatrix} x_2 \\ y_2 \\ w_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \times \begin{bmatrix} x_1 \\ y_1 \\ w_1 \end{bmatrix} \quad (8)$$

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 Figure 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).

For the sake of illustration, we have chosen here to rely on the well-known correlation index between



(a) Top down-view



(b) Large aerial image

Figure 5. Example of top-down view localization in the aerial image. The top-down view T (a) is compared with the large aerial image A (b), and the green area denotes the found localization of T .

the top-down view and the portion of the aerial map corresponding to it (see Section 3.2). The correlation coefficient r between two images a and b of size N is computed as follows:

$$r = \frac{N \sum_{i=1}^N I_{ai} I_{bi} - \sum_{i=1}^N I_{ai} \sum_{i=1}^N I_{bi}}{\sqrt{\left(N \sum_{i=1}^N I_{ai}^2 - \left(\sum_{i=1}^N I_{ai} \right)^2 \right) \cdot \left(N \sum_{i=1}^N I_{bi}^2 - \left(\sum_{i=1}^N I_{bi} \right)^2 \right)}} \quad (9)$$

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:

$$d = \bar{I}_{ai} - \bar{I}_{bi} \quad (10)$$

where \bar{I}_{ai} and \bar{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:

$$z = \frac{|d|}{\max_i (|d|)} - c \cdot r \quad (11)$$

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 several cities in France (see Section 2), namely Vannes, Nantes, and Rennes in Brittany, and Dijon in Burgundy.

Aerial images 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 (for the Vannes site only). Aerial images are dated from 2011 to 2012 while ground-level data were taken either in 2013 (Google Street View) or in 2015 (Google Street View or in situ observations).

Table 1. Confusion matrix for the changed/unchanged classification.

	Unchanged	Changed	Total labeled
Unchanged	34	38	72
Changed	8	20	28
Total classified	42	58	100

100 significant locations were selected for the study site and therefore related ground-level P and aerial A images were included in the experiments.

Figure 6 shows the 100 panoramic images used in our study, while Figure 7 shows the 100 corresponding aerial images. As we can see, the data-set shows significant differences in landscapes and visual content.

Let us recall that our goal is to explore how ground-based (possibly crowdsourced) imagery could help to perform change detection in terms of land cover/land use. Visual interpretation has thus been conducted on the whole set of images (i.e. both a ground-based panorama and an aerial image for each of the 100 locations) to label each location as changed or unchanged. Obtained z values for these 100 images yield a variation between 0.00528137 and 0.417598, 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 Table 1 from which are derived producer accuracy (recall) and user accuracy (precision), as well as overall accuracy (ratio of correctly classified elements among all elements). We can observe that the proposed method achieves an overall accuracy of 54%, but with significant difference between recall and precision rate for the two different classes. More interestingly, the method shows a rather high recall rate (71%) for the changed class, at the cost of a lower precision (34%) though. In other words, a location for which a change in land cover/land use occurs is barely to be missed by our method. This is highly preferred to other situations, e.g. missing changes that can be recovered only by manual analysis of the whole data-set. Conversely, manual refinement of the results consists in filtering out false positives only.

In order to assess more precisely the behavior of the method, we have performed a finer classification, where we distinguished between structured areas containing built-up objects and unstructured ones. Thus, each ground-level image is manually assigned to one of these two classes based on its visual content. This reference classification is then compared against an automatic procedure inspired from Leung and Newsam (2010). To do so, we compute lines descriptors on each image to quantify the distribution of edges at different orientations. Indeed, images of structured areas have a higher proportion of horizontal and vertical lines than images of unstructured scenes. Hough transform

Table 2. Confusion matrix for the unchanged/changed/structured classification.

	Unchanged unstructured	Unchanged structured	Changed structured	Total labeled
Unchanged unstructured	25	0	10	35
Unchanged structured	3	6	28	37
Changed structured	3	5	20	28
Total classified	31	11	58	100

Table 3. Computational complexity of the different steps composing our proposed approach. CPU times have been averaged among 100 runs.

Step	Time (ms)
Top-down view construction	877.80
Localisation & registration	2834.62
Difference & correlation & change detection	90.08

(Hough 1962) was applied to detect lines at roughly horizontal, vertical, 45° diagonal, 135° diagonal, and isotropic (non-orientation) directions. Therefore, each panoramic image is represented by a five dimensional line feature vector. Then, we use a Hidden Conditional Random Field (Quattoni, Collins, and Darrell 2004) classifier to label individual images based on their line descriptors.

We report in Table 2 the confusion matrix for this second classification experiments. Let us note that unchanged unstructured areas have been extracted from Vannes and Dijon cities, while unchanged structured areas and changed structured areas are both coming from Vannes, Rennes, and Nantes cities.

Again, we are focusing on changed (structured) areas. We can observe that including a structured/unstructured preclassification step allows to achieve better accuracy. Indeed, considering only images containing built-up structures, the recall for changed areas is reaching 80%, for a precision of 42%. More generally, we can see that the misclassification between changed/unchanged areas is more important with structured areas than unstructured ones.

Beyond accuracy evaluation, we have also measured the computational efficiency of the proposed approach. The goal is to assess its usability in a crowdsourcing context. To do so, we have averaged computation time over 100 runs, considering a standard PC workstation (CPU: i7-4600@2.10 GHz, RAM: 8 GB). Results are reported in Table 3. We can observe that CPU times are very low, the overall process being performed in 3–4 s. This makes the proposed approach a realistic crowdsourcing solution for change detection.

5. Discussion

An in-depth analysis of situations where the proposed method was failing to identify land use/land cover changes was thus performed. We thus observe the strong



Figure 6. Panorama images used in experimental evaluation.



Figure 7. Aerial images used in experimental evaluation.

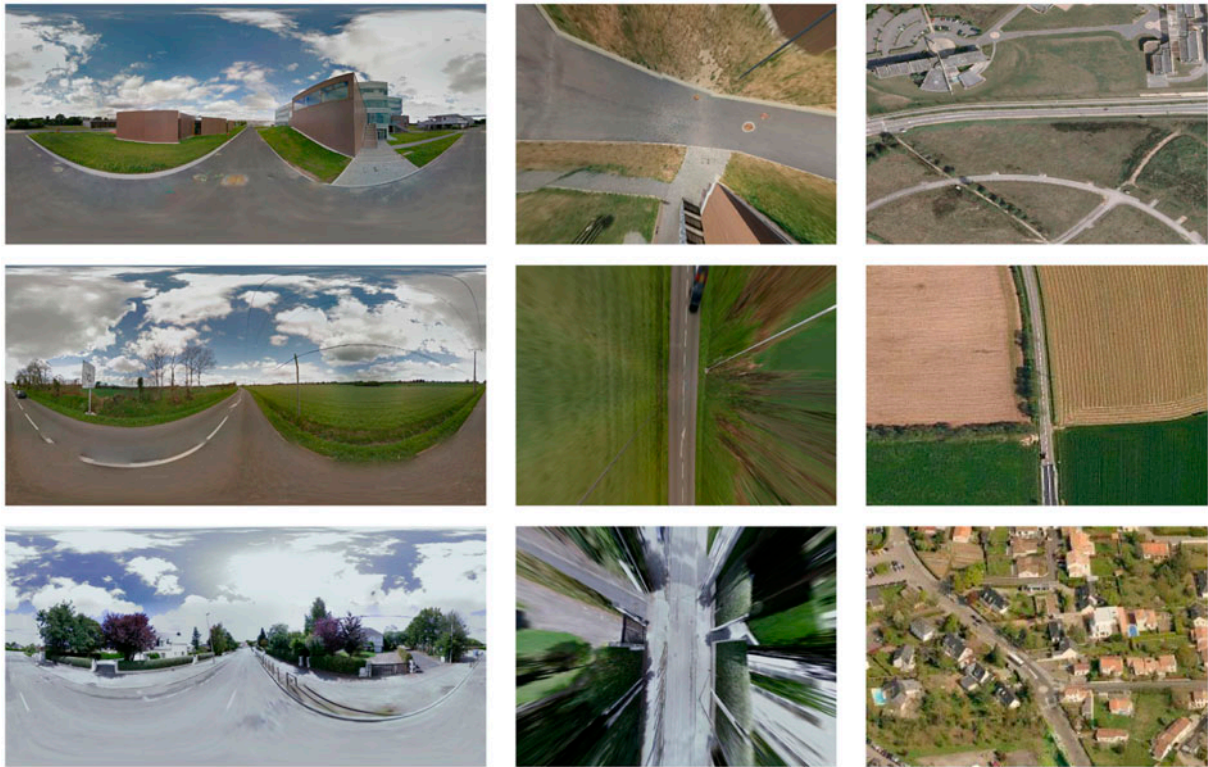


Figure 8. Artifacts brought by built-up structures (from top to bottom): missed changed, false changes due to other structured objects, false changes due to deformation brought by top-down view. For each line are given (from left to right): ground panorama, top-down view, and corresponding aerial image.

effect played by artifacts brought by built-up structures in top-down views, as shown in Figure 8. Critical omission situations were faced when only a few parts of buildings are appearing in the top-down view (Figure 8, first line). Let us note that this issue could be solved through another comparison step based on image features. On the other hand, false positives coming from unstructured areas are caused by the presence of cars or panels in the ground-based image. When mapped on the top-down view, these objects have a similar appearance as built-up structures (Figure 8, second line). Moreover, 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 structured areas also lead to false positives being classified as changed by the proposed method (Figure 8, third line). In the future work, this kind of errors would be removed by considering methods for aerial to ground building matching (Bansal et al. 2011).

6. Conclusions

In the herein presented work, land use/land cover 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 reprojected,

the warped photo is compared to a previously acquired remotely sensed image using a technique combining correlation coefficient and image difference. We have conducted an experiment including 100 images from four different cities in France. The obtained results show a high recall rate for the changed areas, with nevertheless a lower precision rate. 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. Besides, a more careful analysis distinguishing between structured and unstructured areas has been performed to understand the current bottlenecks of the proposed method.

In the aim of enhancing current results, we will now consider more advanced images comparison methods and will complete our preprocessing pipeline by other steps such as photometric correction. Comparing ground-based and aerial imagery is still a challenging issue, as noticed by a recent study from Loschky et al. (2015). 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.

Notes on contributors

Nehla Ghouaiel is a research fellow at the Institute of Research in Computer Science and Random Systems (IRISA), Université de Bretagne-Sud. Her current research interests include image analysis, machine learning, and computer vision.

Sébastien Lefèvre is a professor at the Institute of Research in Computer Science and Random Systems (IRISA), Université de Bretagne-Sud. His current research interests include image analysis and machine learning for remote-sensing data.

ORCID

Nehla Ghouaiel  <http://orcid.org/0000-0001-7419-1004>

Sébastien Lefèvre  <http://orcid.org/0000-0002-2384-8202>

References

- Agarwal, A., C. V. Jawahar, and P. J. Narayanan. 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., N. Journet, and J. P. Domenger. 2013. "Semi-structured Document Image Matching and Recognition." In *SPIE Conference on Document Recognition and Retrieval*, 1–12, Vol. 8658, Burlingame, CA, February 3–4.
- Bansal, M., H. S. Sawhney, H. Cheng, and K. Daniilidis. 2011. "Geo-localization of Street Views with Aerial Image Databases." In *ACM International Conference on Multimedia*, 1125–1128, New York: ACM.
- Crandall, D., L. Backstrom, D. Huttenlocher, and J. Kleinberg. 2009. "Mapping the World's Photos." In *International Conference on World Wide Web*, 761–770. New York: ACM.
- Fischler, M., and R. Bolles. 1981. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24 (6): 381–395.
- Goodchild, M. 2007. "Citizens as Sensors: Web 2.0 and the Volunteering of Geographic Information." *International Review of Geographical Information Science and Technology* 7: 8–10.
- Hays, J., and A. Efros. 2008. "Im2gps: Estimating Geographic Information from a Single Image." In *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 24–26. 1–8.
- Hough, P. 1962. A Method and Means for Recognizing Complex Patterns. US Patent 3,069,654.
- Lee, S., H. Zhang, and D. J. Crandall. 2015. "Predicting Geo-informative Attributes in Large-scale Image Collections Using Convolutional Neural Networks." In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 550–557, Waikoloa Beach, HI, January 6–9.
- Leung, D., and S. Newsam. 2010. "Proximate Sensing: Inferring What-is-where from Georeferenced Photo Collections." In *IEEE Conference on Computer Vision and Pattern Recognition*, 2955–2962, San Francisco, CA, June 13–18.
- Liu, W., and F. Yamazaki. 2011. "Urban Monitoring and Change Detection of Central Tokyo Using High-resolution X-band SAR Images." In *IEEE International Geoscience and Remote Sensing Symposium*, 2133–2136, Vancouver, July 24–29.
- Loschky, L., R. Ringer, K. Ellis, and B. Hansen. 2015. "Comparing Rapid Scene Categorization of Aerial and Terrestrial Views: A New Perspective on Scene Gist." *Journal of Vision* 15 (6): 11. doi:10.1167/15.6.11.
- Lowe, D. 2004. "Distinctive Image Features from Scale-invariant Keypoints." *International Journal of Computer Vision* 60: 91–110.
- Moxley, E., J. Kleban, and B. S. Manjunath. 2008. "Spirittagger: A Geo-aware Tag Suggestion Tool Mined from Flickr." In *ACM International Conference on Multimedia Information Retrieval*, 23–30, Vancouver, October 30–31.
- Muad, A., A. Hussain, S. Abdul Samad, M. Mustaffa, and B. Majlis. 2004. "Implementation of Inverse Perspective Mapping Algorithm for the Development of an Automatic Lane Tracking System." In *IEEE TENCON*, 207–210, Vol. 1, Chiang Mai, Thailand, November, 21–24.
- Muja, M., and D. Lowe. 2009. "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration." In *International Conference on Computer Vision Theory and Applications*, 331–340, Lisbon, Portugal, February 5–8.
- Murdock, C., N. Jacobs, and R. Pless. 2013. "Webcam2Satellite: Estimating Cloud Maps from Webcam Imagery." In *IEEE Workshop on Applications of Computer Vision (WACV)*, 214–221, Clearwater Beach, FL, January 15–17.
- Murdock, C., N. Jacobs, and R. Pless. 2015. "Building Dynamic Cloud Maps from the Ground Up." In *IEEE International Conference on Computer Vision (ICCV)*, 684–692, Santiago, Chile, December 11–18.
- Quack, T., B. Leibe, and L. Van Gool. 2008. "World-scale Mining of Objects and Events from Community Photo Collections." In *International Conference on Content-based Image and Video Retrieval*, 47–56, Niagara Falls, Canada, July 7–9.
- Quattoni, A., M. Collins, and T. Darrell. 2004. "Conditional Random Fields for Object Recognition." In *NIPS (Conference and Workshop on Neural Information Processing Systems)*, 1097–1104, Whistler, BC, Canada, December 17.
- Shirvany, R., M. Chabert, F. Chatelain, and J. Y. Tourneret. 2010. "Maximum Likelihood Estimation of the Polarization Degree from Two Multi-look Intensity Images." In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1198–1201, Dallas, TX, March 14–19.
- Silpa-Anan, C., and R. Hartley. 2008. "Optimised KD-trees for Fast Image Descriptor Matching." In *IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, Vol. 60, Anchorage, AK, June 24–26.
- Viswanathan, A., B. R. Pires, and D. Huber. 2014. "Vision Based Robot Localization by Ground to Satellite Matching in GPS-denied Situations." In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 192–198, Chicago, IL, September 14–18.
- Workman, S., R. Souvenir, and N. Jacobs. 2015. "Wide-area Image Geolocalization with Aerial Reference Imagery." In *IEEE International Conference on Computer Vision (ICCV)*, 3961–3969, Santiago, Chile, December 11–18.
- Xiao, J. 2012. "3D Geometry for Panorama." Accessed March 16 2016. <http://panocontext.cs.princeton.edu/panorama.pdf>
- Xiao, J., K. Ehinger, A. Oliva, and A. Torralba. 2012. "Recognizing Scene Viewpoint Using Panoramic Place Representation." In *IEEE Conference on Computer Vision and Pattern Recognition*, 2695–2702, Providence, RI, June 16–21.

Xie, L., and S. Newsam. 2011. “IM2MAP: Deriving Maps from Georeferenced Community Contributed Photo Collections.” In *ACM SIGMM International Workshop on Social Media*, 29–34, Scottsdale, AZ, November 28–December 1.

Xu, Q., and L. Karam. 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*, 1109–1113, Vancouver, May 26–31.