# Visual Place Recognition with Graph Kernel

## Reference to following papers

Master thesis: Location models for visual place recognition, Elena Stumm

Piasco, Nathan, et al. "A survey on visual-based localization: On the benefit of heterogeneous data." *Pattern Recognition* 74 (2018): 90-109.

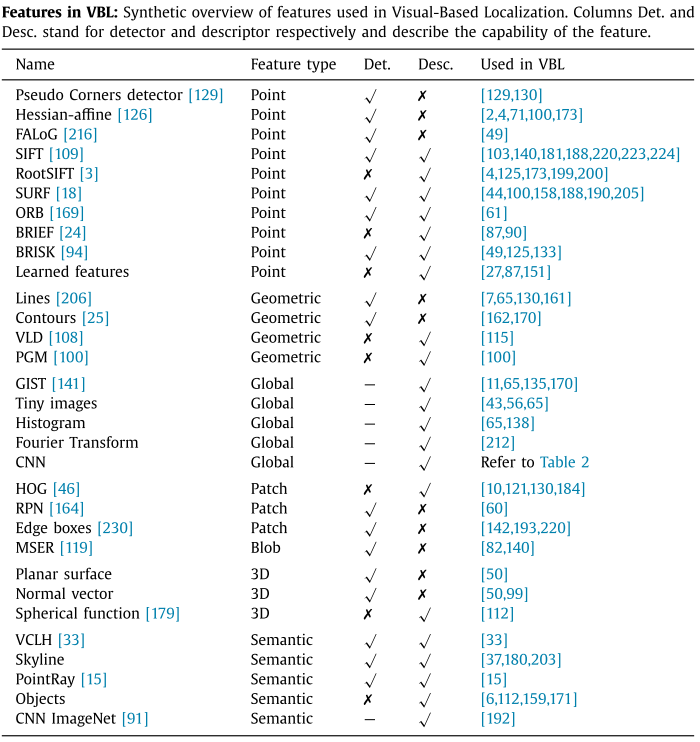
J., Hutter, M., & Siegwart, R. (2016). Robust visual place recognition with graph kernels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

Han, Fei, et al. "SRAL: Shared representative appearance learning for long-term visual place recognition." *IEEE Robotics and Automation Letters* 2.2 (2017): 1172-1179.

## VBL Application

Indirect method: visual place recognition, retrieve image from database and provides coarse location, sensitive to precision.

Direct methods: visual pose localization, retrieve 6-DoF information, recall matters, since it requires instantly pose return back.



## Feature aggregation techniques

* Quantization:

Same as text document search, local features are extracted as the equivalent visual words. Visual database builds a dictionary. Features are clustered to reduce the size of dictionary. The centroid of the clustered feature is visual words (like clustering for k centroids).

* Feature to visual words assignment:

hard assignment from the extracted feature to the nearest visual word in the dictionary. Soft assignment methods have been considered by associating the feature according to a linear combination of the k nearest neighborhoods.

* VLAD:

Instead NN or KNN, difference between feature and its closest visual word is assigned to the final descriptor instead of the visual word.

* Weighting scheme:

Used to find the discriminative features for similarity comparison

Tf-idf weighting: depends on the occurrence frequency of features in database/images.

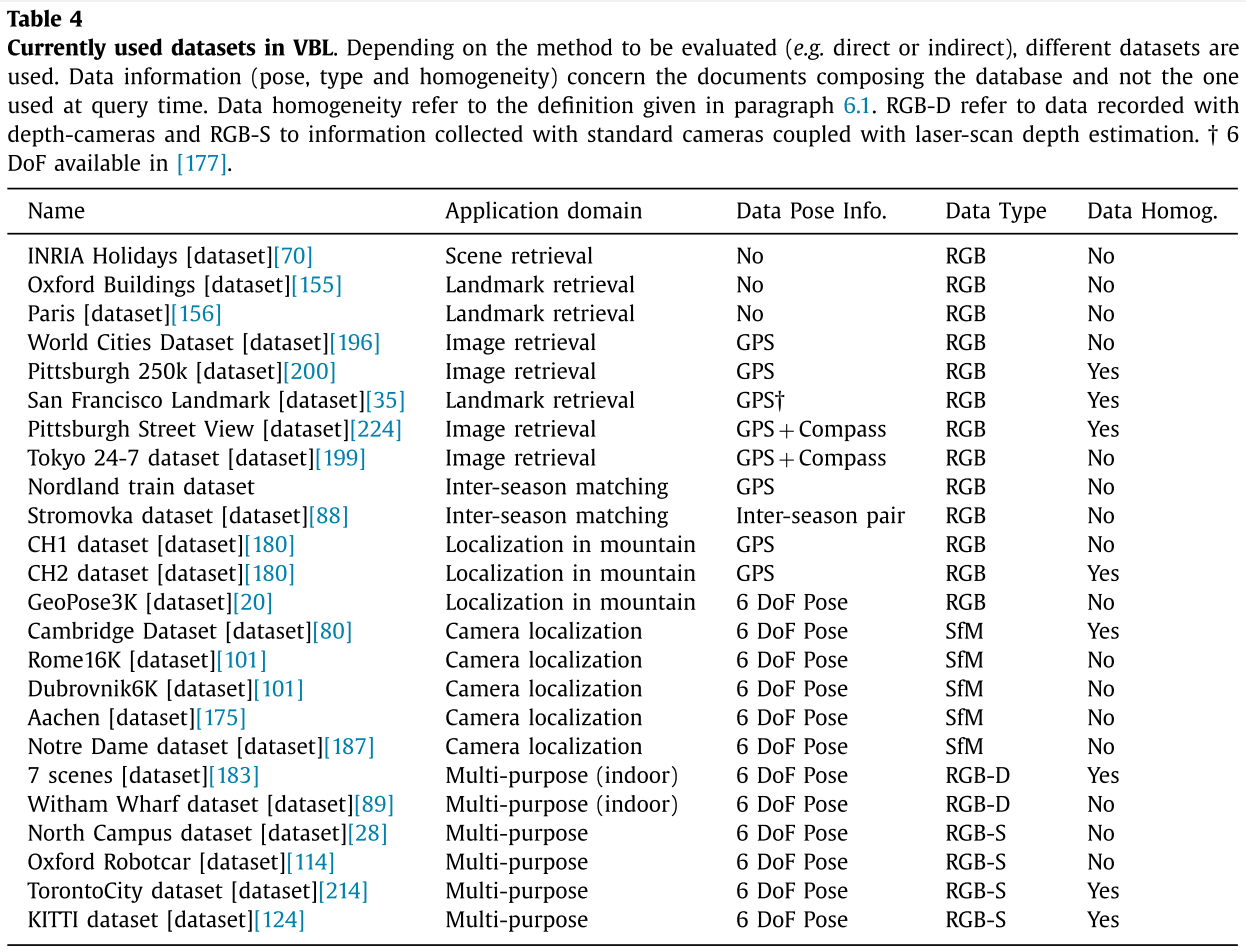
* CNN-based features:

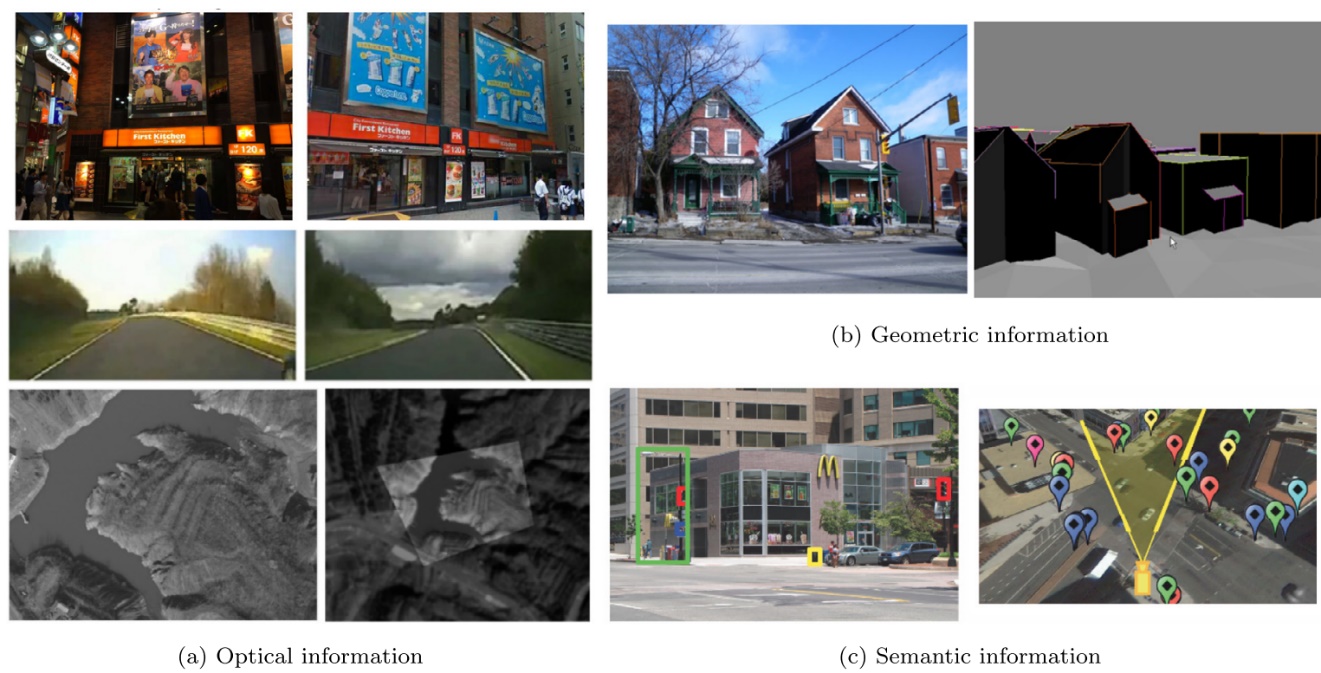
each CNN activation map serves as a feature producer. And several CNN maps produce high dimensional feature vectors for a more powerful discriminative illustration.

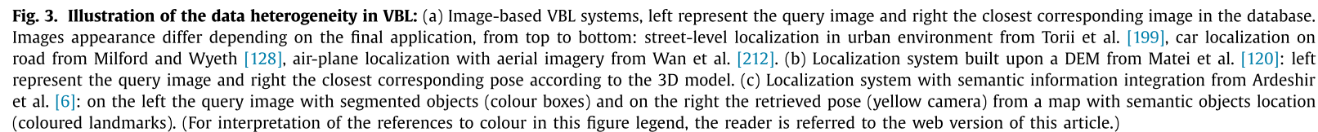
## Changing appearance/Long term SLAM

* Local features are more invariant to certain amount of changes in viewpoint.
* Image rectification for view angle and view points difference
* Illumination & long-term localization
  + Season changes: GRIEF or ORB are more suitable for this condition
  + Shadows: Fourier Transformation
  + Dynamic scene:
    - * Kim et al train SVM classifiers to discriminate strong and weak local features for the VBL task.
      * The method shows promising results where features are more often selected when they are attached to persistent objects, such as facades, and dismissed when they represent ephemeral or changing elements

## Data heterogeneity







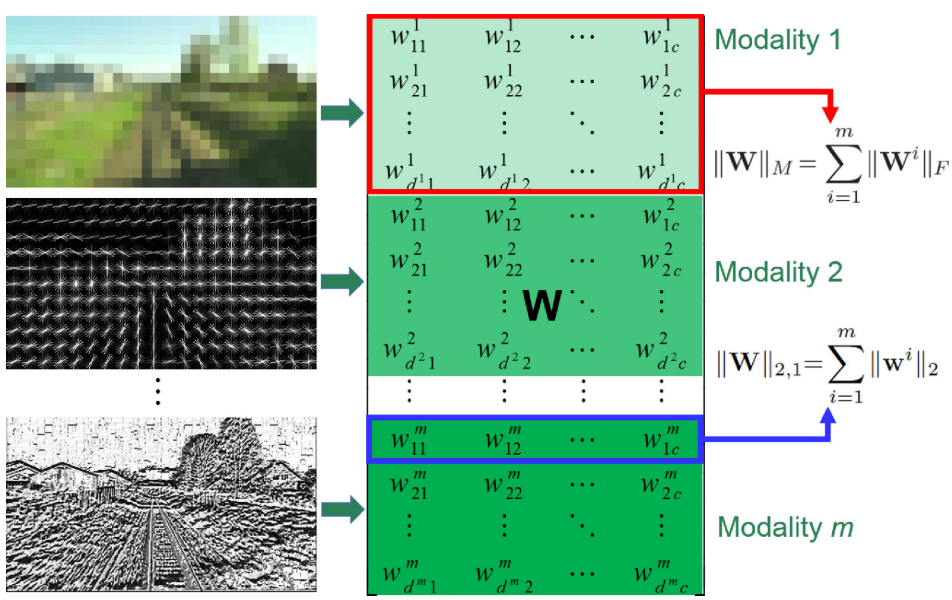
## MultiModal Features

* Global features extract information from the whole image, and a feature vector is often formed based on concatenation or feature statistics (e.g., histograms).
* Learning and fusing heterogenous multimodal features for long term recognition.

A regression model loss that find the relationship between multimodal features and scenarios:

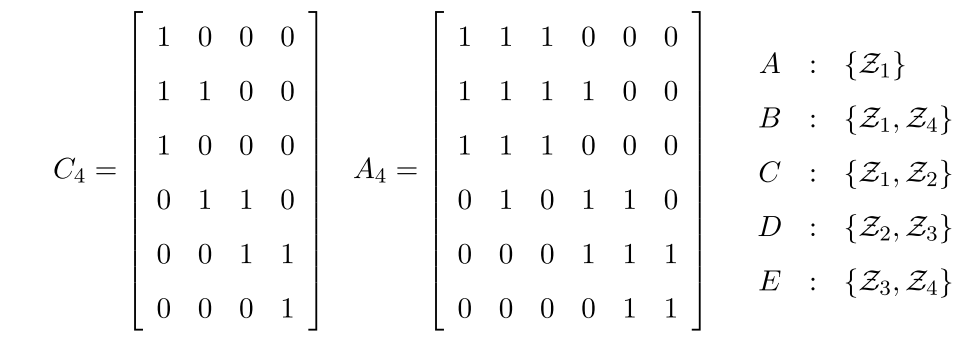


A multimodal features model, different features are concatenated together to serve as labels, then a regression model is trained to give high weights to those discriminative features.



## Graph Kernel

Covisibility map

* Map is updated as new image fed in. Map is a Sparse Clique Matrix
* Adjacency matrix can be easily computed from the SCM
* Inverted index between visual words and observation

Graph kernel:

effectively transform the graph G and G' to a linear feature space.

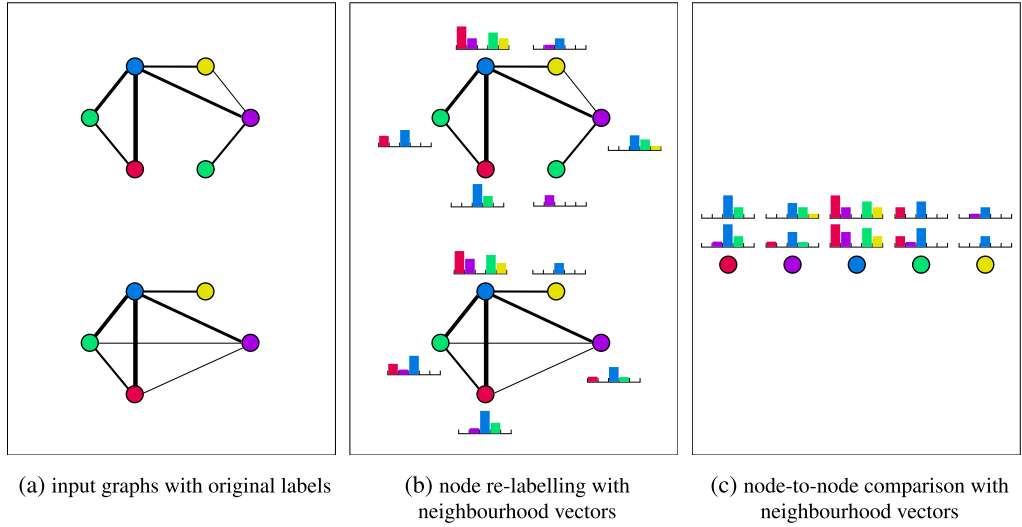
## WL Kernel

## 

## 9. Robust Place Recognition with Graph Kernels

* Rather than relabelling nodes with a single new value, node labels are augmented by a vector corresponding to their neighbourhood. The length of the vector is equal to the size of the label vocabulary (in this case the visual dictionary), and each element is weighted by the strength of the connecting edges in the covisibility graph.
* graph similarity can be measured by taking the dot product between the neighbourhood vectors of corresponding nodes in each graph and summing the results.

Illustrated in the following graph:



The resulting complexity of the observation likelihood calculation is on the order O(nd) (bounded by O(n2)), where n is the number of common nodes, and d is the degree of the graph. Furthermore, due to the sparse nature of visual word observations, a sparse implementation ensures that the complexity does not scale with the vocabulary size (typically on the order of tens or hundreds of thousand words).