Mountain Peak Identification in Visual Content Based on Coarse Digital Elevation Models

Roman Fedorov
Politecnico di Milano
Dipartimento di Elettronica,
Informazione e Bioingegneria
Milan, Italy
roman.fedorov@polimi.it

Piero Fraternali
Politecnico di Milano
Dipartimento di Elettronica,
Informazione e Bioingegneria
Milan, Italy
piero.fraternali@polimi.it

Marco Tagliasacchi
Politecnico di Milano
Dipartimento di Elettronica,
Informazione e Bioingegneria
Milan, Italy
marco.tagliasacchi@polimi.it

ABSTRACT

We present a method for the identification of mountain peaks in geo-tagged photos. The key tenet is to perform an edge-based matching between the visual content of each photo and a terrain view synthesized from a Digital Elevation Model (DEM). The latter is generated as if a virtual observer is located at the coordinates indicated by the geo-tag. The key property of the method is the ability to reach a highly accurate estimation of the position of mountain peaks with a coarse resolution DEM available in the corresponding geographical area, which is sampled at a spatial resolution between 30 m and 90 m. This is the case for publicly available DEMs that cover almost the totality of the Earth surface (such as SRTM CGIAR [4] and ASTER GDEM [10]). The method is fully unsupervised, thus it can be applied to the analysis of massive amounts of user generated content available, e.g., on Flickr and Panoramio. We evaluated our method on a dataset of manually annotated images of mountain landscapes, containing peaks of the Italian and Swiss Alps. Our results show that it is possible to accurately identify the peaks in 75.0% of the cases. This result increases to 81.6% when considering only photos with mountain slopes far from the observer.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Object recognition*; I.4.6 [**Image Processing and Computer Vision**]: Segmentation—*Edge and feature detection*

Keywords

mountain peak identification;User Generated Content; environmental monitoring;geo-tagging;skyline detection

MAED'14, November 7, 2014, Orlando, Florida, USA.
Copyright 2013 ACM 978-1-4503-3123-4/14/11 ...$15.00.
http://dx.doi.org/10.1145/2661821.2661825.

1. INTRODUCTION

Photo hosting platforms and social networks are reaching nowadays unprecedented diffusion in terms of the number of publicly available user generated photographs. An increasing part of these photos is geo-tagged and contains snapshots of the environment in which we live. For example, Flickr has collected more than 8 billion images, with more than 3.5 million new daily uploads [6]. Panoramio has reached 75 million geo-tagged images [1], which mostly contain outdoor landscapes, monuments, etc.

In this paper, we are particularly interested in the analysis of photographs taken in mountain regions. A large fraction of them contains the skyline defined by mountain peaks, slopes, ridges, crests, etc., both as main subject and as background. As such, these photographs implicitly contain precious information related to, e.g., environmental phenomena, which has not been fully exploited so far. For example, the analysis of visual content might reveal the snow cover at different altitudes, in regions where ground measurements are not available. This, in turn, is an extremely important information needed for the prediction of the Snow Water Equivalent (SWE) available to the cities, industries and agriculture [9].

Therefore, it can be very useful to extract this latent environmental information from photographs of the mountains, as they are constantly being added to public repositories. A crucial step to obtain these results is to uniquely and automatically identify mountain peaks in visual content, so as to collect series of photographs containing the same peak, and track its appearance over time.

The main contribution of this paper is a method that, given a geo-tagged photograph and a 360° panoramic view (henceforth called *panorama*) of the terrain synthesized from a Digital Elevation Model (DEM), is able to identify the peaks in the photograph and assign them a label (e.g., *Matterhorn*) which comes with the annotated DEM. The proposed method has some key properties that make it suitable for the analysis of large-scale datasets. First, it is fully unsupervised. Second, it provides accurate results also when using coarse DEMs (at a spatial resolution between 30 m and 90 m) that are made available to the public at no cost, e.g., SRTM CGIAR [4] and ASTER GDEM [10].

This is exemplified in Figure 1a, which clearly demonstrates some of the challenging aspects that need to be addressed: i) due to the limited spatial resolution of the DEM, the contours of the panorama are oversmoothed; ii) due to GPS errors, there might be a mismatch (highlighted in

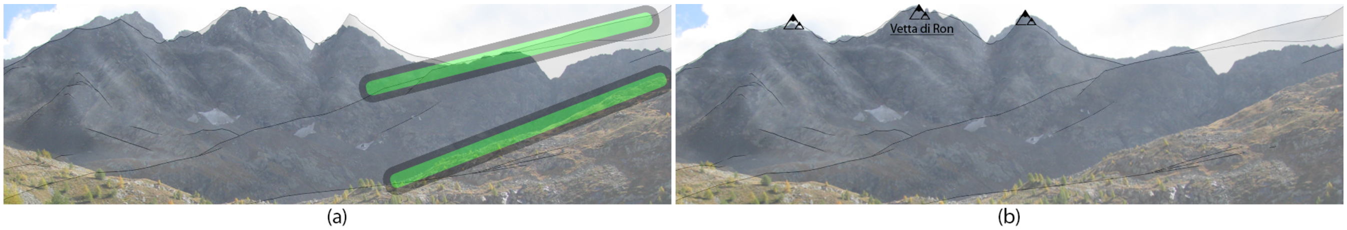

Figure 1: An example of matching between a photograph and a panoramic view synthesized from a DEM. Output of (a) global alignment, with highlighted the mismatch with the DEM; (b) local alignment.

green) between the contours, especially for mountain slopes close to the observer. In general, the mountain profiles of the photos and the panorama might not align exactly, thus requiring a robust matching method.

2. RELATED WORK

The problem of identifying mountain peaks in user generated content has recently attracted the attention of the research community. Baboud et al. [3] propose an algorithm for photo-to-terrain alignment based on a DEM. However, the method is not quantitatively evaluated on a large dataset, and qualitative results are provided only for 28 photographs. The examples reported in the paper reveal a very accurate alignment with the terrain, indicating the use of a high-resolution DEM.

Baatz et al. [2] approach a related problem, that is, the estimation of the geographical position of mountain photographs in the absence of geo-tags by means of content based analysis. However, they do not address how to determine the labels of the mountain peaks. In addition, in some of the examples, the sky-to-terrain segmentation is performed manually, before the photograph is processed by the algorithm.

Liu and Su [8] present an image content search method based on the shape of the skyline. The idea is to match two photographs which contain the same peaks, similarly to landmark search in urban environments. However, labeling of mountain peaks is not supported.

A preliminary version of our work was presented by the authors of this paper in [5]. This work contains a number of novel contributions. First, the global alignment method has been improved, by cascading a refinement step that increases the overall performance. Second, the method is made robust in the presence of clouds, by introducing a skyline detection algorithm. Third, the evaluation is now based on a large number of photographs, which were manually annotated, so that it is possible to observe the impact of clouds and other obstacles on the overall performance. Fourth, different measures are proposed to rigorously evaluate the various steps of the method, which also capture the fine-grained alignment of the individual peaks.

Unlike [3], we provide a quantitative evaluation on a significantly larger dataset and introduce different adjustments in the preprocessing and alignment algorithm, needed when copying with photos taken in diverse weather conditions and in the presence of other objects (trees, mountain slopes in the foreground, etc.). In addition, we adopt a coarse resolution DEM, which is publicly available. Conversely, [2] is based on an extremely precise DEM available only for Switzerland ($swissALTI^{3D}$: $2m$ spatial resolution), and it is not obvious how similar results can be achieved in a different area.

3. MOUNTAIN PEAK IDENTIFICATION

To understand the amount of photographs of mountain landscapes available on content sharing platforms, we crawled from Flickr a 300 × 160 kilometer region across the Italian and Swiss Alps (in the area of Pennine Alps, Lepontine Alps, Rhaetian Alps and Lombard Prealps) downloading 600k photographs with a valid geo-tag. We carried out a crowdsourcing experiment on a random sample of images taken at an elevation of 600 meters or above, obtaining that approximately 21% of the photographs contain a distinctive skyline of the mountain profile.

The proposed method aims at annotating these photographs with the labels of the mountain peaks that appear in them. To this end, given a photograph and the additional metadata extracted from the EXIF container (geo-tag, focal length, camera model and manufacturer), it is possible to perform a matching with a 360° panoramic view of the terrain synthesized from a Digital Elevation Model (DEM). The method proceeds in four steps, which are described below and illustrated in Figure 2.

Preprocessing: The horizontal Field Of View (FOV) of the photograph is calculated from the focal length and the size of the camera sensor. Then, the photograph is rescaled considering that the width of the panorama corresponds to a FOV equal to 360°. This is necessary to ensure that the photograph and the panorama have the same scale in degrees/pixel and matching can be performed without resorting to scale invariant methods.

Due to the different nature of the photograph and the panorama (Figure 2a), it is not possible to exploit conventional descriptors, e.g., color, texture or local features. However, it is still possible to rely on the edges to match the images. Hence, we apply an edge extraction algorithm to both the photograph and the panorama to produce an edge map, which assigns to each pixel the strength of the edge at that point and its direction (Figure 2b).

Global alignment: A skyline detection algorithm is applied [7] to the photos and all the edges above the skyline are removed being considered obstacles or clouds. Next, a filtering procedure is applied to the edges of the photo, by decreasing the strength of the edge points as the vertical position decreases (Figure 2c - top). This procedure allows us to decrease the relevance of edge points that do not correspond to terrain contours, since they might be due to people, buildings, vegetation and other objects. As for the panorama, the edges corresponding to the skyline can

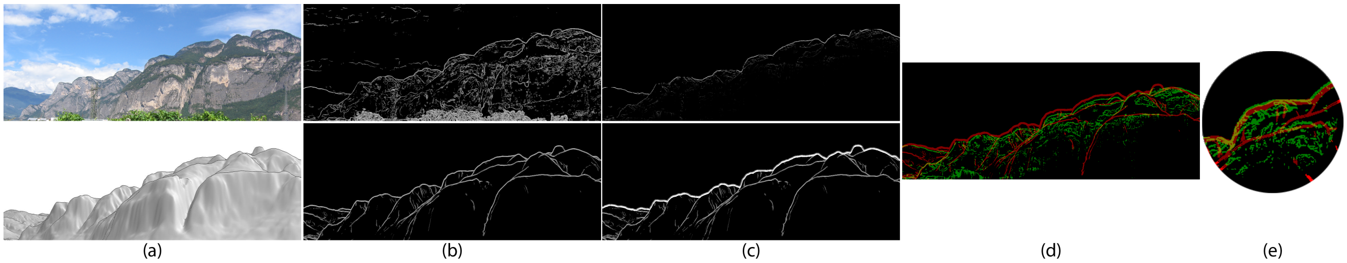

Figure 2: A schematic example of the peak identification method: (a) input photograph (top) and corresponding panorama (bottom), (b) edge extraction, (c) skyline detection, filtering and dilation (d) global alignment with refinement (e) local alignment.

be identified as the upper envelope of the edge map, by keeping, for each column of pixels, the topmost edge point. Then, a morphological dilation is applied to emphasize the edges corresponding the skyline (Figure 2c - bottom).

The matching between the photograph and the corresponding panorama is performed using a Vector Cross Correlation (VCC) technique proposed in [3], which takes into account both the strength and the direction of the edge points. The output of the VCC is a correlation map that, for each possible horizontal and vertical displacement between the photograph and the panorama, indicates the strength of the matching. Then, the top-K local maxima of the correlation map are identified as candidate matches.

Global alignment can match mountain edges also below the skyline and is robust with respect to skyline detection errors. However, the global maximum of the correlation is not necessarily the correct match. This might occur, for example, when some edges of the photograph happen to match the shape of different portions of the panorama. As such, the top-K matches are further analyzed by the refinement step below.

Refining global alignment: For each of the top-K candidate matches, we measure the Hausdorff distance between the skyline edge points of the photo and of the panorama, when the two are overlapped at the candidate matching position. A scoring function is computed, which combines the Hausdorff distance and the rank position computed by the initial global alignment. The candidate with the highest score is then chosen as the best match between the photo and the panorama (Figure 2d).

Local alignment: Our method generates a panorama from a coarse DEM, using a possibly noisy geo-tag. Therefore, in most cases the panorama does not match the photo perfectly, thus increasing the difficulty in finding a correct global alignment. This is clearly shown in Figure 1a: first, the rightmost part of the photograph does not match the actual skyline, due to the occlusion of a mountain slope close to the virtual observer; second, it is not possible to simultaneously match all the three peaks in the leftmost part of the photograph by means of a simple rigid displacement. Therefore, to improve the precision of the position of each mountain peak (Figure 1b), a separate VCC procedure is applied, similar to the one used in the global alignment step. Specifically, for each peak we consider a local neighborhood centered in the location identified by the global alignment. In this way each peak position is refined by identifying the best match in this local neighborhood (Figure 2e).

4. EXPERIMENTS

Dataset: Our method was tested on a set of photographs selected from those crawled in the monitored region. We manually inspected a subset of 200 photographs and the panoramas generated based on the accompanying EXIF metadata to make sure that a plausible matching existed. Indeed, in some cases, we found that the geo-tag was available but incorrect, such that the generated panorama could not be matched to the photograph by any means. Finally, we retained 162 photographs in our test set. Then, the ground truth data was generated by an alignment tool developed ad-hoc, which allows the user to find the correct position of the photograph in the panorama and then to locally warp the image by overlapping each mountain peak present in the photo to the corresponding one in the panorama.

Measures: For each peak $i = 1, \dots n$, let (x_i^p, y_i^p) and (x_i^r, y_i^r), denote the pixel coordinates in the coordinate system of the photo and of the panorama, respectively. When the photo is aligned with a displacement $(\Delta x, \Delta y)$, we define the angular error in the position of the i-th peak as

$$\epsilon_i(\Delta x, \Delta y) = \sqrt{d_x(x_i^r, \Delta x + x_i^p)^2 + d_y(y_i^r, \Delta y + y_i^p)^2}\,, \quad (1)$$

where

$$d_x(x_1, x_2) = (360/w_r)\min(w_r - |x_1 - x_2|, |x_1 - x_2|)$$

is the angular distance (in degrees) between two points along the azimuth, given the circular symmetry of the panorama, and w_r is the number of pixels corresponding to 360°. Similarly

$$d_y(y_1, y_2) = (360/w_r)|y_1 - y_2|$$

where the same angular resolution in degrees/pixel is assumed due to small elevation angles. When creating the ground truth, the images are warped so as to minimize the average angular error

$$\epsilon(\Delta x, \Delta y) = (1/n)\sum_{i=1}^{n} \epsilon_i(\Delta x, \Delta y)$$

and to find the best displacement

$$(\Delta x^*, \Delta y^*) = \arg\min_{\Delta x, \Delta y} \epsilon(\Delta x, \Delta y)$$

Note that $\epsilon^* = \epsilon(\Delta x^*, \Delta y^*)$ cannot always be reduced to 0, due to the coarse nature of the panorama.

Let $(\Delta x_k^G, \Delta y_k^G)$, $k = 1, \dots, K$, denote the displacements of the top-K candidate matches of global alignment. We define $p_K^G(\theta)$ as the fraction of the photos in the test set that

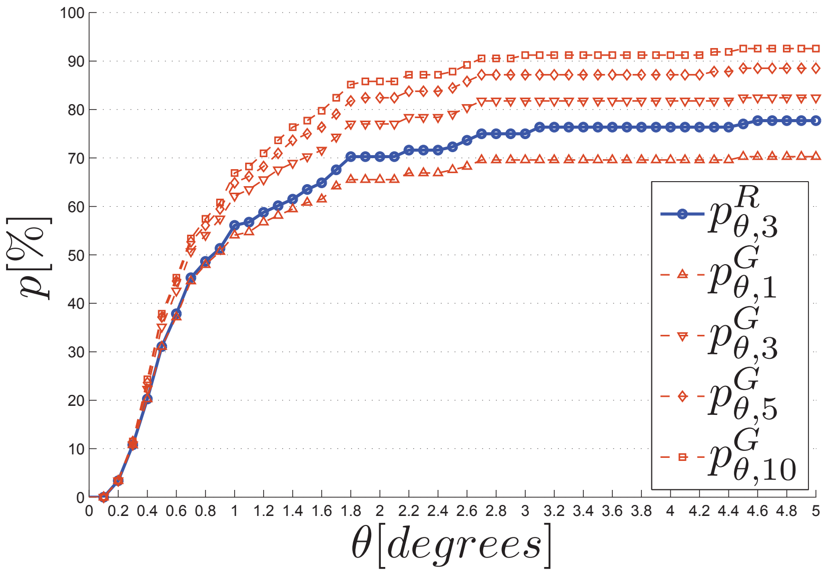

Figure 3: Performance analysis of the global alignment and of the refinement step

have at least one candidate match displacement $(\Delta x^G_k, \Delta y^G_k)$ lying within angular distance θ from the ground truth $(\Delta x^*, \Delta y^*)$. The refinement step selects $(\Delta x^R_K, \Delta y^R_K)$ to be one of the displacements $(\Delta x^G_k, \Delta y^G_k)$ (not necessarily the best). Then, $p^R_K(\theta)$ is the fraction of photographs for which the difference between $(\Delta x^R_K, \Delta y^R_K)$ and $(\Delta x^*, \Delta y^*)$ is below θ. Note that $p^R_K(\theta) \leq p^G_K(\theta)$ by construction, and the equality holds if the refinement step is always able to identify the correct match within the top-K candidates.

The local alignment step computes a different displacement $(\Delta x^L_i, \Delta y^L_i)$ for each of the n peaks. Then, the average error is defined as

$$\epsilon^L = (1/n) \sum_{i=1}^{n} \epsilon_i(\Delta x^L_i, \Delta y^L_i)$$

Results: Figure 3 shows the performance of the global alignment on the whole dataset. It can be observed that $p^G_K(\theta)$ saturates when θ exceeds 3°. Specifically, 69.6% of the photographs are aligned with an average error below 3°, when considering the top-1 match. The fraction of correctly aligned photos grows to 81.8%, 87.2% and 91.2% when K is 3, 5 and 10, respectively. Diminishing returns in the average error are observed when increasing K; thus, we selected $K = 3$ in the refinement step by trial and error method, which results in 78% of correctly aligned photos. The refinement performance curve lies approximately halfway between the top-1 and top-3 curves of global alignment. This shows the benefit of introducing the refinement step and its ability to pick the correct candidate from the top-3 candidates.

Taking a deeper look into the dataset, Table 1 describes the performance of the proposed method depending on the different properties of the visual content, manually annotated in two ways; first, we marked whether the photograph contains clouds (80 out of 162); second, we marked the presence of mountains close to the observer that might occlude the skyline in the background (49 out of 162). The presence of clouds is one of the main obstacles to be addressed. This is due to the fact that, when clouds partially occlude the skyline, the outcome of the skyline detection algorithm might fail. In addition, edge points due to clouds above the skyline might compromise the filtering procedure, since the latter is based on the assumption that there are no edges above the skyline. In the case of global alignment, the fraction of correctly matched photographs grows to 72.4% and 82.9% in the absence of clouds, when considering the top-1 and top-3 candidates, respectively. Conversely, the presence of clouds leads to a reduction of correct matches, which represent, however, at least 66.7% of the cases. The performance of the refinement step is also affected by the presence of clouds, being equal to 77.6% (72.2%) when clouds are absent (present). The impact of clouds is higher in the refinement step than in the top-3 candidates global alignment, since the former relies heavily on the correctness of the estimated skyline. Another issue lies in the presence of mountain slopes nearby the observer. Indeed, in this case small errors in the geo-tag might lead to a panorama which does not correctly represent the viewpoint of the photograph. This situation is clearly visible in the rightmost part of Figure 1a. In the case of global alignment, the fraction of correctly matched photographs grows to 74.8% and 89.3% in the absence of nearby mountains, when considering the top-1 and top-3 candidates, respectively. A similar behaviour is observed for the refinement step (81.6%).

Local alignment further improves the matching between the photograph and the panorama. This is measured by comparing the average angular error between the peak positions after the refinement step, $\epsilon(\Delta x^R_K, \Delta y^R_K)$, with the one obtained after local alignment, ϵ_L. In our experiments, we found that the error decreased from $\epsilon(\Delta x^R_K, \Delta y^R_K) = 0.99°$

	$p^G_{3,1}$	$p^G_{3,3}$	$p^R_{3,3}$
Whole dataset	69.6%	81.8%	75.0%
Absence of clouds	72.4%	82.9%	77.6%
Presence of clouds	66.7%	80.6%	72.2%
Absence of nearby mountains	74.8%	89.3%	81.6%
Presence of nearby mountains	57.8%	64.4%	60.0%

Table 1: Performance results decomposed by dataset categories and photograph content properties

to $\epsilon_L = 0.78°$, i.e., a 21% reduction with the radius of the local neighborhood set to 7.5°.

Unfortunately it was not possible to compare our results with those obtained by other algorithms discussed earlier, due to the lack of a publicly available dataset and unspecified quantitative evaluation metrics [3]. Instead, [2] and [8] address different problems (respectively, geo-tag estimation and relevant image retrieval) and cannot be compared directly with our work.

5. CONCLUSIONS AND FUTURE WORK

We presented a method for the identification of mountain peaks in geo-tagged photographs, which might find use in the automatic annotation of UGC. In our tests the method was able to correctly estimate the alignment between the photograph and the panorama in 75.0% of the cases. Mountain peaks were tagged with an average angular error of 0.78°. Due to the unsupervised nature of the method, we envisage its use in the automatic analysis of massive quantities of UGCs and its exploitation for environmental monitoring.

6. ACKNOWLEDGMENTS

This work is supported by

- EU FP7 CUbRIK Integrating Project, http://www.cubrikproject.eu
- POR-FESR 2007-2013 PROACTIVE Project, http://www.proactiveproject.eu

7. REFERENCES

[1] Panoramio. [Official API results on an entire globe query].

[2] G. Baatz, O. Saurer, K. Köser, and M. Pollefeys. Large scale visual geo-localization of images in mountainous terrain. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part II*, ECCV'12, pages 517–530, Berlin, Heidelberg, 2012. Springer-Verlag.

[3] L. Baboud, M. Cadik, E. Eisemann, and H.-P. Seidel. Automatic photo-to-terrain alignment for the annotation of mountain pictures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), oral presentation*, 2011.

[4] T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, et al. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45, 2007.

[5] R. Fedorov, D. Martinenghi, M. Tagliasacchi, and A. Castelletti. Exploiting user generated content for mountain peak detection. In *Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation (SoHuman 2013)*, pages 21–28, 2013.

[6] A. Jeffries. The man behind flickr on making the service 'awesome again', March 2013. [Online; posted 20-March-2013].

[7] W.-N. Lie, T. C.-I. Lin, T.-C. Lin, and K.-S. Hung. A robust dynamic programming algorithm to extract skyline in images for navigation. *Pattern Recognition Letters*, 26(2):221 – 230, 2005.

[8] W.-H. Liu and C.-W. Su. Automatic peak recognition for mountain images. In Y.-M. Huang, H.-C. Chao, D.-J. Deng, and J. Park, editors, *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*, volume 260 of *Lecture Notes in Electrical Engineering*, pages 1115–1121. Springer Netherlands, 2014.

[9] J. L. McCreight and E. E. Small. Modeling bulk density and snow water equivalent using daily snow depth observations. *The Cryosphere*, 8(2):521–536, 2014.

[10] D. Meyer. Aster global digital elevation model version 2 - summary of validation results, August 2011. [Online; posted 31-August-2011].