Tracing traffic dynamics with remote sensing

S. P. Hoogendoorn & H. J. van Zuylen
Transport & Planning Department, Delft University of Technology

Abstract

Science has not yet been completely successful in explaining the various puzzling but interesting phenomena of congested traffic flows. The reasons that these phenomena are poorly understood are mainly due to the lack of adequate data needed to develop this theory. Common measurement methods collect only ‘snapshots’ of the situation (i.e. at a limited number of cross-sections, or a single, instrumented vehicle). Such data do not provide sufficient information to study the dynamics of individual drivers in their continuously changing traffic and roadway environment.

This paper presents a new innovative traffic data collection system to detect and track vehicles from aerial image sequences. Besides the longitudinal and lateral positions as a function of time, the system can also determine the vehicle lengths and widths. Before vehicle detection and tracking can be achieved, the software handles correction for lens distortion, radiometric correction, and orthorectification of the image. Vehicle detection occurs approximately each tenth of a second, after which vehicles are tracked both forward and backward in time (10 frames in each direction). The resulting data redundancy is then used to improve accuracy and reliability of the traffic data. Post-processing furthermore entails using Iterative Extended Kalman filtering to improve data quality.

The software was tested on data collected from a helicopter, using a digital camera gathering high-resolution monochrome images of a Dutch motorway near the city of Utrecht. From the test, it is concluded that the techniques for analyzing the digital images can be applied automatically without many problems, while nearly 100% of all vehicles could be detected and tracked.

1 Introduction

The objective of the present study is to develop a data collection method to collect vehicle trajectories (longitudinal and lateral position of the centre of the
vehicle represented by a rectangle as a function of time) and individual vehicle characteristics (vehicle length and width) in particular during congested traffic flow conditions. Given the fundamental requirements of underlying research into driver behaviour, specific demands to the monitoring system hold for the temporal and spatial resolution. For one, the final system must have a resolution of at least 40 cm. Since the roadway length that can be observed by a single camera is determined by resolution of the camera, the used high resolution B&W digital camera (resolution of 1300 pixels \( \times \) 1030 pixels), could observe 1300 \( \times \) 0.4 = 520 m roadway length.

Given average headways between vehicles, and their average speeds, it was decided that the time between two observations should not exceed 0.1 s. It can be shown that in case the specifications above are met, the locations of the vehicles can be determined with an accuracy of 1/4 pixel (= 0.1 m). The resolution of the speeds that are determined from the vehicle positions is thus 1 m/s.

2 Airborne traffic data collection

Data was collected from a helicopter, using a digital camera system. The camera that was used in the end, provided greyscale images at a resolution of 1300 by 1030 pixels with a maximum frequency of 8.6 Hz. The area that each pixel represents is determined by the specifications of the camera (light sensitive chip and lens) and the height at which the images are collected. To decrease the probability that clouds obstruct the observations, it was decided to not fly higher than 500 m. It was decided to use a camera with a 2/3” chip, and a lens of 16 mm. A Personal Computer equipped with a frame grabber was attached to the camera enabling real-time storage of the digital images. The camera was attached to the helicopter and was fixed. No gyroscopic mounting was used to attach the camera, assuming that the resulting vibrations and movements of the helicopter would not influence the quality of the collected data too much.

The data was collected from a height of 500 m, implying that approximately 500 m of roadway could be observed at a single time instant. During the measurement, the helicopter tended to slowly drift from its location.

3 Process overview

The collected raw data consist of large sequences of digital greyscale high-resolution images of the roadway and the traffic it carries. The approach to process these data and determine the individual traffic data consist of the following steps, described in the ensuing of the paper:
1. Image processing (correction for lens distortion, radiometric correction, and orthorectification / geo-correction).
2. Determining background image by removing dynamic objects from reference image.
3. Vehicle detection and tracking.
4. Data post-processing (geo-referencing, handling data redundancy, and filtering).
4 Image processing

The first step entails applying image processing techniques, with the aim to ‘standardize’ the images. This implies that the intensities in all images need to be comparable, all images refer to the same plane of observation, and all images have the same orientation. To this end, the following techniques are used respectively:

1. **Radiometric correction** ensures that the effects of changing lighting conditions (e.g. due to clouds, etc.) are reduced.

2. **Lens distortion correction** removes the so-called pincushion effect from the images. Even though the pincushion effect is very moderate, removing it appears crucial to ensure correct orthorectification.

3. **Orthorectification** or **geocorrection** of the images effectively corrects for the movement of the helicopter (changes in height, rotation, pitch, and yawn).

See [1] for a detailed description of radiometric and lens distortion correction. Let us consider orthorectification in some detail. In an aerial photograph of a rectangular object, the image of this object will only be rectangular if the camera is located exactly above the middle of the rectangle (neglecting the lens distortion discussed above). Otherwise, the perspective of the image will be distorted, depending on the location and the angle of the camera. On top of this, the size of the rectangle will depend on the height at which the images are collected, and will the image be rotated around the vertical axis.

During orthorectification, perspective distortion, scale and rotation of the images are adjusted such that the objects on the image are projected at the same location as the same objects in the reference image \( R \). This reference image \( R \) is determined from several images that are collected at different time instants, and thus reflects a relatively part of the roadway (compared to the individual images). These images are stitched together using dedicated image processing software to form the reference image.

Orthorectification needs control points, which are points in the image that are visible in both the reference image and the processed image. In theory, only 4 control points are needed. However, due to the fact that the determination of the location of the control points cannot be achieved with 100% reliability, 10 to 30 points were used instead of four, which also gives an indication of the accuracy of the process.

To handle the fact that the control points in the reference image and the processed image may be far apart, a special process was developed that uses the information of the control point location in image \( I-1 \) to determine the location of these points in image \( I \). This process starts from reference image \( R \). If \( R \) is not the first image of the sequence, the process is also performed backwards (for images \( R-1, R-2, \) etc.). Two sets of control points have been used.

The first set, the so-called characteristic control points are used for coarse matching. These are around these points are unique for the entire image and can thus be used to match to images where the amount of perspective distortion is large (i.e. the objects on the image and the objects on the reference image are far
apart). Typical objects in these sets are lanterns, gantries, etc. The second set – the \textit{roadsurface control points} – contains points of the roadway surface, i.e. on the reference plane such as the lane markings. These points are not unique and are used for finematching only.

The two phases of the process are:

1. \textit{Coarse matching} finds the control points in the characteristic set of the reference image \( R \) in image \( I \). This is achieved using an iterative approach that maximizes the cross-correlation coefficient of the pixels around the control points in the characteristic set. The search window contains \( 50 \times 50 \) pixels. The accuracy of coarse matching is approximately 1 pixel. The results of coarse matching are used to determine the transformation of image \( I^{-1} \) to image \( I \).

2. \textit{Fine matching} uses the same approach as coarse matching. In this case however, the road surface control points are used instead of the characteristic control points, while the search window is only \( 7 \times 7 \) pixels large.

5 Removing dynamic objects from reference image

Vehicle detection is based on removing a \textit{background image} \( B \) (i.e. the empty road) from the individual vehicles, thus leaving the vehicles (dynamic objects). The approach to removing dynamic objects is simple, and based on the assumption that the road will be empty most of the time (referred to as the \textit{histogram approach}). Considering the median intensity of a pixel in a sequence of images thus reflects the intensity of the road surface, i.e. the background, at that particular location. Application of this procedure to all pixels in the reference images effectively removes the dynamic objects from the scene. Fig. 1 shows an example of applying the approach to the data collected.

![Figure 1: Empty road determined for A27 motorway using histogram approach.](image)

6 Vehicle detection and tracking

Having determined the background image \( B \), the next step of the approach is the detection and tracking of the vehicles. During vehicle detection, the location of the vehicle, and its dimensions is determined. Vehicle tracking entails following the detected vehicle both backward and forward through a series of \textit{geocorrected} images.
6.1 Vehicle detection

For any image, vehicle detection is based on the difference between the current image $I$ and the background image $B$. A first approximation is to use a threshold value to decide whether a pixel represents a vehicle or not. If so, neighboring pixels can either be identified as a vehicle or not. In practice, a number of complicating factors will occur:

1. Both light and dark vehicles will cast shadows, which are generally darker than the roadway surface.
2. Light vehicles have dark spots (windshields, etc.).
3. On occasion, a small vehicle completely drives in the shadow of a big vehicle (e.g. a truck or a bus). As a result, the shadow of the small vehicle disappears. Furthermore, the intensity of the vehicle itself may be close to the intensity of the background image.

The biggest problems are caused by vehicles that have the same intensity as the roadway surface or vehicles that have the same intensity as their shadow. Different approaches have been implemented to resolve these issues (morphological grayscale operations, binary morphological operations, split and merge image segmentation, etc.). We refer to [2] for a detailed description. When the vehicles are detected, the positions as well as the length and width of the vehicles could be established with relative ease.

![Figure 2: Vehicle detection and tracking results.](image)

6.2 Vehicle tracking

The aim of vehicle tracking is to follow the vehicles detected in an image, i.e. to determine their position in the other images. In most cases, tracking is done using an approach similar to the control-point approach used in the orthorectification step (coarsematching and finematching). Its application yields a unique label for all vehicles detected during the vehicle detection step, enabling determination of the vehicle trajectories. To improve the accuracy, the original subimage of the detected vehicle (i.e. determined during detection) as well as its
subimage in the previous image \( I-1 \) were jointly used to determine the position of the vehicle in the current image \( I \).

As a final note, it is mentioned that vehicle detection is performed for each frame \( I \), that is at each 10\(^{th}\) of a second. Vehicle tracking entails both forward and backward tracking of all detected vehicles in the next / previous 10 images. As a result, in each image, a vehicle may result in multiple data points: one from vehicle detection (given it is detected), and multiple data points from tracking. Redundant data is used to improve the accuracy of the traffic data.

7 Data post-processing

This post-processing step entails four aspects:
1. For each detected vehicle, combine the redundant data to a unique vehicle trajectory.
2. Translate image coordinates into real-life coordinates.
3. Check and correct data manually (manually detect missed vehicles, remove false vehicle detections).
4. Improving data accuracy by Iterative Extended Kalman filtering.

7.1 Combining redundant data

In theory, for each vehicle 21 data points are determined in the previous steps: one from detection, 20 from forward and backward tracking. Due to missing detection, in practical situations this number may be substantially less. Using dedicated cluster analysis techniques, these points are clustered, identifying eventual outliers. For the remaining points, the median value is considered and considered to be the best estimate for the vehicles position (lateral and longitudinal), as well as its dimension.

7.2 Translation into real-life coordinates

In the final step, the image vehicle coordinates are translated into real-world coordinates. Scaling and translation determine both the longitudinal position of the rear bumper of the vehicle relative to an arbitrary location on the roadway, and the lateral location of the vehicle relative to the right lane demarcation. To this end, maps of the roadway are used.

7.3 Manual data checking and correction

A dedicated tool has been developed, which lets the used check the resulting detection and tracking results. False detections and incorrect tracking can be removed. The tool also offers the user the opportunity to manually detect vehicles. The system in turn takes care of the required vehicle tracking.
7.4 Iterative Extended Kalman filtering

In the current version of the software, data is collected with an accuracy of 1 pixel (i.e. roughly 40 cm in real-life coordinates), which implies that the speeds can be determined with accuracy in the order of 4 m/s. For most applications, this will be too coarse.

![Figure 3: Application of the Extended Kalman filter to individual vehicle measurements (up: vehicle positions; down: longitudinal speeds).](image)

To resolve this problem, an Iterated Extended Kalman Filter was developed that combines rudimentary knowledge of vehicle dynamics and driver behaviour with data from the data collection system. The filter consists of two elements: a so-called state-space equation (the model) describing the vehicle dynamics, and a measurement equation describing how the model output relates to the data. In the situation presented here, a simple car-following rule is used to predict the acceleration of a vehicle as a function of the distance headway and speed of the preceding vehicle, i.e.

\[
\frac{dv_i}{dt} = v_{i-1} - v_i + \kappa (v_{i-2} - v_i) + \varepsilon
\]

where \( \varepsilon \) is some Gaussian noise term describing the errors in the model. Since only the locations are measured direction, the measurement equation is given by

\[
y_i = r_i + \eta
\]
where $\eta$ is also some Gaussian noise term. The Iterative Extended Kalman Filter is then used to combine both equations and to determine an estimate for the location, speed and acceleration. A similar approach is used for the lateral dimension.

Figure 3 illustrates application of the approach for one vehicle, which has been detected and tracked. The figure shows how the filter effectively smoothes the speed measurements.

7.5 Example results

Figure 4 below shows results of application of the approach outlined in this paper. The data in the figure were collected at weaving area on the A2 motorway near the Dutch city of Utrecht.

![Figure 4: Example vehicle trajectories for situation on A2 motorway near Utrecht, The Netherlands.](image_url)

8 Conclusions

This paper describes a new data collection system that was developed to determine individual vehicle trajectories from high-resolution grayscale images. These images were collected using a digital camera mounted underneath a helicopter, and stored on a personal computer. Having applied a number of photogrammetric operations on the images, nearly all vehicles could be detected
and tracked, yielding both vehicle positions, and vehicle dimensions. The spatial resolution of the raw data was 40 cm; the temporal resolution is 0.12 s. Using the measurement set-up and detection and tracking algorithms described in this paper, it was possible to track vehicles on an area of 500 m length. The accuracy and reliability of the data was increased substantially by multiple detection and tracking of the vehicles, and by using an Iterated Extended Kalman filter.

Acknowledgements

The research described here is part of the research programme “Tracing Congestion Dynamics – with Innovative Traffic Data to a better Theory”, sponsored by the Dutch Foundation of Scientific Research MaGW-NWO. It is co-sponsored by the Traffic Research Center AVV of the Dutch Ministry of Transportation, Public Works and Water Management.

References