Joint GPS and Vision Direct Position Estimation

Yuting Ng, Student Member, IEEE and Grace Xingxin Gao, Senior Member, IEEE

BIographies

Yuting Ng is a graduate student in the Aerospace Engineering Department at the University of Illinois at Urbana-Champaign. She received her B.S. degree in electrical engineering, graduating with university honors, from the same university. Her research interests are in advanced signal tracking algorithms, navigation, control, RADAR and UAVs.

Grace Xingxin Gao received her B.S. degree in mechanical engineering and her M.S. degree in electrical engineering from Tsinghua University, Beijing, China in 2001 and 2003. She received her PhD degree in electrical engineering from Stanford University in 2008. From 2008 to 2012, she was a research associate at Stanford University. Since 2012, she has been an assistant professor in the Aerospace Engineering Department at University of Illinois at Urbana-Champaign. Her research interests are systems, signals, control, and robotics. She is a senior member of IEEE and a member of ION.

Abstract—GPS and vision position sensing are complementary. Under open sky environments, GPS sensing is superior due to unimpeded strong signal reception while vision sensing suffers from the lack of unique features. In urban settings, however, vision sensing becomes superior with the abundance of unique characterizing vision features while the GPS sensing performance is hindered by obstruction and multipath. To better leverage upon the complementary nature of the two sensing modes, we propose the Joint GPS and Vision Direct Positioning (GPS+V DP).

GPS+V DP achieves meaningful integration between the two sensing modes, directly estimating the receiver position from the entire raw GPS signal and vision features extracted from the raw camera image. GPS+V DP consists of two synchronized lines of processing: GPS Direct Positioning (GPS DP) and Vision Direct Positioning (Vision DP). GPS DP searches for the composite signal replica that gives the highest correlation against the observed raw GPS signal. This best matched composite signal replica is most likely generated from the most optimal receiver parameters of 3D position, clock bias, 3D velocity and clock drift. Vision DP searches for the geo-tagged reference image that gives the lowest composite feature distance against the observed image. This best matched reference image is most likely generated from the most optimal camera latitude, longitude, heading and tilt. The measurements from both GPS DP and Vision DP are concatenated and used to directly estimate and track the sensors’ position parameters.

We implemented GPS+V DP using our research platform (PyGNSS) and an open source computer vision library (OpenCV). We tested our GPS+V DP receiver architecture using experimental data collected on campus. We demonstrate the functionality of our algorithm through our experimental results.

Vision sensing can provide positioning via camera localization [1]. Examples are Parallel Tracking and Mapping (PTAM) [2], Large-Scale Direct monocular SLAM (LSD-SLAM) [3] and Semi-direct monocular Visual Odometry (SVO) [4]. However, these algorithms require good position initialization, unique and robust vision features, and come at high computational cost. GPS is able to address the above drawbacks by providing the initial position estimate, providing position measurements when vision features are unavailable, and also reduce computational cost by introducing position constraints [5]. On the other hand, GPS sensing is degraded, due to signal obstruction and multipath, in urban environments where vision sensing is superior. Thus we propose the deep integration of the two complementary modes of position sensing via Joint GPS and Vision Direct Positioning (GPS+V DP).

GPS+V DP consists of two synchronized lines of processing, GPS DP and Vision DP, where the key idea is feature matching, direct position estimation, followed by position tracking. In GPS DP, a vector correlator is used to search for the best matched composite signal replica that gives the highest correlation against the observed raw GPS signal [6], [7]. GPS DP then estimates the 3D position, clock bias, 3D velocity and clock drift of the GPS sensor. In Vision DP, image feature extraction [8], feature matching [9]–[11] and homology verification [12]–[14] is used to search for the best matched geo-tagged reference image against the observed camera image [5]. Vision DP then estimates the latitude, longitude, heading and tilt of the vision sensor. Together, the measurements from GPS DP and Vision DP are concatenated and used to track the overall GPS and Vision position sensing parameters of 3D position, clock bias, 3D velocity, clock drift and attitude, obtaining a robust navigation solution.

The rest of the paper is organized as follows. Section II describes the development and implementation of our GPS+V DP framework. Section III describes the experiment setup followed by results and analysis. Finally, Section IV summarizes the paper.

II. JOINT GPS AND VISION DIRECT POSITION ESTIMATION (GPS+V DP)

The Direct Positioning concept for both GPS DP and Vision DP is similar. Both GPS DP and Vision DP are based on Maximum Likelihood Estimation (MLE) of position. In other words, we are searching for the underlying position parameters that generated the observations, which are the GPS raw signal and camera image. We conduct this search by generating geo-tagged replicas and performing feature matching between the replicas and the observation. In GPS DP, the composite signal
Fig. 1. Block diagram describing GPS+V DP. GPS+V DP involves feature matching, direct position estimation and measurement combination in a state-based navigation filter.

Fig. 2. Constrained initialization to reduce computational costs. The initial prediction of the navigation parameters is indicated by the blue cross.

Fig. 3. GPS feature matching using the vector correlator. Vision feature matching using vision features and homography analysis.

Following the initialization of the navigation guesses is feature matching, as shown in Fig.3. GPS DP uses the vector correlation to get the likelihood distribution across the navigation guesses, estimating the mean and variance [6]. Vision DP is unable to get a similar continuous likelihood distribution. The replica sampling in Vision DP is discrete and depends on the reference database. In addition, vision feature matching is susceptible to spurious results. As such, a robust Vision DP algorithm is required.

The robust Vision DP algorithm used in our implementation is inspired by a combination of prior work by other researchers [5], [11]. Two rounds are involved. In the first round, 2D features are extracted from the reference images [17], feature matching is then performed between the reference images and the observed image. The reference images are then ranked according to their overall feature distance to the observed image. The lower the overall feature distance, the better the match. A threshold on the overall feature distance is used to select potential reference images. In the second round, homography analysis is used to verify that the reference image and observed image is generated from a similar camera view [12]–[14].

The GPS DP and Vision DP measurements are then weighted based on their measurement variances and used as direct position measurements in a state-based navigation filter. In our implementation, we used a Kalman Filter (KF) with a 34 millisecond update interval. This update interval is set by the image frame rate of our camera.

III. EXPERIMENT AND ANALYSIS

A. Experiment Setup

The hardware setup is shown in Fig.4. The GPS received signal was collected using a SiGe GNSS v3 Sampler connected to a laptop. The antenna is a small patch antenna. The images were videoed using a handheld Nexus 5X mobile phone on the lowest available resolution of 720x1280 pixels with a frame rate of 30 frames per second (approximately 34 milliseconds between consecutive frames).

GPS DP was implemented using our research platform - PyGNSS. Vision DP was implemented using Open Source Computer Vision (OpenCV) [19], also in Python, and Google
3

Fig. 4. Hardware setup for collecting the GPS signal. Images were collected using a handheld Nexus 5X mobile phone by the front passenger.

Street View API [15], [16] which allows for retrieval of geo-tagged reference images based on latitude, longitude, heading and camera tilt. Both GPS DP and Vision DP were combined to obtain GPS+V DP with PyGNSS.

B. Results and Analysis

A simple experiment was conducted to demonstrate the performance of GPS+V DP. We first analyze the individual effectiveness of GPS DP and Vision DP. Following that, the GPS sensor and vision sensor is synchronized and their measurements deeply integrated in the navigation and tracking loop filter of GPS+V DP.

GPS DP Results

Fig.5 and 6 show the correlation and spectra outputs from GPS DP. Fig.5 shows the uncombined channel correlation and spectra outputs. These are centered near 0 position and clock bias residual, 0 velocity and clock drift residual. Channels where the signal strengths are higher have greater correlation amplitudes and spectra magnitudes. The shapes of the correlation and spectra are also as expected - triangular with a pointy peak for the correlations, raised sinusoidal with a rounded peak for the spectra. Fig.6 shows the same result for the vector correlation and vector spectrum, a noncoherent combination of the channel correlations and spectrums. The magnitude and shape of the vector correlation and vector spectrum is also much greater and smoother than that of the individual channels. As such, the vector correlation and vector spectrum provide a more robust estimate of the navigation parameters than the correlations and spectrums from the individual channels.

The navigation results of GPS DP as displayed on Google Maps is given in Fig.7. The results show that GPS DP is functioning with accurate tracking through most tall buildings and with slightly inaccurate but still robust tracking through the tallest buildings.

Vision DP Results

Fig.8 shows the first round of the image feature matching results from Vision DP where the observed image on the left is compared against the geo-tagged reference images from Google Street View on the right. Fig.9 shows the performance improvement of image feature matching with homography verification. The blue curves show results from feature matching while the red curves show results from feature matching followed by homography verification. Fig.9 also shows the Vision DP result for the urban area on campus.

GPS+V DP Results

After verifying the individual effectiveness of GPS DP and Vision DP in position estimation, the two algorithms are deeply integrated in GPS+V DP. The results are as shown in Fig.10. The two algorithms, GPS DP and Vision DP was first run separately until a navigation trend was found and used for synchronization. Using PyGNSS, GPS DP was run at an update rate of 34 milliseconds to accommodate the update rate of Vision DP which is set by the frame rate of the camera.

GPS+V DP successfully accurately navigated the GPS and vision sensor through an urban area of campus.

IV. CONCLUSION

We propose the deep coupling of GPS and Vision position sensing through Joint GPS and Vision Direct Positioning (GPS+V DP). We designed and implemented the GPS+V DP system. To evaluate the performance of GPS+V DP, we conducted experiments under both open sky and urban scenarios. We then analyzed the GPS DP, Vision DP and GPS+V DP results, demonstrating the effectiveness of deeply coupling GPS and vision via GPS+V DP.

ACKNOWLEDGMENT

Special thanks to Ganshun Lim for his continuous support, encouragement and implementation advice, John Joseph (J.J.) Vergere and Akshay Shetty for assistance in data collection, Anneliese Xin Hui Fang for kindly editing the abstract and proofreading the paper.
Fig. 5. Plots of correlation amplitude against code phase difference on the left and spectra magnitude against carrier frequency difference on the right. 0 code phase residual and 0 carrier phase residual is with respect to the current 3D position and clock bias, current 3D velocity and clock drift.

Fig. 6. Vector correlation and vector spectrum, generated as a noncoherent sum of the correlations and spectrums across all channels. The peaks are centered near 0 residuals indicating that the measurements are unbiased. In addition, the vector correlation and vector spectrum has the same peak magnitude, as expected. Finally, the shape of the vector correlation and vector spectrum is also as expected. GPS DP searches for the peaks.

Fig. 7. GPS DP results as displayed on Google Maps and Google Street View. GPS DP maintains accurate tracking through most tall buildings. GPS DP maintains robust but slightly inaccurate tracking through the two tallest buildings shown in Google Street View.
Fig. 8. The image on the left is captured using a handheld Nexus 5X mobile phone. The images on the right are geo-tagged reference images from Google Street View. The closest geo-tagged reference image is the one with the lowest overall feature distance. Showing the first 5 “matching feature points” for illustration purposes.

REFERENCES


Fig. 9. Vision DP results showing effective positioning via observed images.

Fig. 10. GPS+V DP results showing point of synchronization and subsequent successful position tracking.


