

# Accurate Visual Localization for Automotive Applications

Eli Brosh , Matan Friedmann , Ilan Kadar , Lev Yitzhak Lavy , Elad Levi , Shmuel Rippa ,  
Yair Lempert, Bruno Fernandez-Ruiz, Roei Herzig, Trevor Darrell  
Nexar Inc.

## Abstract

Accurate vehicle localization is a crucial step towards building effective Vehicle-to-Vehicle networks and automotive applications. Yet standard grade GPS data, such as that provided by mobile phones, is often noisy and exhibits significant localization errors in many urban areas. Approaches for accurate localization from imagery often rely on structure-based techniques, and thus are limited in scale and are expensive to compute. In this paper, we present a scalable visual localization approach geared for real-time performance. We propose a hybrid coarse-to-fine approach that leverages visual and GPS location cues. Our solution uses a self-supervised approach to learn a compact road image representation. This representation enables efficient visual retrieval and provides coarse localization cues, which are fused with vehicle ego-motion to obtain high accuracy location estimates. As a benchmark to evaluate the performance of our visual localization approach, we introduce a new large-scale driving dataset based on video and GPS data obtained from a large-scale network of connected dash-cams. Our experiments confirm that our approach is highly effective in challenging urban environments, reducing localization error by an order of magnitude.

## 1. Introduction

Robust and accurate vehicle localization plays a key role in building safety applications based on Vehicle-to-Vehicle (V2V) networks. A V2V network allows vehicles to communicate with each other and to share their location and state, thus creating a 360-degree ‘awareness’ of other vehicles in proximity that goes beyond the line of sight. According to the National Highway Traffic Safety Administration (NHTS), such a V2V network offers the promise to significantly reduce crashes, fatalities, and improve traffic congestion [1]. The increasingly ubiquitous presence of smartphones and dashcams, with embedded GPS and camera sensors as well as efficient data connectivity, provides

Figure 1. Method Overview: Given a video stream of images, a hybrid visual search and ego-motion approach is applied to leverage both image representation and temporal information. The VL-GIST representation is applied to provide a coarse localization fix of the image, while the visual ego-motion is used to estimate the vehicle’s motion between consecutive video images. Fusing vehicle dynamics with the coarse location fixes further regularizes the localization error and yields a high accuracy location data stream.

an opportunity to implement a cost-effective V2V “Ground Traffic Control Network”. Such a platform would facilitate cooperative collision avoidance by providing advance V2V warnings, e.g., intersection movement assist to warn a driver when it is not safe to enter an intersection due to high collision probability with other vehicles. While GPS is widely used for navigation systems, its localization accuracy poses a critical challenge for proper operation of V2V safety networks. In some areas such as urban canyons environments, GPS signals are often blocked or partially available due to high-rise buildings [19]. In Fig. 2 we show the accuracy of GPS readings from crowd-sourced data of over 250K driving hours taken in New York City (NYC). The figure demonstrates that the number of rides that suffer from urban canyon effects resulting in GPS errors of 10 m or above is 40%, and that of 20 meters is 20%.

In this work, we propose a hybrid coarse-to-fine approach for accurate vehicle localization in urban environments based on visual and GPS cues. Fig. 1 shows a



















