Features Attention and Anticipation with Stereo Vision (#16)

Carnegie Mellon University

Kai-Hung Chang (kaihungc), Jerry Hsiung(shsiung)

The Robotics Institute

Motivation

Visual odometry (VO) algorithms have been major research interests as cameras become cheaper and more ubiquitous. Among them the most common ones are feature-based methods shown in Figure 1.

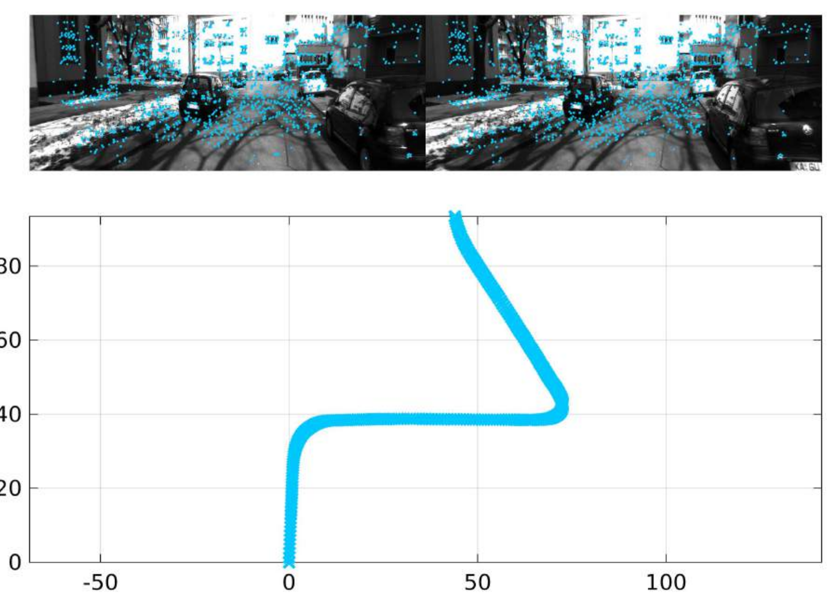


Figure 1. A typical stereo visual odometry includes feature matching and calculating frame-to-frame transformation [1].

However, traditional feature-based pipelines might require extensive computational costs and suffer from real-time operation without the use of powerful CPUs or GPUs. The motivation of this project is to find a novel formulation so that the VO uses only **minimum subset of total features** to best **minimize future-covariances** while **maintaining constant computational costs**. The results allows feature-based visual odometry methods to be more efficient and suitable for micro robotics platforms.

Problem Formulation

Our pipeline first predicts the next several states based on the past linear motion. It also predicts future landmark projections. From the predictions, we select the best subset of features.

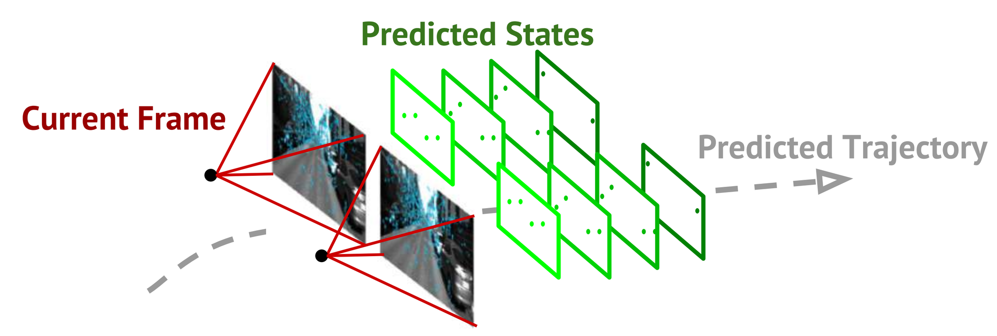


Figure 2. Our formulation predicts the future states based on the past odometry, and calculate the associated future state covariance to be minimized

We select the features based on the information matrix of state estimation and landmark information. We formulate this problem as an optimization problem that solves two kinds of objectives.

(1) Worst Estimation Error:

$$f_\lambda(\mathrm{S}) \;=\; \lambda_{\min}\left(\bar{\Omega}_{k:k+H} + \textstyle\sum_{l\in\mathrm{S}} p_l \Delta_l\right)$$

(2) Volume and Mean Radius of the Confidence Ellipsoid:

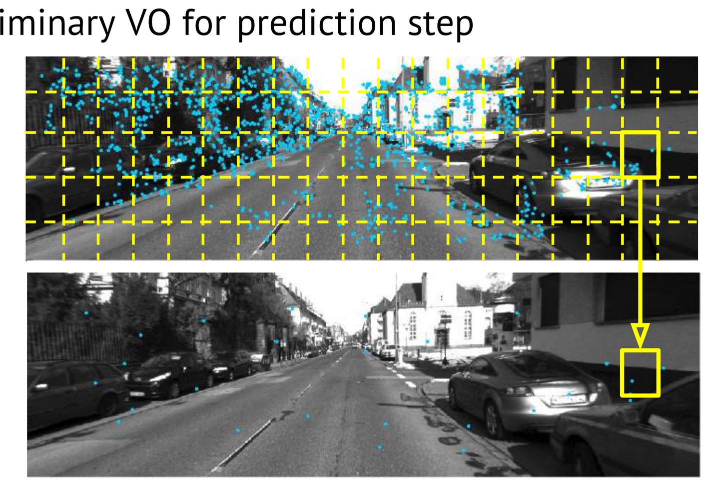
$$f_{\det}(\mathrm{S}) \;=\; \log\det\left(\bar{\Omega}_{k:k+H} + \textstyle\sum_{l\in\mathrm{S}} p_l \Delta_l\right)$$

where S is the subset of features, k:k+H is the future prediction horizon, and $\Omega_{k:k+H} \doteq P^{-1}_{k:k+H}$ is the information matrix associated with future pose states.

Approach

Our proposed algorithm follows the following steps:

(1) Select Bucket Features:
Select maximum n features per bucket to calculate preliminary VO for prediction step

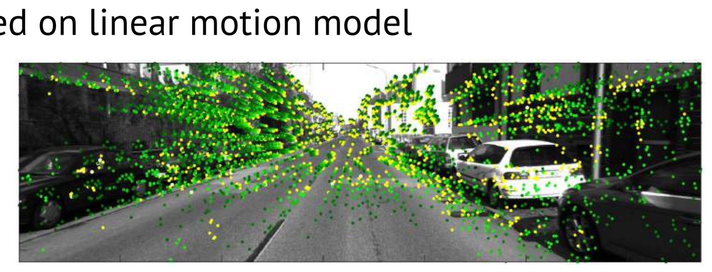


(2) Predict Bucket Features:
Calculate predicted position and use future poses and landmarks to form a Bundle Adjustment equation. Use the Schur-Complement to extract state information.

$$\text{Complete State Information} = \begin{bmatrix} F^TF & F^TE \\ E^TF & E^TE \end{bmatrix}$$

$$\bar{\Omega}_{k:k+H} = F^TF - F^TE(E^TE)^{-1}E^TF$$

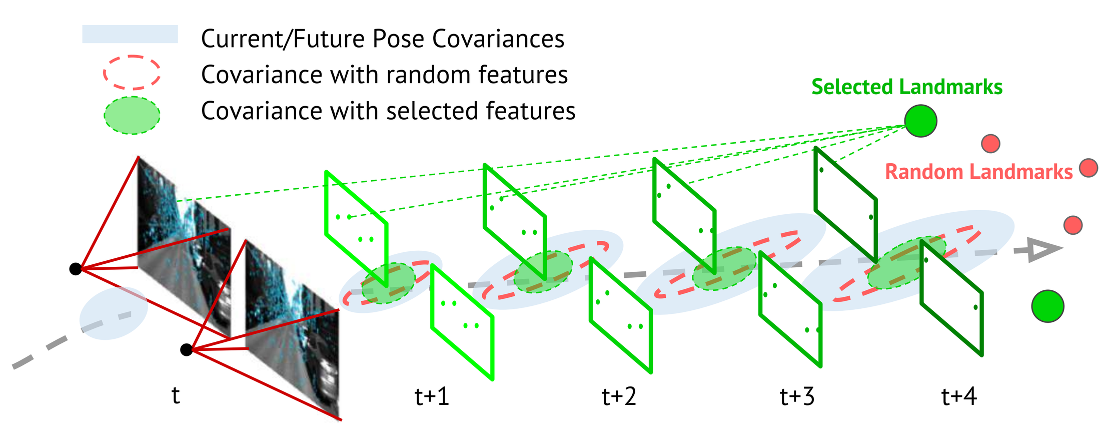
(3) Predict Rest of Features (Pool features):
For the rest of the points, predict their feature locations based on linear motion model



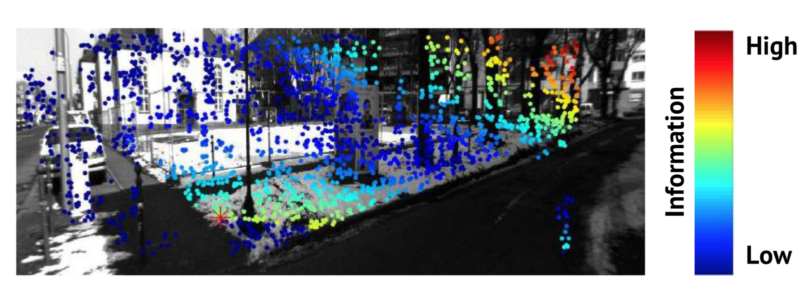
(4) Feature-wise information gain:
From the selection pool, we calculate the state information gain for each feature,

$$\Delta_l = F_l^TF_l - F_l^TE_l(E_l^TE_l)^{-1}E_l^TF_l$$

(5) Feature Selection:
Greedy algorithm to select features best minimize the two objective functions.



(5.1) Greedy Algorithm:
Based on the upper-bound of the objective functions, we prioritize the features with higher values



(6) Complete VO:
Use bucket features and the selected features to perform complete VO.

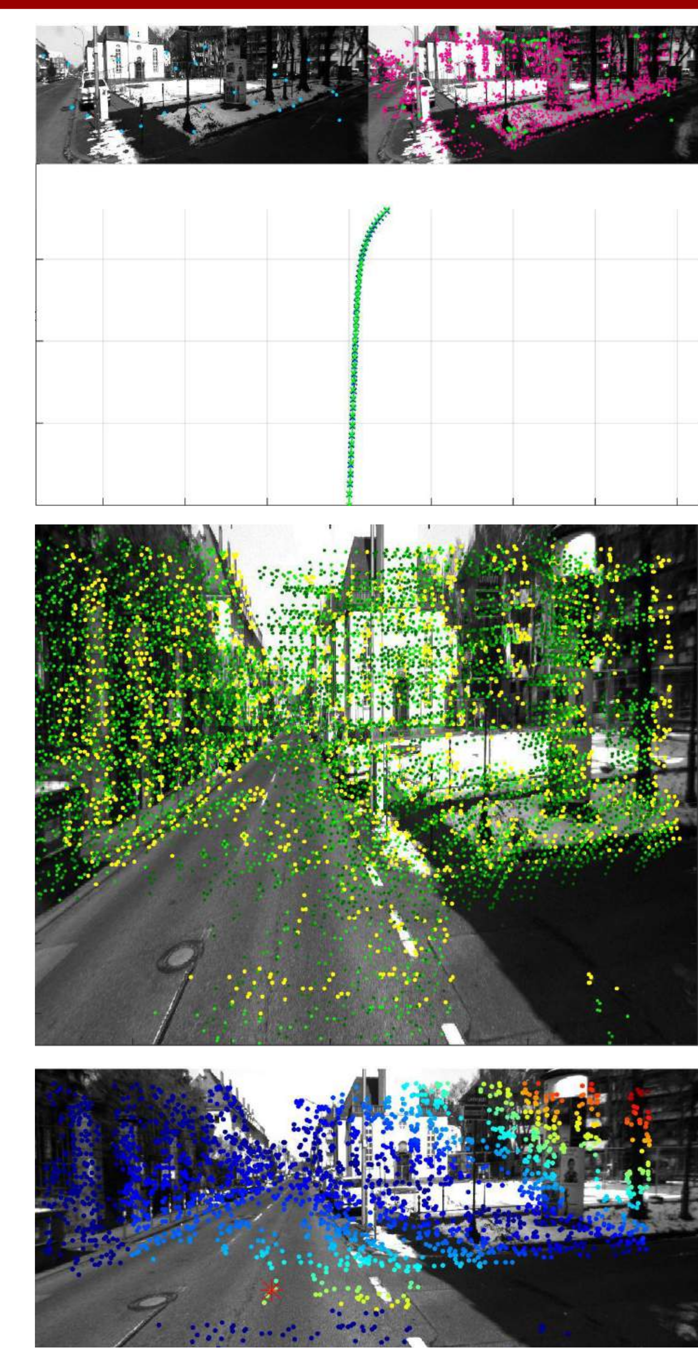
Experimental Setup

We implemented our algorithm using the Matlab toolbox and C++ with LibViso2 [1]. The general pipeline is shown in Figure 3.

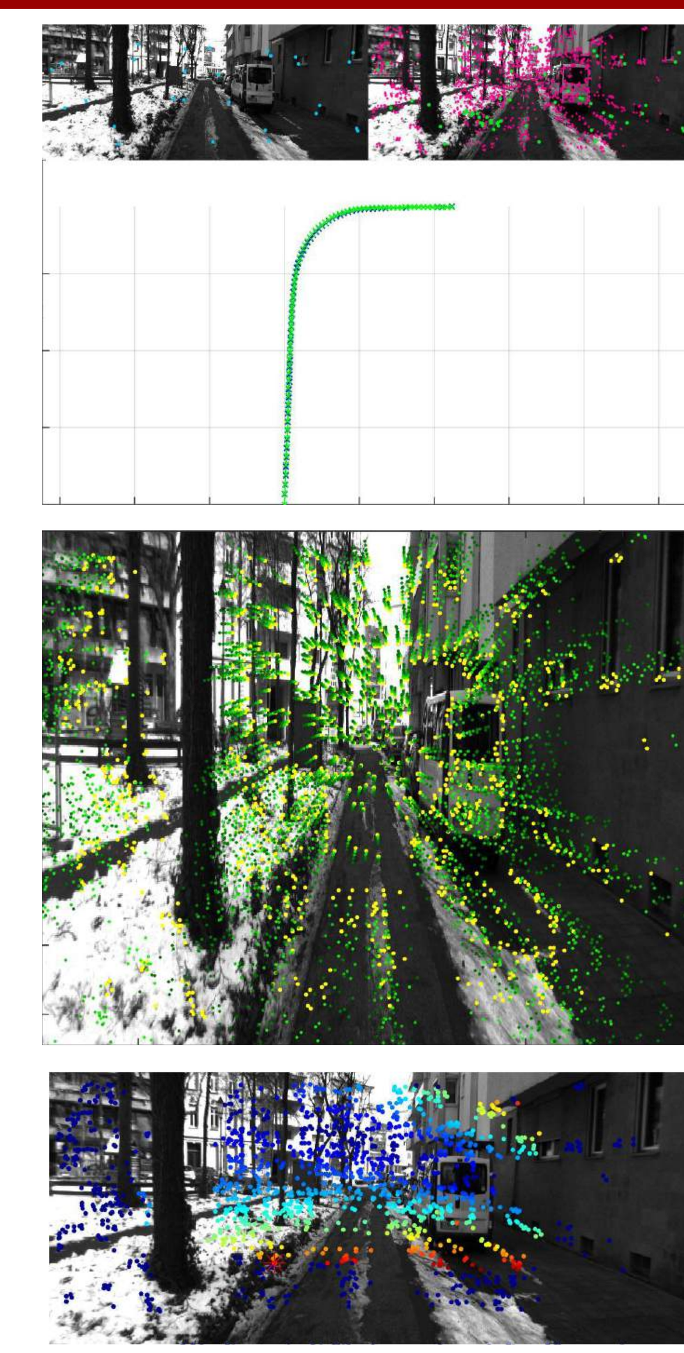
Figure 3. The above block diagram shows the main pipeline of our system..

To validate our algorithm, we used the KITTI dataset [3] with stereo camera input.

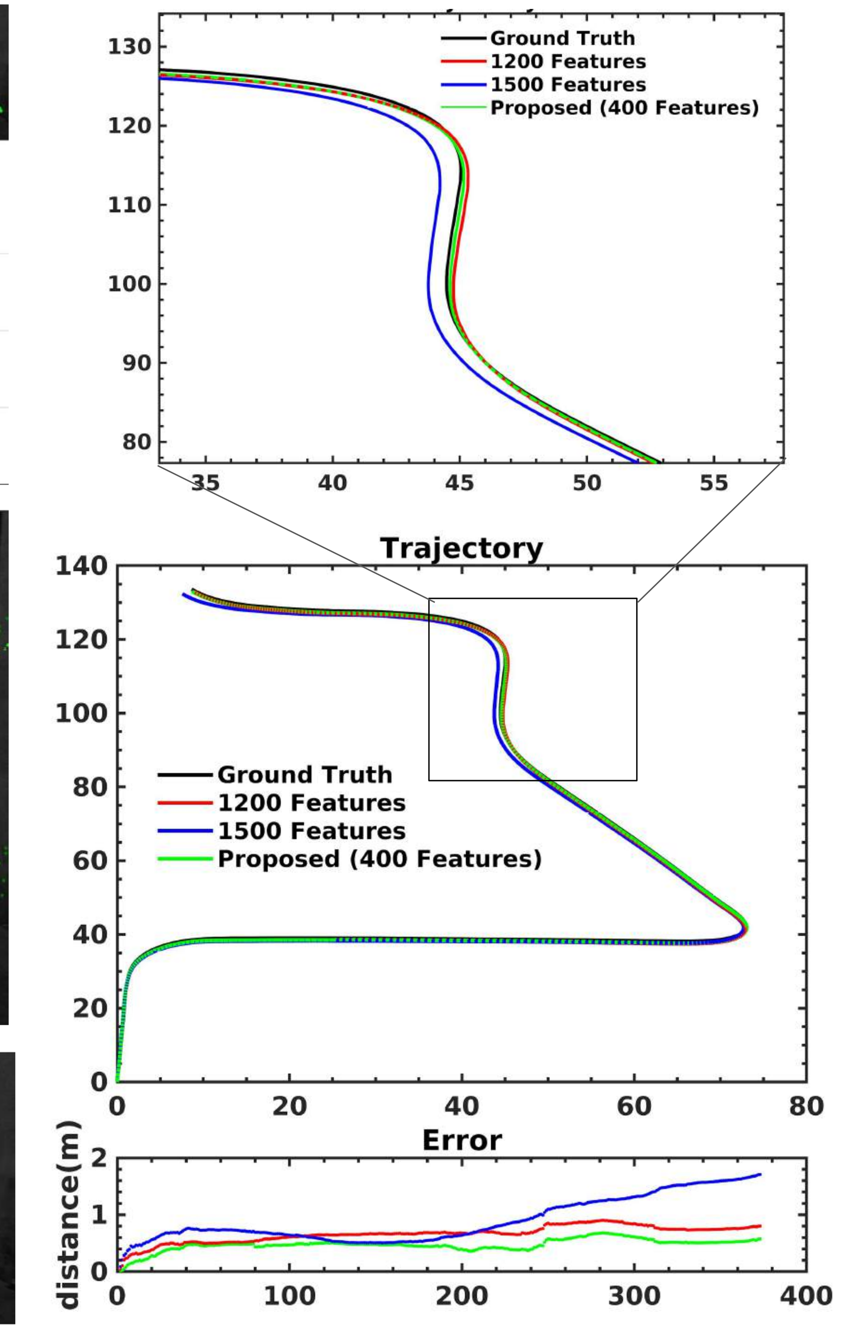
Results



(Top) The vehicle is turning to the right. (Middle) The graph shows that all the feature prediction is moving to the right with rotations. (Bottom) The energy map shows the top right corner and mid-bottom features are the most informative.



(Top) The vehicle is moving straight forward. (Middle) The features are spreading outward radially as vehicle is moving forward. (Bottom) The energy map shows the mid-range ring of features contain the best information because all the closer features will be out of sight.



The top and middle trajectory graphs compare the proposed method to other two original LibViso2 methods, which use three to four times more features. The error plot shows our proposed method out performs both method in absolute error comparing to the ground truth.

Discussions:

Our algorithm relaxes the constraint of IMU requirement in [2], decreases memory usage as well as computation cost, and utilizes pure geometry information to perform state propagation and correction. Two potential applications are (1) when IMU is not available, (2) when under memory or processor limitation, and when IMU breaks during robot operations. The performance of our algorithm can be further improved by providing more information. For example, nonlinear camera motion model, or the confidence of future linear motion can be incorporated to better select predict future frames. The visual appearance of the features, which is not considered in our algorithm can also be included to award distinct features with higher scores.

References:

[1] B. Kitt, A. Geiger, and H. Lategahn, “Visual Odometry Based on Stereo Image Sequences with RANSAC-Based Outlier Rejection Scheme,” in IV, 2010

[2] L. Carlone and S. Karaman, “Attention and anticipation in fast visual-inertial navigation,” arXiv preprint arXiv:1610.03344, 2016.

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” Int. J. Robot. Res., vol. 32, no. 11, pp. 1231–1237, 2013.