Camera Elevation Estimation from a Single Mountain Landscape Photograph

Martin ČADÍK,[1] Jan VAŠÍČEK,[2] Michal HRADIŠ,[1] Filip RADENOVIĆ,[3] Ondřej CHUM[3]

[1]{cadik, ihradis}@fit.vutbr.cz, [2]xvasic21@stud.fit.vutbr.cz,
[3]{filip.radenovic, chum}@cmp.felk.cvut.cz

FACULTY OF INFORMATION TECHNOLOGY
Brno University of Technology
Brno, Czech Republic
http://cphoto.fit.vutbr.cz/

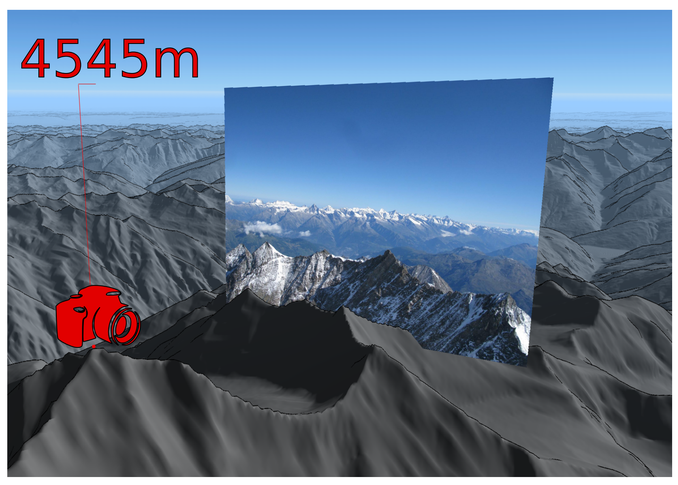
Faculty of Electrical Engineering
Czech Technical University
Prague, Czech Republic
http://cmp.felk.cvut.cz/

INTRODUCTION

In outdoor environments one of the important attributes is the *elevation*: the height of a geographic location above the sea level. Elevation data are important for a number of applications, including earth sciences, climate change research, hydrology, outdoor navigation and localization. Traditionally the assessment of elevation was the domain of geodesy, which offered several means to measure altitude. This work addresses the problem of camera elevation estimation from a single photograph.

Contributions:

- a new dataset: Alps100K
- experiment evaluating human performance of elevation estimation
- methods for automatic elevation estimation from image content
  - based on convulational neural networks (CNN)
  - based on bag-of-words (BOW) image retrieval
  - hybrid combination of CNN



AUTOMATIC ELEVATION ESTIMATION

Convolutional neural network (CNN) was initialized from Placess-CNN [2] and then finetuned on the Alps100K dataset. While the original architecture mimics Caffe reference network, we replaced the second fully-connected layer fc7 with a randomly initialized layer with 2048 neurons, and layer fc8 was reduced to a single neuron representing the predicted elevation. The network was trained by mini-batch Stochastic Gradient Descent with momentum and MSE objective function.

Sparse high-dimensional bag-of-words (BOW) approach using inverted file to efficiently retrieve images performs well for specific object and place recognition, especially when combined with a spatial verification step.

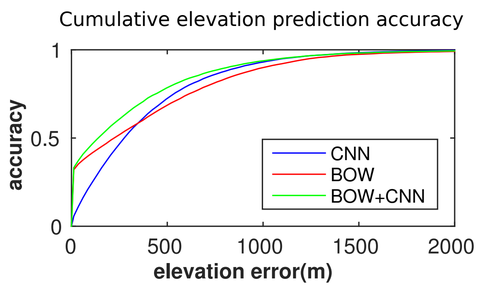
Compact image representations (mVocab) [1] image retrieval approach using a joint dimensionality reduction from multiple vocabularies shows certain level of generalization power.

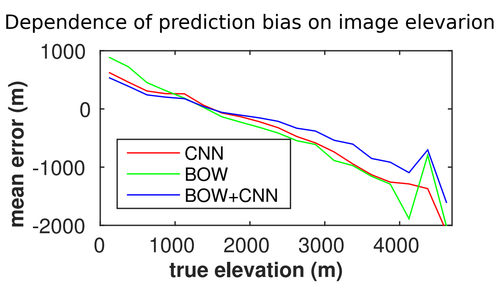
Hybrid method tries to estimate the elevation by recognizing the location (using BOW), and if that fails, i.e., no spatially verified image is retrieved, then by a secondary estimator, either mVocab or CNN.

EXPERIMENTAL RESULTS

Test: 13148 images (13% of Alps100K); **Training:** 84988 images (rest)
Performanse measure: overall root-mean-square error (RMSE) of elevation predictions with regards to the known ground truth elevations

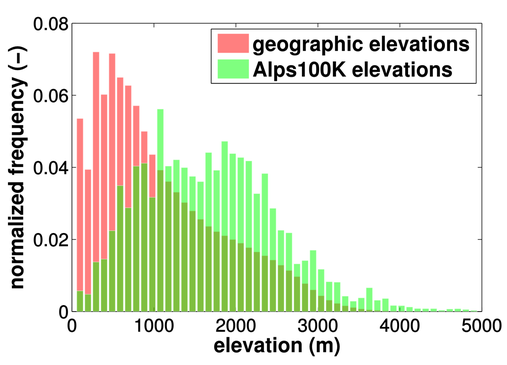
Method	test dataset (13148 images)	user experiment set (50 images)
Baseline	801.49; 786.42	1383.64; 1154.43
Human	-	879.95
CNN	537.11	709.10
BOW	601.63	757.76
mVocab	610.36	811.00
BOW+mVocab	564.14	646.89
BOW+CNN	500.44	531.05





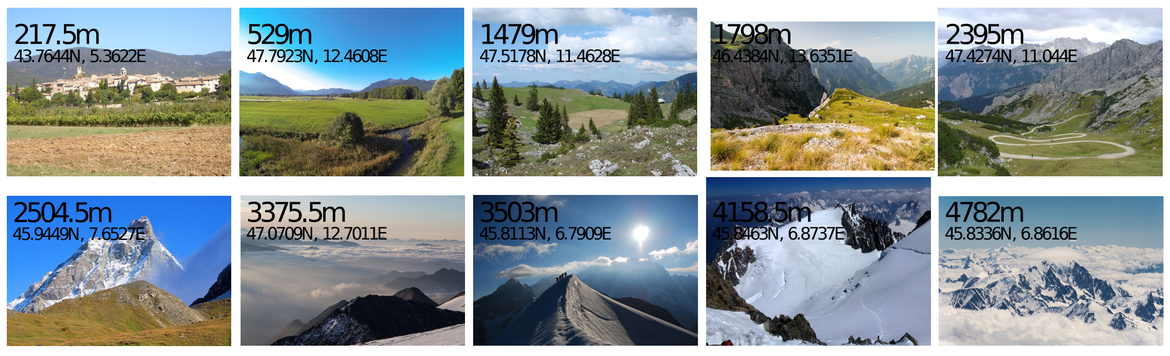
Alps100K: A NEW DATASET

Alps100K is a novel dataset of 98136 annotated (GPS coordinates, elevation, EXIF if available) outdor images from mountain environments. The collection covers vast area of Alps, the highest range in Europe.

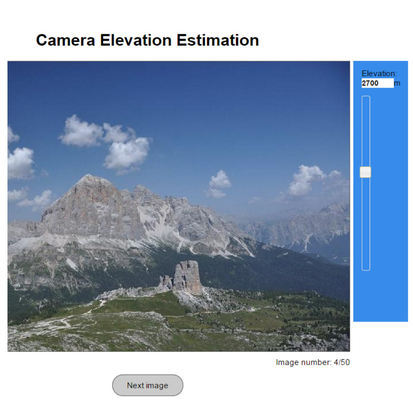


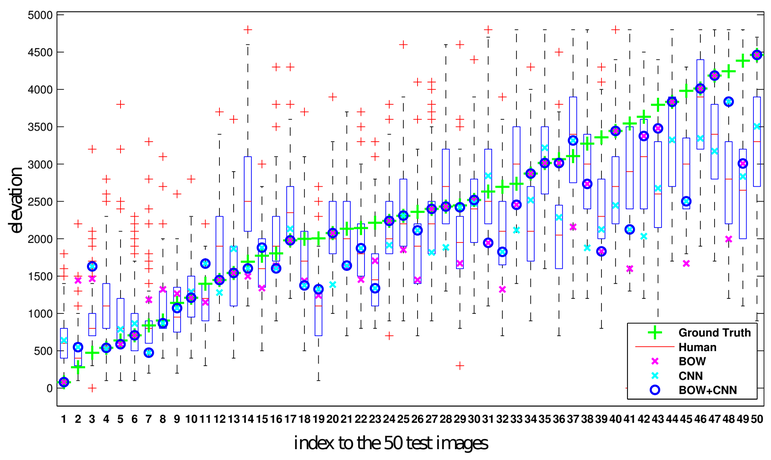
Dataset acquisition:

1. A list of all hills and mountain peaks was created using OpenStreetMap
2. The list of hill names was used to query the Flickr photo hosting service
3. Scene classifier was aplied to cull irrelevant (non-landscape) images
4. Elevation of photo was inferred using GPS and digital elevation model



HUMANS ESTIMATING ELEVATION





User experiment:

- 100 participants were tested using custom web-based interface
- 50 test images, spanning elevation range [79m, 4463m]
- presented in random order, on average 10 mins to finish the experiment

Results:

- humans are able to estimate camera elevation from visual information
- RMSE=879.95m, people underestimate high elevations

CONCLUSIONS

A new benchmark dataset of elevation-annotated images was collected. Two approaches were proposed to automatically estimate the camera elevation from a single landscape photo. In an user experiment, human performance on this task was measured and evaluation showed that the proposed methods outperform human abilities in camera elevation estimation.

Alps100K dataset is available for download:
http://cphoto.fit.vutbr.cz/elevation/

References

[1] F. Radenović, H. Jégou, and O. Chum. Multiple measurements and joint dimensionality reduction for large scale image search with shortvectors. In *Proc. ICMR. ACM*, 2015.
[2] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition using Places Database. *NIPS*, 2014.

M. Čadík, J. Vašíček, M. Hradiš, F. Radenović, and O. Chum. Camera Elevation Estimation from a Single Mountain Landscape Photograph. In *Proceedings of BMVC 2015*, Swansea, United Kingdom.

This work was supported by SoMoPro II grant (financial contribution from the EU 7 FP People Programme MarieCurie Actions, REA 291782, and from the South Moravian Region), SGS15/155/OHK3/2T/13, and MSMT LL1303 ERC-CZ grants. The content of this poster does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.