Predicting Good Features for Image Geo-Localization Using Per-Bundle VLAD

Hyo Jin Kim, Enrique Dunn, Jan-Michael Frahm

UNC DEPARTMENT OF COMPUTER SCIENCE

CV²R Consortium for Vision and Virtual Reality

Motivation

Are all features useful for geo-localization?

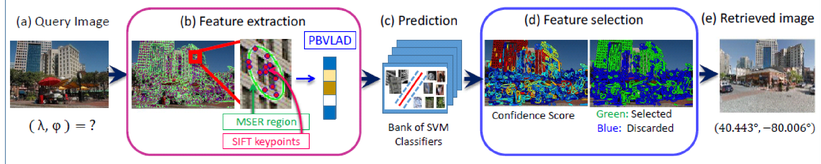
Selecting features intelligently would Allow us to achieve better performance while reducing the number of features.

Contributions

(1) We offer a way to predict features that are good in a data-driven sense. We show that by selecting features based on predictions from learned classifiers, geo-localization performance can be improved.

(2) We propose per-bundle vector of locally aggregated descriptors (PBVLAD) as a novel representation for bundled local features that is effective for both learning to predict features and image retrieval.

Overview



We discover features that are useful for recognizing a place in a data-driven manner, and use this knowledge to predict useful features in a query image prior to the geo-localization process.

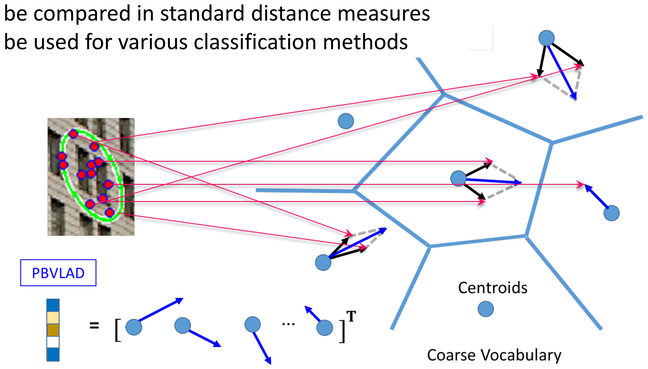
Feature Representation

Per-Bundle VLAD (PBVLAD)

MSER is described by a vector of locally aggregated descriptors (VLAD) on multiple scale-invariant features detected within the region.

Advantages:

- Robust to photometric and geometric changes
- Compact representation with fixed-size
- Can be compared in standard distance measures
- Can be used for various classification methods



$$p_a^w = \sum_{d_i : NN(d_i)=w, d_i \in B_a} \frac{d_i - c^w}{\|d_i - c^w\|} \qquad p_a = [p_a^1, p_a^2, ..., p_a^{|W|}]$$

Image Similarity

Matching score of feature to image: $f(p_q, I_r) = \max_{p_r \in I_r} M(p_q, p_r),$

Similarity between two images: $Sim(I_q, I_r) = \sum_{p_q \in I_q} f(p_q, I_r)$

Predicting Features for Geo-Localization

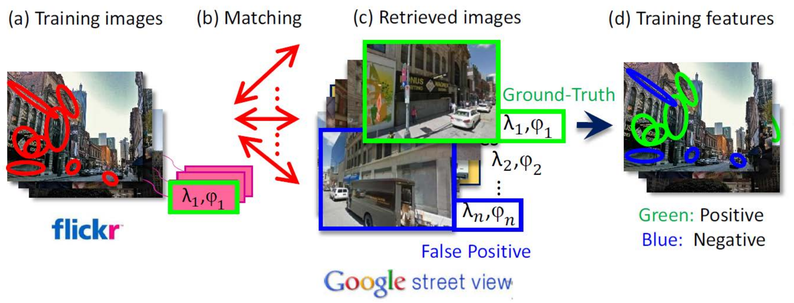
Automatic Training Data Generation

Utilize GPS-tagged images on the web to generate training data.

(1) Perform image geo-localization to get false positive images
- False positive images: Images in the shortlist whose GPS location is far away
- Ground-truth images: Spatially verified images near given GPS location

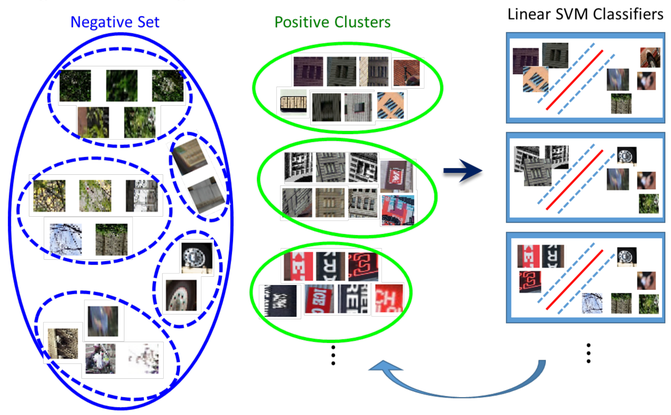
(2) Generate training data (features) comparing its matching score in ground-truth and false positives

- Positive: $f(p_t, I_{GT}) - f(p_t, I_{FP}) > m$
- Negative: $f(p_t, I_{GT}) - f(p_t, I_{FP}) < -m$



Closed-Loop Training of SVM Classifiers

Due to label conflicts and high intra-class variation, perform bottom-up clustering while training SVM on each cluster.

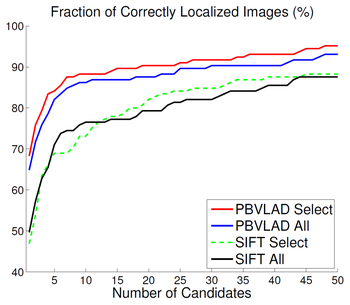


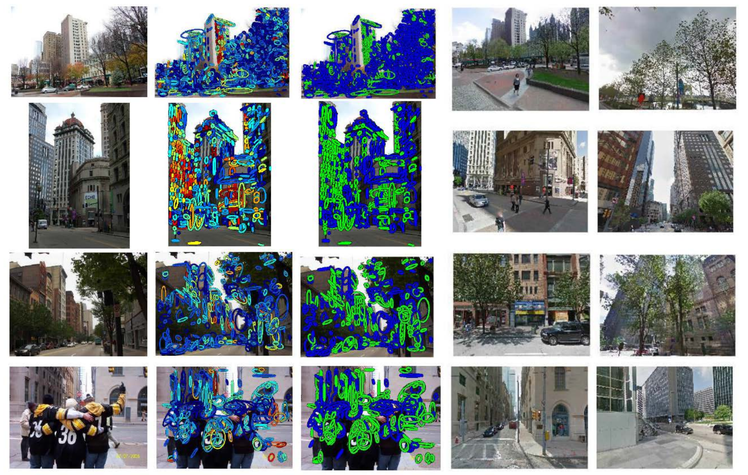
Experiments

Image Geolocalization in Pittsburgh, PA

Query Images: Unconstrained photos from photo sharing websites. [1]
Database Images: Google Streetview with 16 reprojcted images from each panorama

Method	% Correct
PBVLAD All	64.83
PBVLAD Select	**68.28**
PBVLAD Random	33.38
PBVLAD Select^C	19.31
SIFT All [1]	49.66
SIFT Select	46.90
Chance	0.20





Oxford Buildings Dataset

	Full	Dim Reduced		
Dim	16384	8192	4096	2048
mAP	**0.369**	0.364	0.334	0.264

Affect of dimension reduction

Descriptor	# Vocabulary	mAP
BoW	200,000	0.364
BoW	20,000	0.319
Fisher	64	0.317
VLAD	128	0.339
PBVLAD	128	**0.369**

Reference

[1] A. R. Zamir and M. Shah. Accurate image localization based on google maps street view. ECCV 2010

[2] Z. Wu, Q. Ke, M. Isard, and J. Sun. Bundling features for large scale partial-duplicate web image search. CVPR 2009

Acknowledgement

Supported by the Intelligence Advance Research Projects Activity (IARPA) via Air Force Research Laboratory (AFRL), contract FA8650-12-C-7214. The U.S Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government. The authors would also like to thank Alex Berg and Amir Zamir for helpful discussions.