Estimating Vehicle Speed from Timelapse Video
ASM591-AI Project Proposal

Timothy J Wiegman (ID)

2024-10-08

Contents

1   Problem Description

The author has a research dataset that has sparse patches from different sensor modalities, and it is desired to fuse the multiple data sources to help answer questions about the dataset as a whole. As a result, this project will attempt to use one sensor modality (timelapse video) to estimate a value that is given by another sensor modality (GPS movement). The input dataset will consist of a timelapse video (frames recorded at 1 Hz) from a camera mounted to the windshield of a tractor towing a mower attachment. The output labels will be the average ground speed of the tractor at each point in time, with ground truth data calculated from the GPS positions reported over time by an agricultural telematics logger that operated concurrently in some of the monitored tractors. The solution proposed in this project is based on existing techniques, though it is being used to address a novel and significant challenge in the application of computer vision.

This problem is significant because it advances the science of fusing multimodal data sources in a physical environment, and, assuming the artificial intelligence's speed estimates prove to be reliable and accurate, because it allows the original research dataset to be retroactively expanded to cover more operations. While there exist some methods for estimating vehicle velocity from video input [1], they typically rely on grounded external cameras [2] or they estimate only the relative velocity of other vehicles [3], rather than the ground

velocity of the vehicle carrying the camera itself. In addition, there is novelty in the fact that this project aims to use *timelapse* video, rather than conventional video; since timelapses capture images across longer periods, these kinds of videos are working with information that is more temporally sparse than in conventional videos by a factor of at least 24. Timelapse imagery has been used in some existing studies [4], but not applied to active vehicles as in this study.

2 AI Methods

The proposed methods for this project include Deep Learning, Convolutional Neural Networks (CNNs), and Long Short-Term Memory recurrent neural networks (LSTMs). Deep Learning was chosen because it allows combining multiple types of neural networks (like CNNs and LSTMs), and it can automatically learn features of a complex dataset like video in a way that would be much more difficult for humans to manually design. CNNs were chosen because they allow the model to combine information gathered from wider areas of the frame, enabling larger motion patterns to be detected. LSTMs were chosen because they allow the model to combine information from previous frames, as a temporal dimension is necessary for detecting speed. While these methods may be computationally intensive, a properly architected model should still be within the capabilities of the hardware available to the author. If the completed model ends up too inefficient, then the input video can be downsampled to lower resolutions in order to reduce computational requirements and increase processing speed.

3 Prior Work

Estimating the speed of a vehicle from video footage is a well-characterized problem in several applications [5]. However, most of these focus on detecting the speed of *other* vehicles on the road, either by stationary traffic cameras [2] or by onboard driver-assist-type systems [3]. Using an onboard camera to detect the vehicle's *own* ground speed is a less common goal, since most relevant systems would have access to the output of an odometer or speedometer that can directly measure the ground speed. However, since this project assumes that telematics (including recorded speedometer data) are unavailable, velocity must be estimated from other data sources (namely, vision).

This type of problem is referred to as "visual odometry" and there exist some approaches to it in classical computer vision [6], [7]. However, these often rely on stereo cameras for depth estimation [7], while only monocular video exists for the research dataset considered here. Perhaps more relevantly, none of these methods use artificial intelligence and so are unfit for this task by default. However, the general approaches

described by classical vision algorithms may prove useful for this project; one review [8] explains the two main classical approaches to this task: one feature-based and one appearance-based method. The feature-based method involves extracting image features, matching them in each frame, and estimating motion from their relative movements. The appearance-based approach, on the other hand, directly estimates speed from the images themselves, using either optical-flow or template-matching algorithms. The literature notes that appearance-based approaches are more suitable for low-texture environments such as soil, asphalt, and concrete [8], so an AI-based method that acts in similar fashion is recommended for the roadside mowing dataset being studied in this project.

4 Deliverables

The deliverables for this project are threefold: first, a python program that performs supervised training on a properly architected neural network (that uses CNNs and LSTMs for deep learning) on an appropriate dataset. Second, an exported set of fully-trained weights for that model, which allow predicting results on new data. Third, the results of applying the fully-trained model on the gaps in the author's research dataset; these may be summarized into graphs of tractor speeds over time, and the accuracy of the predicted speeds should be calculated based on any periods when GPS data was available for comparison.

5 Feasibility

The project's proposed timeline is shown in the Gantt chart below.

	2024-10-11	2024-10-18	2024-10-25	2024-11-01	2024-11-08	2024-11-15	2024-11-22
Submit Proposal	Complete						
Design NN Architecture	Start	Complete					
Train NN on Supervised Dataset			Start	Complete			
Predict for Rest of Data					Start & Complete		
Prepare Presentation						Start & Complete	
Write Final Report							Start & Complete

References

[1]   A. Marban, V. Srinivasan, W. Samek, J. Fernández, and A. Casals, "Estimating position and velocity in 3d space from monocular video sequences using a deep neural network," in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, pp. 1460–1469.

[2]   D. Bell, W. Xiao, and P. James, "Accurate vehicle speed estimation from monocular camera footage," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, pp. 419–426, 2020.

[3]   I. García-Aguilar, J. García-González, D. Medina, R. M. Luque-Baena, E. Domínguez, and E. López-Rubio, "Detection of dangerously approaching vehicles over onboard cameras by speed estimation from apparent size," *Neurocomputing*, vol. 567, p. 127057, Jan. 2024, doi: 10.1016/j.neucom.2023.127057.

[4]   H. Hendrickx, X. Blanch, M. Elias, R. Delaloye, and A. Eltner, "AI-Based Tracking of Fast-Moving Alpine Landforms Using High Frequency Monoscopic Time-Lapse Imagery," *EGUsphere*, pp. 1–20, Aug. 2024, doi: 10.5194/egusphere-2024-2570.

[5]   D. Fernández Llorca, A. Hernández Martínez, and I. García Daza, "Vision-based vehicle speed estimation: A survey," *IET Intelligent Transport Systems*, vol. 15, no. 8, pp. 987–1005, 2021, doi: 10.1049/itr2.12079.

[6]   D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006, doi: 10.1002/rob.20103.

[7]   A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2008, pp. 3946–3952. doi: 10.1109/IROS.2008.4651147.

[8]   M. O. A. Aqel, M. H. Marhaban, M. I. Saripan, and N. Bt. Ismail, "Review of visual odometry: Types, approaches, challenges, and applications," *SpringerPlus*, vol. 5, no. 1, p. 1897, Oct. 2016, doi: 10.1186/s40064-016-3573-7.