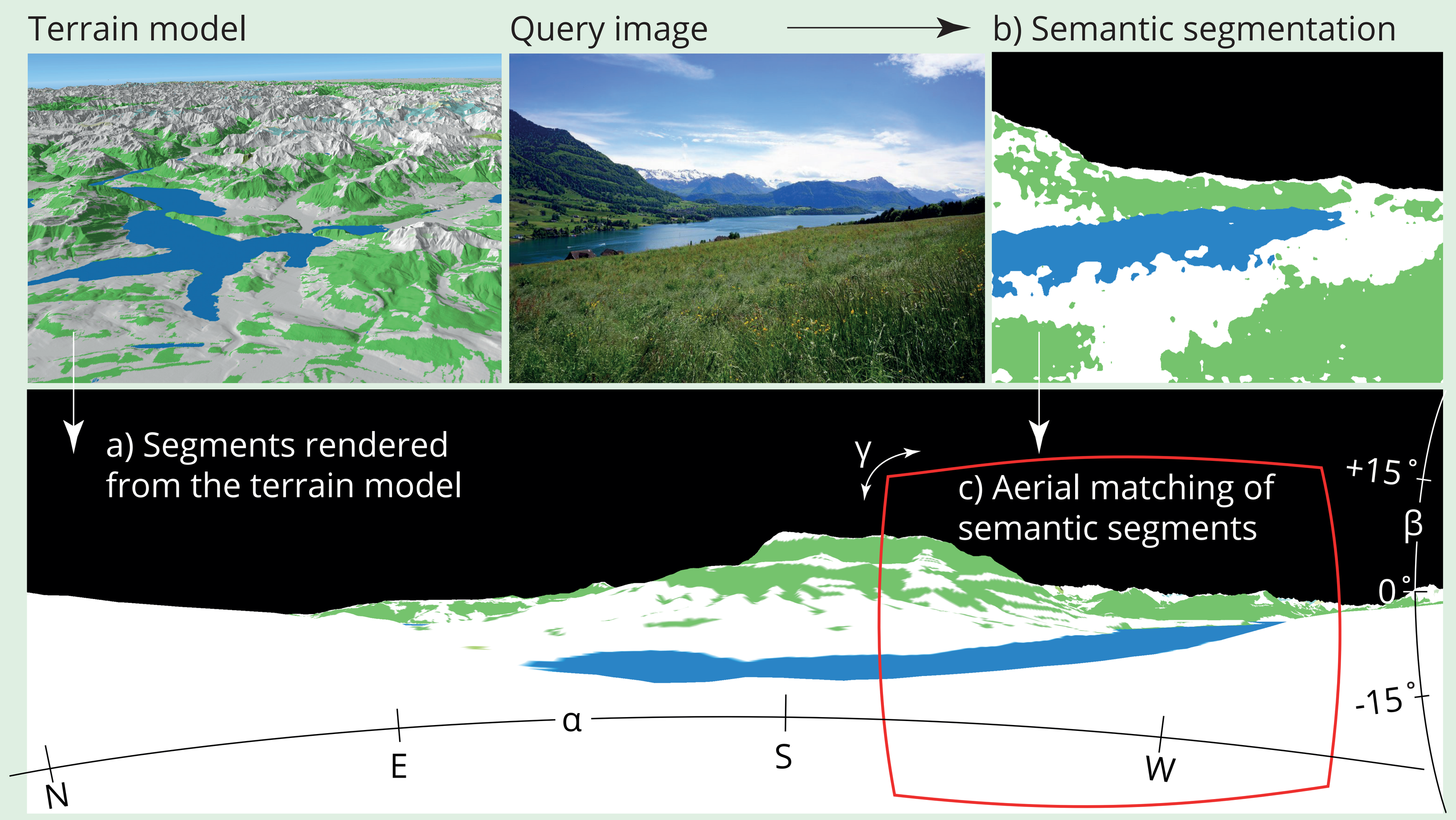
Camera Orientation Estimation in Natural Scenes Using Semantic Cues

Jan Brejcha ibrejcha@fit.vutbr.cz
Martin Čadík cadik@fit.vutbr.cz
CPhoto.fit.vutbr.cz
FACULTY OF INFORMATION TECHNOLOGY department of computer graphics and multimedia

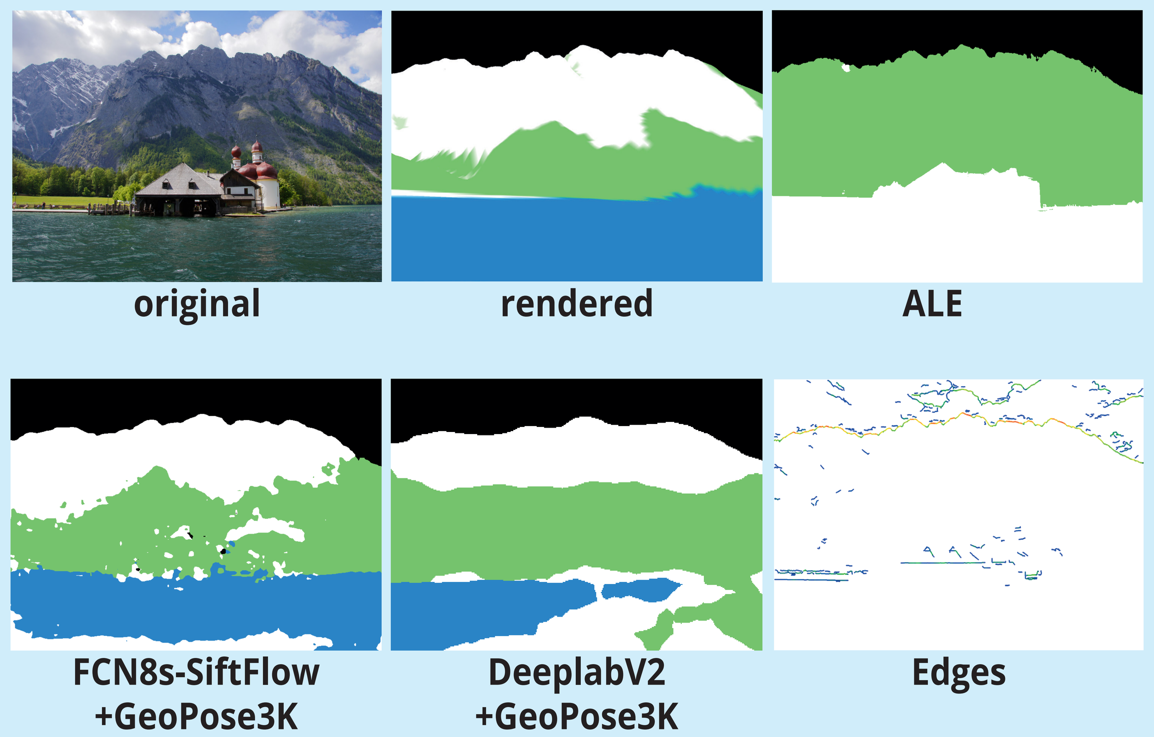
Overview

We propose a novel method for aligning a single image to a digital terrain model with the use of semantic segments. (a) Synthetic semantic segments and edges are rendered using terrain model and geospatial database. (b) Query image is segmented via semantic segmentation method. (c) Semantic segments from query image are aligned with synthetic semantic segments and camera orientation (α, β, γ) is recovered.



Semantic segmentation & edge detection

We fine-tuned and trained three semantic segmentation methods on a GeoPose3K [1] dataset. We show that training on synthetic data is important step to obtain reasonable segmentations for our application. Furthermore, to detect edges, we use edge detector trained on rendered silhouettes.



Spherical cross-correlation

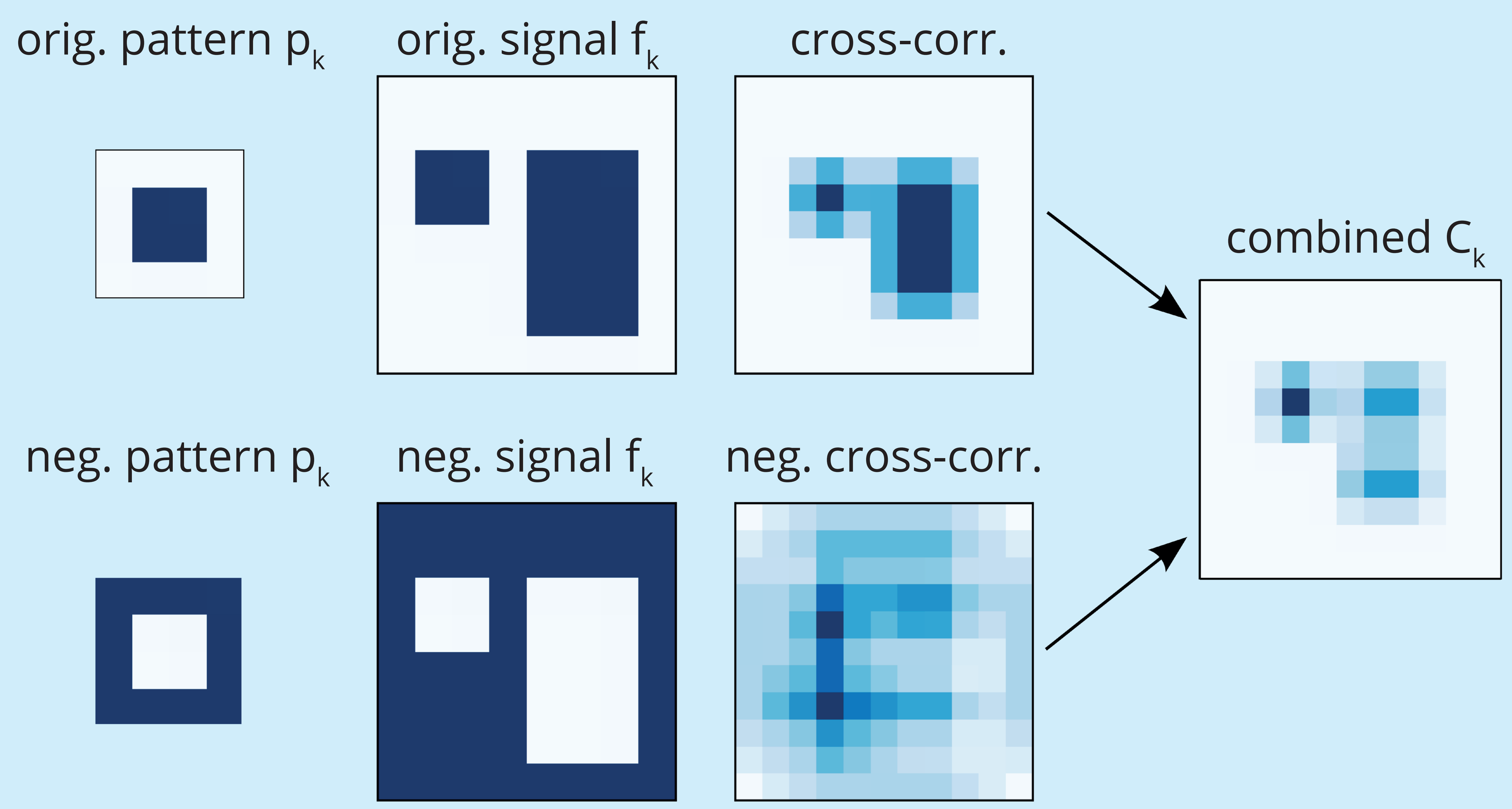
We propose to use spherical cross-correlation as the similarity measure to match semantic segments. We exploit the cross-correlation theorem for efficient computation in the Fourier domain.

$$\forall g \in SO(3): f \star p(g) = \int_{S^2} f(\omega)p(g^{-1}\omega)d\omega$$

Cross-correlation as a measure of confidence

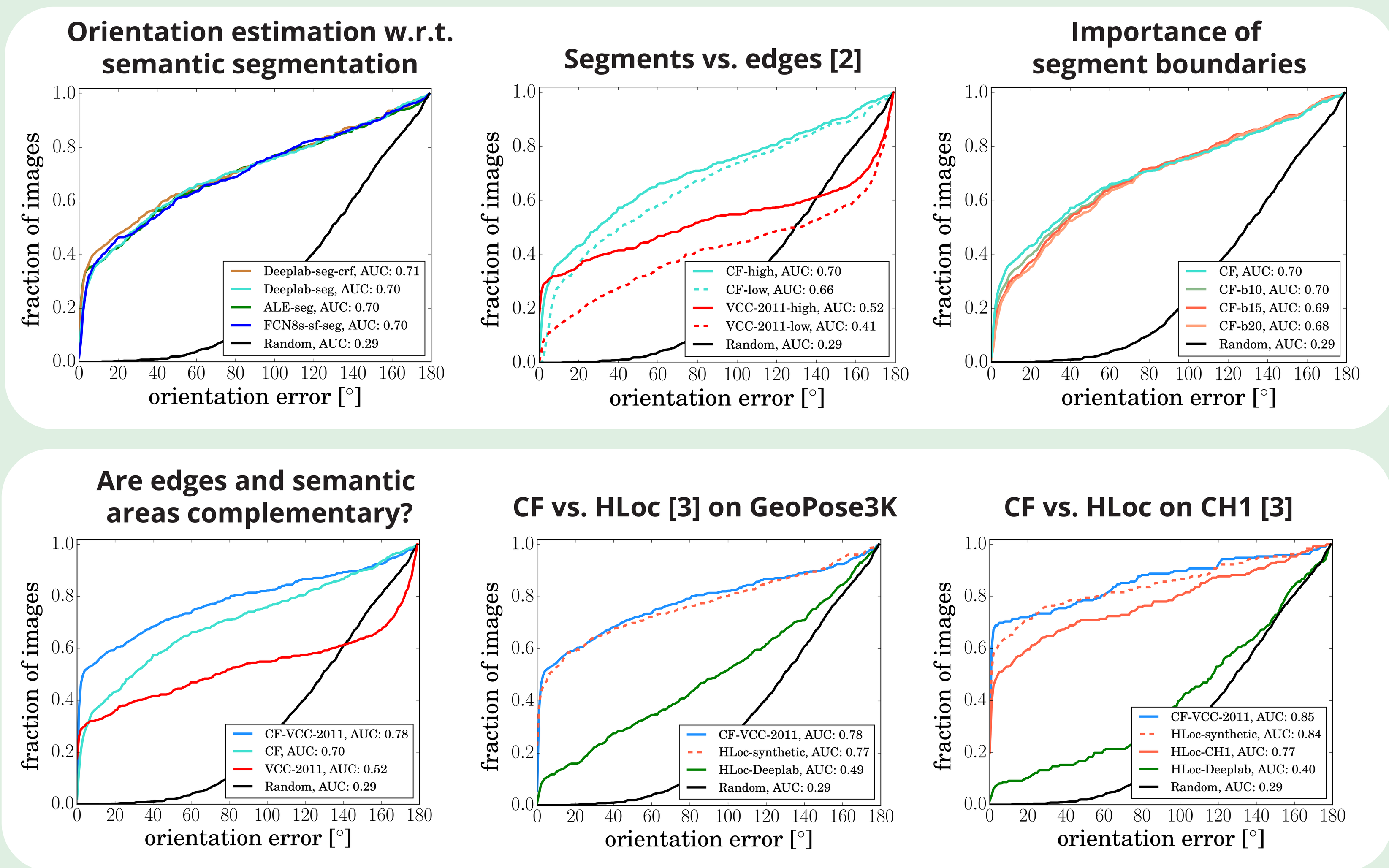
To estimate positive and accurate matching confidence, we propose to separate the cross-correlation into a positive and negative part. When the positive and negative cross-correlations are combined, the maximum value is correctly in place where both the pattern and its surroundings overlap the largest areas.

$$\forall g \in SO(3): C_k(g) = (f_k \star p_k(g))(f'_k \star p_k(g)')$$



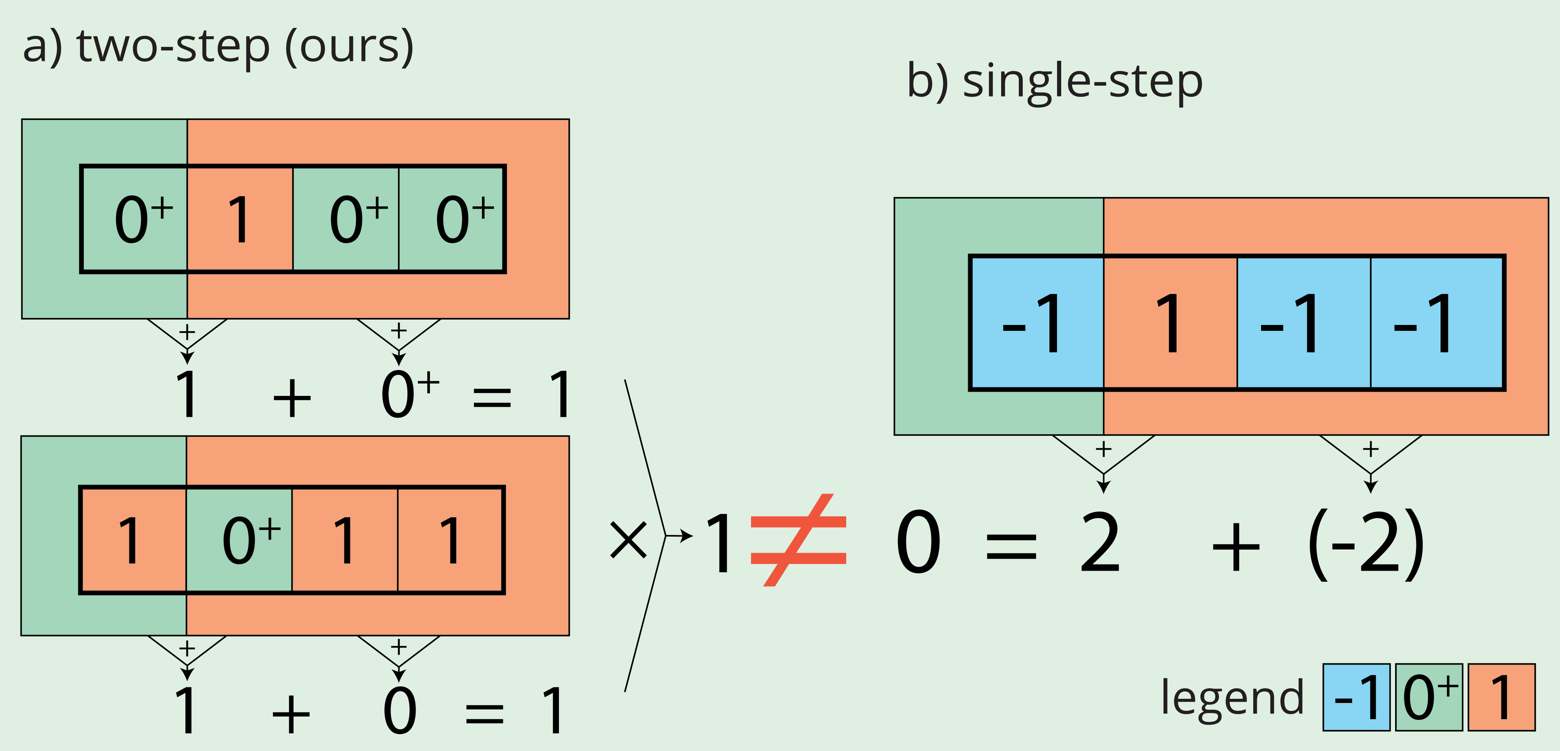
Experiments

- Semantic segments are complementary to edges.
- Boundaries of semantic segments are less informative than their areas.
- Combination of matching with edges and semantic segments exhibit better performance.



Two-step vs. single-step cross-correlation

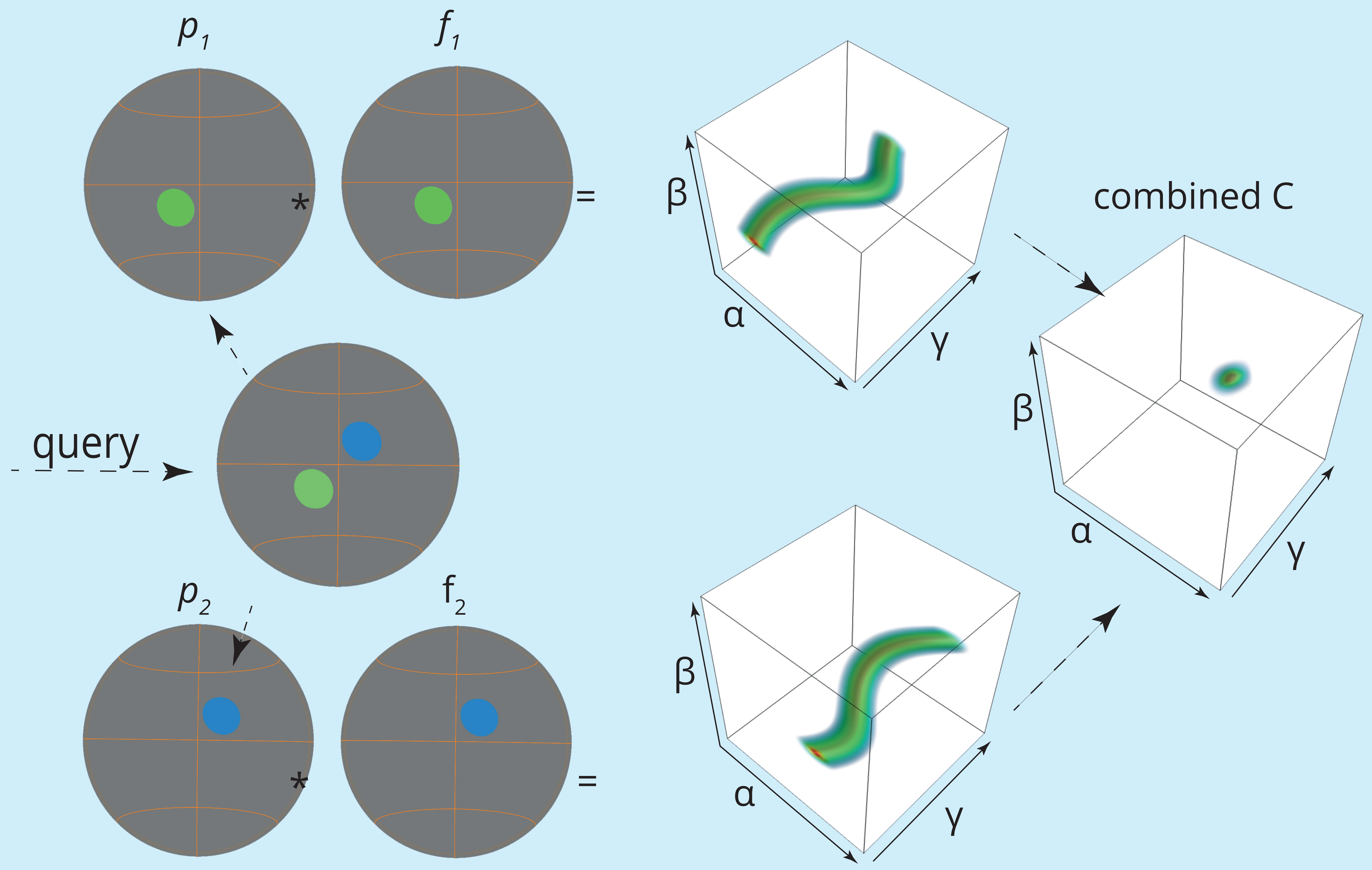
Our two-step cross-correlation of positive functions cannot be simply reduced to a single cross-correlation of real-valued functions, as it provides different results.



Confidence fusion

To match confidences of multiple semantic segment classes, we propose Confidence Fusion (CF) framework, which is a weighted geometrical average of subsequent confidences. This allows easy integration with different modalities, e.g., existing edge-matching [2] can be added as one confidence layer into our framework.

$$\forall g \in SO(3): C(g) = \prod_k (C_k(g))^{w_k}$$



References

[1] J. Brejcha and M. Čadík. GeoPose3K: Mountain Landscape Dataset for Camera Pose Estimation in Outdoor Environments. Image and Vision Computing, 66:1 – 14, 2017.
[2] L. Baboud,M. Čadík, E. Eisemann, and H.-P. Seidel. Automatic Photo-to-terrain Alignment for the Annotation of Mountain Pictures. In proceedings of CVPR 2011, pages 41–48, Washington, D.C., USA, 2011. IEEE Computer Society Press.
[3] O. Saurer, G. Baatz, K. Köser, L. Ladicky , and M. Pollefeys. Image based geo-localization in the Alps. International Journal of Computer Vision, 116(3):213–225, 2016.

Acknowledgements

This work was supported by V3C – "Visual Computing Competence Center" by Technology Agency of the Czech Republic, project no. TE01020415; by The Ministry of Education, Youth and Sports from the "National Programme of Sustainability (NPU II) project IT4Innovations excellence in science - LQ1602"; and by the IT4Innovations infrastructure which is supported from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center - LM2015070". We thank Jakub Pelikán for semantic segmentation.