

# Geo-referenced UAV Localization

Mo Shan

Paopao Robot Talk

March, 2018

# Outline

## Geo-referenced localization

Feature based image matching

Gradient based image matching

# Geo-referenced localization

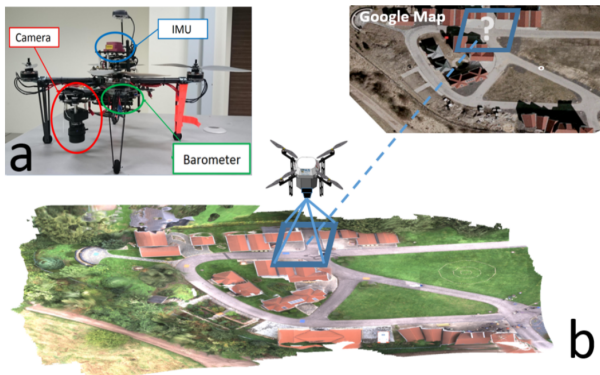
## Motivation

- ▶ UAV flies in outdoor environment, over houses, roads, etc
- ▶ GPS alone may be insufficient, eg jamming, disaster management
- ▶ The operating zone is usually known
- ▶ An easily accessible, memory efficient prior map could be used as reference, eg Google Map

# Geo-referenced localization

## Problem overview

- ▶ UAV relies on camera, IMU, barometer, prior map



# Geo-referenced localization

## Problem definition

- ▶ Given a prior map  $\mathcal{M}$ , a sequence of images  $\mathcal{X} = \{x_1, \dots, x_{t-1}\}$ , IMU data  $\mathcal{Y} = \{y_1, \dots, y_{t-1}\}$ , where  $y_i$  contains angular velocities and roll, pitch, yaw angles, and altitude  $\mathcal{D} = \{d_1, \dots, d_{t-1}\}$ , where  $d_i \in \mathbb{R}_{>0}$ ,  $t \in \{1, \dots, T\}$
- ▶ Calculate the maximum likelihood location
$$l_t = \underset{l}{\operatorname{argmax}} P(l | \mathcal{M}, \mathcal{X}, \mathcal{Y}, \mathcal{D})$$
- ▶ Simplified as: given the previous state, the current state is independent of the history
- ▶ 
$$l_t = \underset{l}{\operatorname{argmax}} P(l | \mathcal{M}, x_{t-1}, x_t, y_{t-1}, y_t, d_{t-1}, d_t, l_{t-1})$$

# Geo-referenced localization

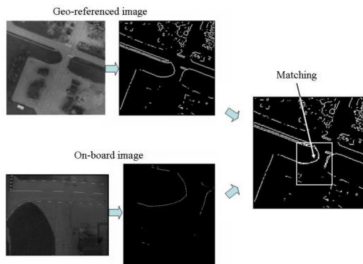
## Challenges

- ▶ Significant scene changes due to difference in modality, viewpoint, weather, etc
- ▶ Lack of visible features in certain regions of low resolution map
- ▶ Large illumination variation for on-the-fly images

# Geo-referenced localization

## Literature review

- ▶ Image registration technique realized by edge matching
- ▶ The registration is robust to change in scale, rotation and illumination to a certain extend
- ▶ However, during the whole flight there are few successful matches

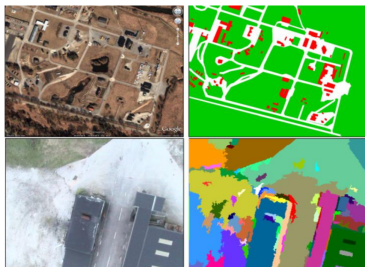


Source: Conte and Doherty

# Geo-referenced localization

## Literature review

- ▶ UAV images are segmented into superpixels and then classified as grass, asphalt and house
- ▶ Circular regions are selected to construct the class histograms, which are rotation invariant
- ▶ However, discarding rotation gives rise to the classification uncertainty



Source: Lindsten et al.



# Geo-referenced localization

## Initial position

- ▶ Correlation filter is used for global localization
- ▶  $F$  is 2D Fourier transform of the input image
- ▶  $H$  is the transform of the filter
- ▶  $\odot$  denotes element wise multiplication and  $*$  indicates complex conjugate.
- ▶ We correlate the current frame and the map.
- ▶ Transforming  $G$  into the spatial domain gives a confidence map of the location.

$$G = F \odot H^* \quad (1)$$

# Geo-referenced localization

## Initial position

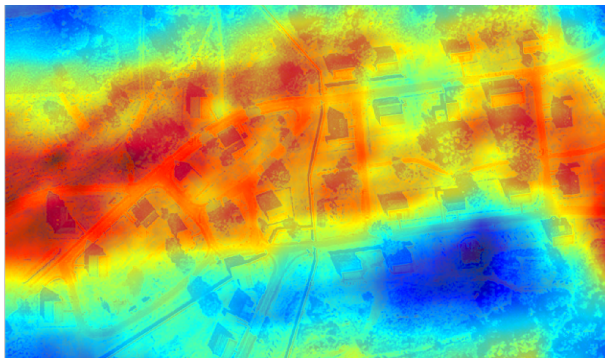
- ▶ Onboard image at take off position, and its corresponding rectangular region in the map



# Geo-referenced localization

## Initial position

- ▶ The confidence map of the frame
- ▶ The black area represents the highest confidence
- ▶ However, this may fail if the image contains little distinctive feature



# Geo-referenced localization

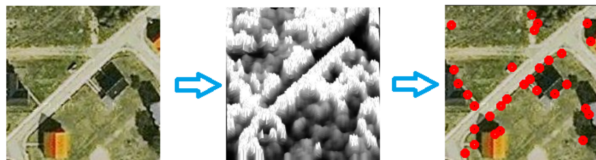
## Position prediction

- ▶ The current position is predicted to confine template matching
- ▶ The features are selected and tracked based on optical flow
- ▶ Compute the motion field using angular velocities and depth as in PIX4FLOW
- ▶ Inter-frame motion can also be obtained from homography decomposition

# Geo-referenced localization

## Feature based approach

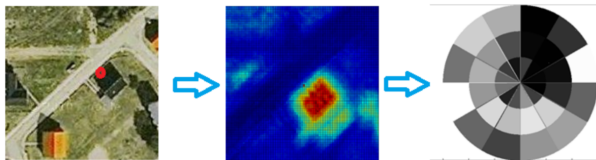
- ▶ Maximal Self Dissimilarity (MSD) measures the self-dissimilarity of a pixel according to the rarity of the central patch
- ▶ The similarity metric is Sum of Squared Distance (SSD)
- ▶ The image is transformed into a saliency map based on the rarity of the patch, and then keypoints are detected at maximum in the map



# Geo-referenced localization

## Feature based approach

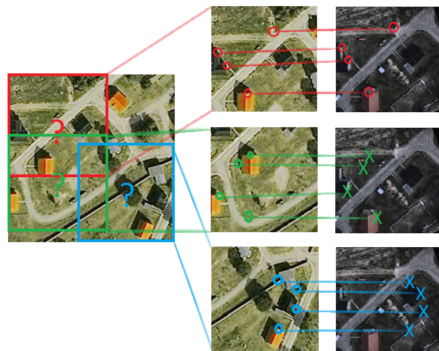
- ▶ Local Self Similarity (LSS) descriptor is formed by comparing the image patch with its surrounding regions using SSD
- ▶ The correlation surface is transformed to the descriptor by log-polar binning



# Geo-referenced localization

## Feature based approach

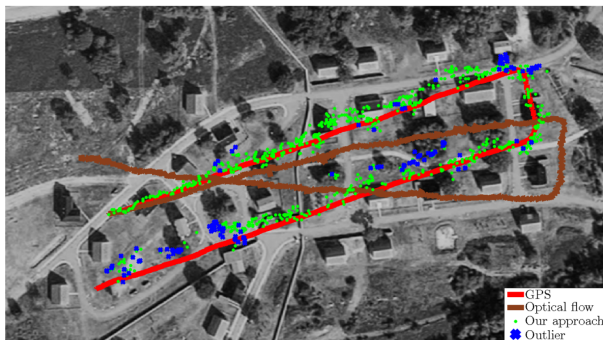
- ▶ Only the keypoints in the reference map will be used due to inconsistency for different modalities
- ▶ For correct window, all keypoints will overlap those in the template, achieving minimum L2 distance over the feature descriptors



# Geo-referenced localization

## Feature based approach

- ▶ Feature based approach follows GPS closely
- ▶ But SSD computations in MSD and LSS are time consuming





# Geo-referenced localization

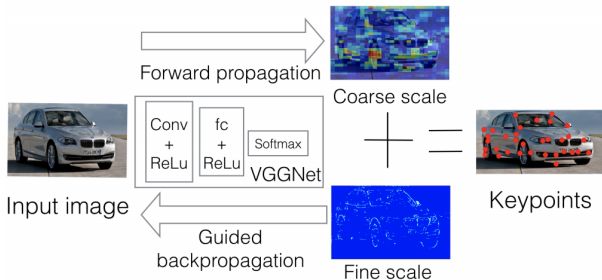
## Feature based approach

- ▶ Hand-crafted keypoint detection may lack semantic consistency
- ▶ However, training CNNs often require large annotated dataset
- ▶ Is it really necessary to label each keypoint for CNNs?
- ▶ Class labels could provide weak supervision

# Geo-referenced localization

## Feature based approach

- ▶ The input image is fed to a pretrained network on classification
- ▶ Use an occluder to obtain the coarse scale heatmap
- ▶ Guided backpropagation is performed to get the fine scale heatmap



# Geo-referenced localization

## Feature based approach

- ▶ At coarse scale, the contribution of each patch in the input image for object classification is analyzed by covering it and examine the change in the confidence of class prediction
- ▶ If the confidence of the correct class drops dramatically due to the occlusion of a patch, then the probability of the patch containing a discriminative feature is very high

# Geo-referenced localization

## Feature based approach

- ▶ The network is denoted by a mapping  $f : \mathbb{R}^N \mapsto \mathbb{R}^C$ ,  $x \in \mathbb{R}^N$ ,  $y \in \mathbb{R}^C$ , where  $x$  is an image of  $N$  pixels, and  $y = [y_1, \dots, y_C]^T$  denotes the classification score of  $C$  classes, with  $y_i$  being the probability of the  $i$ th class. The pixels inside an occluder  $b$  of image  $x$  are replaced by a vector  $g$ , and this occlusion function is denoted by  $h_g$ . Hence the change in classification score is  $\delta_f(x, b) = \max(f(x) - f(h_g(x, b)), 0)$ .
- ▶ To avoid creating edges, random colors are used as  $g$  instead of mono color
- ▶ Since only the class with maximum probability is considered, the decrease of score is  $d(x, b) = \delta_f(x, b)^T \mathbb{I}^C$ , where  $\mathbb{I}^C \in \mathbb{N}^C$  is an indicator vector whose elements are zero except at the predicted class  $c$ .

# Geo-referenced localization

## Feature based approach

- ▶ For the fine scale, guided backpropagation is performed on the unit that has maximum activation from the softmax layer
- ▶ It reveals which pixel positively influences the class prediction, by maximizing the probability of the predicted class while minimizing that of other classes, ie it locates the pixel where the least modification has to be made in order to affect the prediction the most
- ▶ It's called guided backpropagation because the gradient is guided by the input from below and by the error from above

# Geo-referenced localization

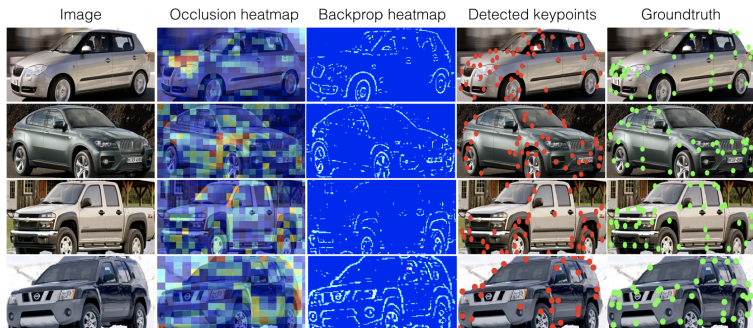
## Feature based approach

- ▶ The activation at layer  $l + 1$  could be obtained from the activation at layer  $l$  through a ReLU unit as  $f_i^{l+1} = \text{ReLU}(f_i^l) = \max(f_i^l, 0)$ .
- ▶ The backpropagation is  $R_i^l = (f_i^l > 0) \cdot R_i^{l+1}$ , where  $R_i^{l+1} = \frac{\partial f^{out}}{\partial f_i^{l+1}}$ .
- ▶ For guided backpropagation, not only the input is positive, but also the gradient, i.e.  $R_i^l = (f_i^l > 0) \cdot (R_i^{l+1} > 0) \cdot R_i^{l+1}$ . In this way only the positive gradients are retained in backpropagation

# Geo-referenced localization

## Feature based approach

- ▶ The coarse scale and fine scale are combined linearly
- ▶ The heatmaps are transformed into log-likelihood keypoint distributions used as the confidence score



# Geo-referenced localization

## Feature based approach

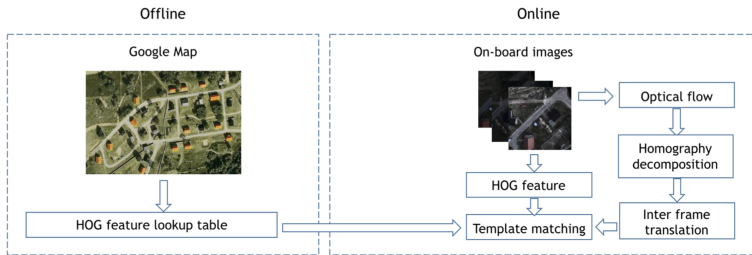
- ▶ The most important patches are usually those centered around the keypoints, such as those near the rear view mirrors, head lights as well as the wheels, which are semantically consistent
- ▶ The rear view mirrors as well as car logos are always highlighted in the gradient images from guided backpropagation, which confirms the close relevance of keypoints and high activations
- ▶ This approach could detect semantically consistent keypoints in the reference map and the onboard image, eg corners of the man-made structures, and sliding window search could be avoided. However, it's still difficult to obtain real-time performance due to forward and backward passes



# Geo-referenced localization

## Gradient based approach

- ▶ Histograms of Oriented Gradients (HOG) descriptors are used to encode the gradient information in multi-modal images
- ▶ The HOG features for the map are computed offline
- ▶ During onboard processing, we use global search to initialize the UAV position
- ▶ Then for each frame, we track the pose by position prediction and image registration



# Geo-referenced localization

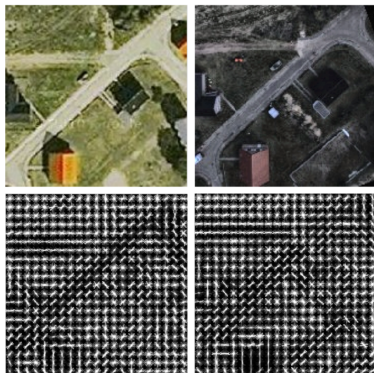
## Gradient based approach

- ▶ To construct HOG, 1D point derivative masks are convolved with the image to get the gradients
- ▶ Magnitude-weighted gradient orientation histograms are constructed in cells and blocks
- ▶ Clipped  $L_2$  norm normalization scheme is performed to the histogram of every block to compensate for illumination variance
- ▶ Because the blocks are overlapped, every cell contributes to multiple blocks, significantly improving the performance of HOG
- ▶ Eventually the histograms are vectorized to form a 1D feature

# Geo-referenced localization

## Gradient based approach

- ▶ The gradient patterns for houses and roads are quite similar in HOG glyph
- ▶ The structures of road and house are clearly preserved even under dramatic photometric variations



# Geo-referenced localization

## Gradient based approach

- ▶ Several metrics are compared to compute the similarity of HOG descriptors
- ▶ Correlation and Intersection measures similarity while Chi-Square and Bhattacharyya measures distance
- ▶ We transform similarity values to distance by  $d = 1 - \text{correlation}$
- ▶ The distance values are then normalized with respect to the ground truth value
- ▶ Correlation is the best for differentiating the outliers, since the distances of 1.829, 2.428 are the largest

|               | GT | Outlier 1 | Outlier 2 |
|---------------|----|-----------|-----------|
| Correlation   | 1  | 1.829     | 2.428     |
| Chi-Square    | 1  | 1.810     | 2.009     |
| Intersection  | 1  | 0.954     | 0.970     |
| Bhattacharyya | 1  | 1.260     | 1.248     |

# Geo-referenced localization

## Gradient based approach

- ▶ Weighted coarse to fine search is used to avoid sliding window search
- ▶ There are  $N$  particles, and for each particle  $p$ , its properties include  $\{x, y, H_x, H_y, w\}$ , where  $(x, y)$  specify the top left pixel of the particle,  $(H_x, H_y)$  is the size of the subimage covered by the particle and  $w$  is the weight. The  $(x, y)$  is generated around the predicted position, while  $(H_x, H_y)$  equals to the size of the onboard image
- ▶ The optimal estimation of the posterior is the mean state of the particles. Suppose each  $p$  predicts a location  $l$ , then the estimated state is

$$E(l) = \sum_{i=1}^N w_i l_i \quad (2)$$

# Geo-referenced localization

## Gradient based approach

- ▶ Based on the predicted state  $(x_p, y_p)$  of where the UAV could be in the next frame, we calculate the likelihood that UAV location  $(x_c, y_c)$  is actually at this location.
- ▶ After the particles are drawn, the subimages of the map located at the particles are compared with the current frame. To estimate the likelihood, we use Gaussian distribution to normalize these distance values, where  $d$  is the distance between the two images under comparison,  $\sigma$  is the standard deviation,  $\hat{w}$  is then normalized based on the sum of all weights to ensure that  $w$  is in the range  $[0, 1]$ .

$$\hat{w} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-d^2}{2\sigma^2}\right) \quad (3)$$

# Geo-referenced localization

## Gradient based approach

- ▶ The search is conducted from coarse level to fine level to reduce the computational burden
- ▶ For the coarse search,  $N$  particles are drawn randomly in a rectangular area, whose width and height are both  $s_c$ , with a large search interval  $\Delta_c$ .
- ▶ The fine search is carried out in an smaller area with size  $s_f$  and search interval  $\Delta_f$ .
- ▶ HOP relies mainly on coarse search which is often quite accurate. If the minimum distance of coarse search is larger than a threshold  $\tau_d$ , then the match is considered invalid. Only when coarse search fails to produce valid match do we conduct fine search

# Geo-referenced localization

## Gradient based approach

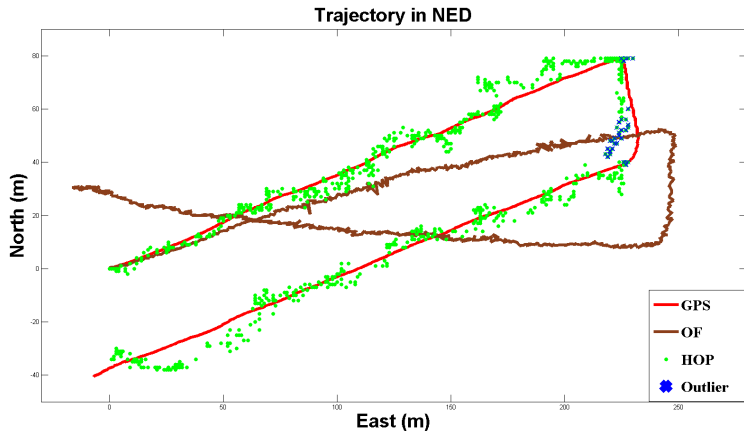
- ▶ The most important parameters are  $N$  and  $s_c$ .
- ▶ More  $N$  increases the accuracy of the weighted center but demands more computational resources.
- ▶ Likewise, larger  $s_c$  ensures the matching is robust to jitter while smaller  $s_c$  reduces the time consumed.
- ▶ Hence, we trade off the robustness and efficiency when determining those parametric values.



# Geo-referenced localization

## Gradient based approach

- ▶ The root mean square error (RMSE) of HOP is 6.773 m
- ▶ It runs at 15.625 Hz on average





# Geo-referenced localization

## Key insights

- ▶ Low resolution Google Map could be used to provide prior information for localization
- ▶ CNNs trained with weak supervision may provide consistent keypoints
- ▶ HOG is an effective descriptor for multi-modal image registration

## Related papers

- ▶ Geo-referenced UAV localization is presented in [1]
- ▶ Keypoint detection using CNNs is presented in [2]
-  Google Map Aided Visual Navigation for UAVs in GPS-denied Environment
-  Weakly supervised keypoint detection