Strain, T. J., Wilson, R. E., & Littleworth, R. (2020). Computer Vision for Rapid Updating of the Highway Asset Inventory. *Transportation Research Record*. https://doi.org/10.1177/0361198120928348

Peer reviewed version

Link to published version (if available):
10.1177/0361198120928348

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Sage Publications at https://doi.org/10.1177/0361198120928348. Please refer to any applicable terms of use of the publisher.
Computer Vision for Rapid Updating of the Highway Asset Inventory

Tom Strain¹, R. Eddie Wilson¹ and Roger Littleworth ²

Abstract
In this paper, a decision support system is proposed to assist an analyst in updating the highway roadside asset inventory. The feasibility of the system is tested with assets along an eight kilometre section of the A27 highway on the south coast of England, UK. Survey data from a vehicle equipped with a single forward-facing camera and a GPS-enabled inertial measurement unit (IMU), aerial imagery of the highway, and the asset inventory is fused to develop the system. The camera on the vehicle is calibrated, so that assets may be automatically located within the survey images. The assets are then classified by a state of the art convolutional neural network. Therefore, those assets recorded correctly in the inventory, and those needing further manual inspection are automatically identified. Three different asset types are considered (traffic signs, matrix signs and reference marker posts) and overall 91% of the assets in a withheld test-set are verified automatically. Thus the analyst is presented with a much smaller set of assets for which the inventory is incorrect and which require further inspection. We therefore demonstrate the value in fusing multiple data sources to develop decision support systems for transportation asset monitoring.

Keywords
Computer vision, transport asset management, decision support systems, sensor fusion.

Introduction
An operational and safe highway network is critical to daily life, business and ultimately the economy. Therefore, effective transportation asset management (TAM), defined as “the strategic and systematic process of operating, maintaining, upgrading and expanding physical assets effectively throughout their lifecycle” is vital. To regulate and implement TAM, highway agencies typically follow practises detailed in a TAM plan (TAMP), constructed by government or national level departments. For example, highway agencies from the Netherlands, Ireland, Italy and the UK are currently collaborating with research centres across Europe on the asset monitoring for infrastructure (AM4INFRA) project. The initiative aims to provide an asset management framework to enable “consistent and coherent cross-asset, cross-modal and cross border decision-making”, to ensure standardised, safe and value for money TAM across Europe. Similarly, the US Department of Transport (DOT) provides extensive guidance on how to develop an effective TAMP for state agencies.

The process of asset inventory and field data collection is identified as an important step in a TAMP. Typically, the inventory contains geographical, physical and condition data for each of the assets. Field data may range from manual inspection of the assets to Lidar point clouds. Reportedly, over 70% of U.S. state agencies survey the highway to collect data on assets such as the pavement, signs, guardrails, and lighting units. However, the development of asset monitoring systems for high-cost, low-quantity assets has been prioritised due to funding constraints. Instead, for roadside assets, an analyst usually inspects the survey data, and updates the asset inventory manually. Furthermore, in some regions of the UK, it is common for assets to be monitored via physical inspection in which the inventory is updated on site with a ruggedised tablet (Jacobs, personal communication). Consequently, asset monitoring becomes infrequent and costly.

Cameras provide a relatively cheap sensing capability, therefore, imagery data of assets is often collected in surveys. There are a large number of roadside assets, and a number of reasons why the inventory may be incorrect. For example, the asset may have been moved or damaged between surveys. Therefore, it is difficult for the analyst to comprehensively and accurately maintain the inventory from manual inspection of the images. Thus there is a requirement for tools and systems to assist the analyst in their decision-making when monitoring roadside assets. Furthermore, rich data sources concerning the assets already exist to develop and test the tools.

Usually, the number of assets recorded incorrectly is small compared with those that are correct. Consequently, the analyst spends only a small fraction of their time inspecting those assets that require further attention and updating. Therefore, a decision support system that might automatically identify only those assets that are recorded incorrectly has the potential to reduce the manual workload, and improve upon the current inventory updating process.

¹Department of Engineering Mathematics, University of Bristol, UK
²Jacobs, UK

Corresponding author:
Tom Strain, University of Bristol, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, UK.
Email: tom.strain@bristol.ac.uk
In this paper, one such system that employs computer vision techniques is developed and demonstrated. To prove its feasibility, an inventory of roadside assets installed along an eight kilometre section of the A27 on the south coast of England is considered. To develop the system, imagery and position data from a survey, captured by a vehicle equipped with a single forward-facing camera, and a GPS-enabled inertial measurement unit (IMU) is used. Additionally, data from the asset inventory, and aerial imagery of the highway are also considered. The decision support system is tested on three types of assets: traffic signs, matrix signs and reference marker posts. In developing this system, we demonstrate the value in fusing multiple data sources to support the TAM decision-making process, and to reduce the workload expected of the analyst. Furthermore, the single camera and IMU represents a cheap and deployable sensing capability. It is envisioned that the system will form a future distributed and ubiquitous monitoring capability, so that assets installed on all national highways may be frequently monitored and easily maintained. To the author’s knowledge, this is the first TAM system that automatically verifies both the asset type and position recorded in the inventory, from a single camera and IMU.

**Literature Review**

Computer vision can be broadly understood as two distinct problem areas. Firstly, there are techniques that consider the reconstruction of a 3-dimensional scene in a 2-dimensional image (or collection of 2-dimensional images). Models and algorithms described rigorously in Hartley and Zisserman⁷ such as the pinhole camera, structure from motion and computation of the projection and fundamental matrices, allow such rich views of the world to be created. Secondly, there are pattern recognition based techniques that extract features in images to perform classification via either supervised or unsupervised learning. An overview of this latter problem area is given by Gonzalez and Woods⁸ and Prince⁹. Further, a variety of computer vision techniques employed by intelligent transportation systems are described in Loece et al.¹⁰.

A great deal of work has been undertaken in applying computer vision for the automatic detection of assets in images. In particular, there is a large body of literature on traffic sign detection and classification in images. Arnoul et al.¹¹ developed and deployed a Kalman filter-based system to detect roadside assets from their motion relative to a camera installed on a survey vehicle. Further, Greenhalgh and Mirmehdi¹² presented a support vector machine to classify histogram of oriented gradients (HOG) features generated from images of traffic signs. However, state of the art performance is achieved by deep neural networks; the classifiers proposed by Sermanet and LeCun¹³ and Cireşan et al.¹⁴ achieve respectively 99.17% and 99.46% accuracy on the German Traffic Sign Detection Benchmark dataset. However, these systems classify images in which the sign occupies the majority of the image. Improved methods that can detect and classify traffic signs in whole scene imagery (in the wild) are presented by Zhu et al.¹⁵ and Kryvinska et al.¹⁶.

A number of systems that also estimate the asset position have been developed. Wang et al.¹⁷ presented a stereo camera system to estimate the position of assets identified in imagery and tested different configurations (one vehicle with two cameras and two vehicles each with one camera). In addition, Balali et al.⁶ proposed a method that leverages Google Street View imagery to map the position of the assets.

Furthermore, methods that perform a dense 3-dimensional reconstruction of the highway and its assets are presented in¹⁸,¹⁹. These approaches employ structure from motion techniques to generate point clouds purely from imagery. In addition, semantic texton forests are then employed to perform 2-dimensional segmentation of the image, providing asset classification and 3-dimensional mapping at the pixel level. Unfortunately, point cloud generation from imagery is computationally expensive, however, a cheaper LIDAR-based method for highway asset inventory monitoring is presented in Sairam et al.²⁰. Examples of three roadside asset management systems in operation are shown in Figure 1.

Further, TAM systems that assess the condition of highway assets have been proposed. Systems that automatically evaluate the retroreflectivity of traffic signs, and detect defective road studs (reflective cats eyes) are presented respectively in Chengbo and Yichang²¹ and Mcloughlin et al.²².

In addition to roadside assets, computer vision based methods have been used successfully to monitor the pavement. Zhang et al.²³ developed a deep learning architecture to automatically identify cracks in images of asphalt pavement at the pixel level. The system is developed and tested on illuminated overhead images of the pavement and thus requires a specialised vehicle to be employed. Similarly, the UK-based company Gaist provides pavement condition data, derived from high-resolution imagery to highway agencies²⁴. From dialogue with the company, it is understood that their products do not currently incorporate automation. Rather, they manually label polygons in images with condition data so that future products might be trained to evaluate the highway surface automatically. Gaist also collaborate with the University of York in a project in which refuse trucks are equipped with cameras to monitor the pavement condition²⁵. The trucks typically traverse the same routes on a regular basis and therefore it is hoped that pavement deterioration might be modelled from...
their data. Other readily deployable pavement monitoring capabilities include the system presented in Radopoulou and Brilakis, which considers imagery data from a parking camera mounted on the bumper of a vehicle.

Data Sources

We now describe the development of our own asset monitoring decision support system. We first describe each of the data sources that we use. Samples of the data sources are illustrated in Figure 2.

Asset Inventory

The asset inventory in the UK must adhere to standards set by Highways England and contains information such as the asset position, condition, installation date and maintenance history. However, TAM is often performed by subcontractors and consequently the inventory can be inaccurate and inconsistent. The position, as a UK Ordnance Survey Easting-Northing coordinate \((e, n)\) of each asset is recorded in the inventory. For some assets, physical attributes such as the size (width and height), and the mounting height are also provided. For example, the sizes of traffic signs and matrix signs are recorded in the inventory, but the size of the reference marker posts are not. Therefore, the reference markers are assumed to have a 1m height, 0.5m width and 0m mounting height. In total, 590 individual assets (373 traffic signs, 172 reference markers and 45 matrix signs) are taken forward to develop and test the system.

Aerial Imagery

Overhead imagery of the highway can be viewed using software tools such as ArcMap. The Easting-Northing coordinate for any point on the highway can be obtained from the software, however, the software provides no height relief.

Survey Vehicle

The vehicle is equipped with a GPS-enabled IMU and a single forward-facing camera. As the survey is performed, an image is taken every 2m along the highway. Simultaneously, the vehicle’s heading angle and position are recorded by the IMU. Thus, each image is accompanied with metadata describing the position and direction of the vehicle as the image is taken. The vehicle’s position as an Easting-Northing \((e, n)\) coordinate, and its heading are recorded with a resolution of 0.01m and 0.01° respectively.

Data Pre-processing

The position of each asset is given in the world coordinate system \((e, n, z)\) where \(z\) is the height above the highway surface. However, the decision support system considers the assets as viewed by the camera on the vehicle. Therefore, the survey data is processed so that we may consider the assets relative to the vehicle.

Coordinate Transformation

We first define a coordinate transformation of the arbitrary coordinate \(p = (e_p, n_p, z_p)\) in the world frame, to a coordinate \(p_v = (x_v, y_v, z_v)\) relative to the IMU on the vehicle, where the \(x\)-coordinate axis points along the heading direction of the vehicle, the \(y\)-coordinate axis points orthogonally across the width of the highway, and the \(z\)-coordinate axis continues to point vertically upwards. The coordinate transformation is achieved by a translation so that the IMU is at the origin of \((e, n)\) plane, and a rotation through the vehicle heading angle \(\theta\) about the \(z\)-coordinate axis. Therefore, the coordinate relative to the vehicle is given by \(p_v = R_z(p - o_v)\), where \(R_z\) is the 3-dimensional rotation matrix about the \(z\)-coordinate axis through the heading angle \(\theta\). The translation vector \(o_v = (e_v, n_v, 0)\) is the origin of the new coordinate system, and lies directly beneath the IMU on the highway surface, see Figure 3. The coordinates \(e_v\) and \(n_v\) are the Easting and Northing coordinate recorded by the IMU respectively. The position of the assets relative to the vehicle may then be computed by this transformation.

Heading Correction

The heading direction is computed by the IMU in real-time from the straight-line between the vehicle’s current position and its position as the previous image was taken. The heading angle is then defined as the angle between the heading direction and the Easting-coordinate axis. This method is inaccurate when the vehicle turns a corner, and therefore we re-compute the headings after the survey. Specifically, we compute the vehicle’s heading direction as the \(j^{th}\) image is taken by using a central difference scheme applied to the position of the vehicle when
the \((i - 1)^{th}\) and \((i + 1)^{th}\) image are taken. Our corrected heading angle is thus given by

\[
\theta_i \approx \arctan \left( \frac{n_{i+1} - n_{i-1}}{e_{i+1} - e_{i-1}} \right). \tag{1}
\]

This scheme provides a better approximation for the heading compared to the IMU value that is computed in real time.

**Camera Calibration**

Shortly we will automatically locate assets within the survey images. This is achieved by calibrating the camera on the vehicle, such that the pixel coordinates of an asset in an image from the survey may be computed, given the position of the asset relative to the camera.

Formally, during camera calibration, we aim to recover the camera parameters that project the arbitrary coordinate \(p_v\) to a pixel coordinate \((u, v)\) on an image taken by the camera. The camera parameters are categorised into intrinsic parameters; that is, internal properties of the camera, and extrinsic parameters; that is, how the camera is positioned and orientated on the vehicle relative to the IMU.

It is likely that different surveys will use different cameras, which may be installed onto the vehicle in a number of configurations. Consequently, we may not make any assumptions about the camera or its position and orientation on the vehicle in the survey considered in this paper. Therefore, to locate the assets within the survey images, all intrinsic and extrinsic camera parameters must be calibrated.

**Camera Model**

Both the IMU and camera are positioned on the vehicle at height \(h\). The camera is offset from the IMU at position \((x_0, y_0, h)\) and has roll angle \(\alpha\), tilt angle \(\beta\) and pan angle \(\gamma\); namely, rotations about the \(x\)-coordinate, \(y\)-coordinate and \(z\)-coordinate axis respectively, shown in Figure 3.

The camera on the vehicle is assumed to be a standard pinhole camera, with no radial or tangential distortion, zero skew, and the centre of projection is assumed to be at the centre of the image plane. Therefore the coordinate \(p_v\) is projected to the pixel \((u, v)\) according to

\[
f(p_v; \mathbf{c}) = \lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & 0 & 0 \\ 0 & f_v & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R \\ t \end{bmatrix} \begin{bmatrix} y_v \\ z_v \\ 1 \end{bmatrix}. \tag{2}
\]

Here, \(f_u\) and \(f_v\) are the focal lengths of the camera in pixels, and the translation vector is given by \(t = (y_0, h, x_0)^T\).

The matrix \(R\) is the product of the three rotation matrices through the camera’s pan, tilt and roll angles, about the \(z\)-coordinate, \(y\)-coordinate and \(x\)-coordinate axes respectively. Henceforth, for the sake of brevity, we represent the set of camera parameters \(f_u, f_v, \alpha, \beta, \gamma, x_0, y_0\) and \(h\) by the vector \(\mathbf{c}\). The arbitrary scaling constant \(\lambda\) describes the ray of coordinates relative to the camera that project onto the same pixel coordinate.

**Control Points**

To estimate the camera parameters \(\mathbf{c}\), we collect \(n\) control points; that is, coordinates \(\hat{p}_v\), relative to the vehicle of an object in an image, and the pixel coordinates \(\hat{u}_i = [\hat{u}_i, \hat{v}_i]^T\) of that object in the image for \(i = 1, 2, \ldots, n\). The true camera parameters are then estimated by minimising the Euclidean distance between the pixels computed by Equation 2 for each control point coordinate, and the corresponding control point pixel coordinates, that is

\[
\mathbf{c}^* = \arg \min_{\mathbf{c}} \sum_{i=1}^{n} |\hat{u}_i - f(\hat{p}_v; \mathbf{c})|.
\tag{3}
\]

The nonlinear optimisation is performed with the interior-point method. In practice, software routines such as Matlab’s fmincon may be employed to perform the minimisation. The solver is robust, but needs to be constrained by an upper and lower bound on each of the camera parameters. Each control point provides two equations; one for the \(u\) pixel coordinate and one for the \(v\) pixel coordinate. Therefore, a minimum of four control points are required to solve for the eight camera parameters. However, by using a larger number of control points, and thus over-determining the system, the minimisation becomes more robust to measurement error in the control points.

**Implementation**

In total, 24 control points from multiple images are used to calibrate the forward-facing camera on the survey vehicle. The control points are a mixture of assets from the inventory and road markings on the highway surface. For the assets, the control point coordinates are obtained from the inventory, and for the markings, the coordinates are found from the aerial footage of the highway. The coordinates are collected...
in \( (e, n, z) \) form and validated using the aerial imagery. The coordinates are then subsequently transformed to the coordinate system \( (x, y, z) \) relative to the vehicle. The pixel coordinates for all control points are found by manual inspection of the images. Table 1 shows the results of the calibration and the upper and lower bounds enforced on each parameter. It is assumed that the camera parameters are constant for the whole survey, and therefore the calibration process is only performed once.

Table 1. The camera parameters found by Equation 3. The interior-point method is used to find the optimal set of camera parameters within the lower and upper bounds shown in the table. The directions of the offsets \( x_0 \) and \( y_0 \) are respectively towards the front of the survey vehicle, the central reservation of the highway.

| Camera parameter | Value | Bounds |
|------------------|-------|--------|
| Focal length \( f_x \) [pixels] | 2058 | [0, 10000] |
| Focal length \( f_y \) [pixels] | 2207 | [0, 10000] |
| Roll \( \alpha \) [degrees] | -1.604 | [-30, 30] |
| Tilt \( \beta \) [degrees] | -5.004 | [-30, 30] |
| Pan \( \gamma \) [degrees] | -0.052 | [-30, 30] |
| Offset \( x_0 \) [m] | 1.765 | [-4, 4] |
| Offset \( y_0 \) [m] | 0.088 | [-2, 2] |
| Height \( h \) [m] | 2.301 | [0, 10] |

**Asset Identification**

**Asset Localisation**

With the survey camera calibrated, roadside assets may be automatically located within the survey images. Consider the asset positioned relative to the vehicle at \( a_v = (x_a, y_a, z_a) \), with width \( a_w \), height \( a_h \) and mounting height \( a_m \). The entire asset occupies positions

\[
a_v = (x_a, y_a, z_a)
\]

where \( y_a - a_w/2 \leq \hat{y}_a \leq y_a + a_w/2 \) and \( a_m - \hat{z}_a \leq a_m + a_h \). Further, a kind-of bounding box around the asset may be constructed by the perimeter around the pixels computed by \( f(\hat{a}_v; c') \). However, the physical attributes for some of the assets in the inventory were found (by manual inspection) to be inaccurate. Therefore the asset width and height are extended by 0.3m to ensure that the asset appears within the computed bounding box. This very simple capability already assists the analyst; instead of manually cross referencing the images with the inventory to find a suitable image of the asset, assets are now automatically located within the survey images and presented to the analyst.

**Image Classification**

The work now proceeds to classify the asset within the identified bounding box as either a traffic sign, matrix sign or reference marker and thus, confirm that the inventory entry is correct. The success of deep convolutional neural networks (CNN) for image classification tasks is well documented\(^{31-33}\). To classify an image, broadly put, it is fed through a series of convolutional layers in which features within the image are extracted. The features are then classified using a fully connected neural network. The CNN is parameterised by a set of weights that are optimised with a labelled training set via stochastic gradient descent\(^{34}\).

The architectures and weights for a number of state of the art CNNs, trained on multiple GPUs for several weeks, are freely available online\(^{35}\). However, we may only use the CNNs to classify an image into one of the classes that the CNNs are originally trained on. To exploit the efficacy of CNNs for asset classification, we utilise transfer learning\(^{35}\). The fully connected neural network is modified so that the number of nodes in the final classification layer is the desired number of classes. Then, the network is retrained using a training set specific to our task. The weights of the convolutional layers are not re-trained, as the features extracted by these layers are thought to generalise well to any image classification task.

To re-train the CNN, we construct a training data set with \( 60\% \) of the assets in the inventory. Each asset is located within multiple images, and the thumbnail within the bounding box is cropped out of the image. In total, 1,000 thumbnails of each asset type are extracted from the survey images. So that the CNN might detect if there is no asset present, and thus, find incorrect entries in the inventory, an additional 1,000 random thumbnails that do not contain any assets are also extracted. Furthermore, data augmentation is performed on the training data; that is, copies of the asset thumbnails are randomly sheered, stretched, squeezed and

**Table 2.** Number of each asset type used for training and validation of the CNN, and to test the decision support system. The training and validation set used to train the CNN are constructed from multiple views of \( 60\% \) of the assets in the inventory. The remaining \( 40\% \) of the assets form a test set to determine the effectiveness of the system.

| Asset type    | Train | Validation | Test |
|---------------|-------|------------|------|
| Traffic sign  | 800   | 200        | 149  |
| Reference marker | 800   | 200        | 69   |
| Matrix sign   | 800   | 200        | 18   |
| Random        | 800   | 200        | 0    |
| Total         | 3200  | 800        | 236  |
rotated to create a larger data set. To test the classification accuracy of the CNN, 20% of the training data is withheld as a validation set. Examples of each class in the training set are shown in Figure 4. The remaining 40% of the assets in the inventory are taken forward as a test set. The effectiveness of the system will be determined by considering whether those assets in the test set are identified as correct or requiring further manual inspection. A summary of the labelled training, validation and test data sets, is provided in Table 2.

We choose the GoogleNet architecture, trained on roughly 1.2 million images from 1,000 different classes, as the initial CNN to be re-trained. The network ranked first in the ILSVRC 2014 classification challenge, and is available in the Matlab deep learning toolbox. The original fully connected neural network is changed so that the final classification layer has 4 nodes, and is therefore suitable for asset classification. The CNN is re-trained for 20 epochs, with a batch size of 32 and a learning rate of $1 \times 10^{-4}$. A classification accuracy of 98% is achieved on the withheld validation set.

### Rapid Asset Inventory Update

We now test the method by considering those assets in the test set that are verified, and those identified for further manual inspection. Each asset is considered individually. Firstly, the survey image for which the vehicle is closest to the asset and contains the asset’s bounding box is found. The asset is then cropped out of the image and classified by the CNN. For those assets classified as the asset type recorded in the inventory, we verify that the assets are recorded correctly in the inventory. Alternatively, assets that are classified as a different asset type than that recorded in the inventory, or as the empty class, are identified as assets that require further manual inspection. The inventory may be incorrect if the asset has been moved or removed from the highway, but the inventory has not been updated correspondingly, or, the asset’s geographical or physical attributes are incorrect in the inventory. Additionally, an asset is identified for further manual inspection if the view of the asset is impeded, or the CNN classifies the asset incorrectly. The system does not determine the reason for an incorrect classification. Rather, the relatively small number of assets for which this is the case are rapidly identified and presented to the analyst, who may then correspondingly update the inventory. Table 3 shows the number of assets that are verified and the reason for incorrect classification. Overall, 91% of highway-side assets in the test set are verified automatically. Examples of the system in operation are shown in Figure 5.

| Asset type         | Number | Verified | Percentage | CNN error | Impeded | Incorrect inventory |
|--------------------|--------|----------|------------|-----------|---------|---------------------|
| Traffic sign       | 149    | 136      | 91         | 5         | 2       | 6                   |
| Reference markers  | 69     | 64       | 93         | 2         | 2       | 1                   |
| Matrix sign        | 18     | 15       | 83         | 2         | 0       | 1                   |
| Total              | 236    | 215      | 91         | 9         | 4       | 8                   |

### Discussion

The system automatically verifies 91% of the assets in the withheld test set, which would otherwise be manually inspected by an analyst. The reduction in the number of assets that the analyst is required to manually inspect demonstrates the beginnings of a promising operational asset monitoring capability. We do not consider whether the asset has been correctly verified. However, if an asset is correctly localised and classified by the CNN (which achieved a 98% classification accuracy), it is highly likely that the asset is recorded correctly in the inventory. The results show that the system performs well for all the asset types considered — although of the 21 assets identified for further manual inspection, 13 were due to an incorrect classification by the CNN or an impeded view of the asset. However, going forward, we believe that training the CNN with a larger data set, and considering multiple images of the asset is likely to address these incorrect asset classifications.

There are a number of other potential refinements that might improve our method. Currently, three asset types have been considered, however, all roadside assets might be monitored by re-training the CNN with a labelled training data set containing all asset types. Should the vehicle be equipped with a more sophisticated IMU, a more accurate heading direction might be computed, and thus, the heading correction would not be necessary. Additionally, using a camera with known intrinsic parameters may simplify the camera calibration method, as only the six extrinsic camera parameters would be unknown. Furthermore, the camera parameters are assumed to be constant throughout the entire survey, and thus the camera is only calibrated once. However, it is possible that the camera angles might change should the vehicle jolt. Therefore, a sequential calibration method might be employed, so that the camera parameters are continually updated. Alternatively, auto-calibration methods such as those presented in Golparvar-Fard et al. and Uslu et al. might remove the need to manually collect control points.

Our system has several notable differences to existing TAM systems. Typically, assets (commonly traffic signs) are automatically detected in the images and subsequently processed. In contrast, our system employs a calibrated camera to project the assets (irrespective of asset type) into the survey imagery, and thus our system is more robust (assuming an accurate calibration). On the other hand, along with a number of existing systems, we also exploit the efficacy of CNNs on image classification problems to perform asset classification.

Secondly, existing TAM systems consider primarily a data source from an external sensor (camera or LIDAR for example) and subsequently detect and localise assets. However, our system is built upon the inventory, that is, we...
firstly consider each asset as it is recorded in the inventory and subsequently verify its entry. Therefore, our method more accurately reflects the role of an analyst, and can thus provide rapid decision support within the current asset monitoring process.

Conclusions

In this paper, a decision support system designed to assist those responsible for the maintenance of highway asset inventories is proposed. An inventory of assets along an eight kilometre section of the A27, data from a single forward-facing camera and a GPS-enabled IMU installed on a survey vehicle, and aerial imagery of the highway, was used to develop and test the system. By collecting control points, the camera was calibrated so that assets may be automatically localised within survey imagery, and a state of the art convolutional neural network was re-trained to classify the assets. The inventory entry for an asset may then be verified if that asset is successfully localised and classified as the asset type recorded in the inventory — consequently the assets that require further manual inspection from the analyst are rapidly identified. The effectiveness of the system is determined by considering those assets in a withheld test set that are automatically verified, and those identified for further manual inspection. Overall, 91% of the assets are automatically verified, which would otherwise generate manual work. We have therefore proven the feasibility of the system, and its benefit to an analyst.

There are a number of limitations in the currently presented system, and therefore opportunities for further research and development. Currently, the system does not consider the asset condition, which is typically recorded in the inventory. Therefore, to further assist the analyst, future research should consider automatic evaluation of the asset condition. Furthermore, the system only considers assets on the highway that are in the inventory (correctly or incorrectly). Therefore, the system might also be extended to automatically identify assets that are not in the inventory at all; assets that have recently been installed, for example. To achieve this, improving the CNN so that assets may be localised and classified from the whole scene of the highway, rather than from cropped asset thumbnails, is likely to be the way forward. Assets identified in the survey imagery may then be searched for in the inventory, and thus, assets missing from the inventory might be identified.

In addition, the system relies on a GPS-enabled IMU installed on the survey vehicle and therefore may perform poorly in closed environments such as tunnels or urban areas. In this paper, we do not consider such environments and the system is assumed to be in operation on an open, multi-lane highway. Therefore, to make the system more robust, future work should also focus on fusing the IMU with a source that might provide an estimate of the vehicle’s position in closed environments. Such an estimate might be provided by a visual odometry-based (VO) method in which the relative camera motion is estimated purely from the survey imagery.

Acknowledgements

The authors would like to acknowledge the data, support and funding from Jacobs, Highways England and the Engineering and Physical Sciences Research Council that have made the work presented in this paper possible.

Author Contributions

The authors confirm that T.J. Strain, R.E. Wilson and R.M. Littleworth contributed to all sections in the paper. All authors reviewed the results and approved the final version of the manuscript. The authors do not have any conflicts of interest to declare.

References

1. Office of Road and Rail. *Highways monitor: update on*
Highways England’s capital planning and asset management. TSO, 2017.

2. National Academies of Sciences, Engineering, and Medicine. *Use of transportation asset management principles in state highway agencies*. The National Academies Press, 2013.

3. European Commission. Common framework for a European life cycle based asset management approach for transport infrastructure networks, 2017. https://cordis.europa.eu/project/rcn/204966/factsheet.en. Accessed July 5, 2019.

4. American Association of State Highway and Transportation Officials. AASHTO transportation asset management guide: a focus on implementation. American Association of State Highway and Transportation Officials, 2011.

5. US Department of Transport. Common Q’s and A’s pertaining to transportation asset management, 2018. https://www fhwa dot gov/asset/faq cfm. Accessed June 6, 2019.

6. Balali V, Ashouri Rad A and Golparvar-Fard M. Detection, classification, and mapping of u.s. traffic signs using google street view images for roadway inventory management. *Visualization in Engineering* 2015; 3(1): 15.

7. Hartley R and Zisserman A. *Multiple View Geometry in Computer Vision*. 2 ed. Cambridge: Cambridge University Press, 2004.

8. Gonzalez RC and Woods RE. *Digital Image Processing*. 3 ed. Upper Saddle River, NJ: Prentice Hall, 2008.

9. Prince SJD. *Computer Vision: Models Learning and Inference*. Cambridge: Cambridge University Press, 2012.

10. Loce RP, Bala R and Trivedi M. *Computer Vision and Imaging in Intelligent Transportation Systems*. 1st ed. Hoboken, NJ: Wiley-IEEE Press, 2017.

11. Arnoul P, Viala M, Guerin JP et al. Traffic signs localisation for highways inventory from a video camera on board a moving collection van. In *Proceedings of Conference on Intelligent Vehicles*. pp. 141–146.

12. Greenhalgh J and Mirmehdi M. Real-time detection and recognition of road traffic signs. *IEEE Transactions on Intelligent Transportation Systems* 2012; 13(4): 1498–1506.

13. Sermanet P and LeCun Y . Traffic sign recognition with multi-scale convolutional networks. In *The 2011 International Joint Conference on Neural Networks*, pp. 2809–2813.

14. Cireşan D, Meier U, Masci J et al. Multi-column deep neural network for traffic sign classification. *Neural Networks* 2012; 32: 333 – 338.

15. Zhu Z, Liang D, Zhang S et al. Traffic-sign detection and classification in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2110–2118.

16. Kryvinska N, Poniszewska-Maranda A and Gregus M. An approach towards service system building for road traffic signs detection and recognition. *Procedia Computer Science* 2018; 141: 64–71.

17. Wang KC, Hou Z and Gong W. Automated road sign inventory system based on stereo vision and tracking. *Computer-Aided Civil and Infrastructure Engineering* 2010; 25(6): 468–477.

18. Golparvar-Fard M, Balali V and de la Garza JM. Segmentation and recognition of highway assets using image-based 3D point clouds and semantic texnot forests. *Journal of Computing in Civil Engineering* 2015; 29: 04014023.

19. Uslu B, Golparvar-Fard M and de la Garza JM. Image-based 3d reconstruction and recognition for enhanced highway condition assessment. In *Congress on Computing in Civil Engineering, Proceedings*. pp. 67–76.

20. Sairam N, Nagarajan S and Ornitz S. Development of mobile mapping system for 3d road asset inventory. *Sensors* 2016; 16(3): 367.

21. Ai C and Tsai YJ. An automated sign retroreflectivity condition evaluation methodology using mobile lidar and computer vision. *Transportation Research Part C: Emerging Technologies* 2016; 63: 96 – 113. DOI:https://doi.org/10.1016/j.trc.2015.12.002. URL http://www.sciencedirect.com/science/article/pii/S0968090X15004179.

22. McLoughlin S, Deegan C, Mulvihill C et al. Mobile mapping for the automated analysis of road signage and delineation. *IET Intelligent Transport Systems* 2008; 2(1): 61–73.

23. Zhang A, Wang KCP, Li B et al. Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering* 2017; 32(10): 805–819.

24. Gaist. Highwayview roadscape information has changed forever, 2017. https://www gaist co.uk/ highwayview. Accessed June 17, 2019.

25. Department for Transport. £200 million funding boost for england’s roads, 2017. https://www.gov.uk/government/news/ 200-million-funding-boost-for-englands-roads. Accessed June 17, 2019.

26. Radopoulou SC and Brilakis I. Automated detection of multiple pavement defects. *Journal of Computing in Civil Engineering* 2017; 31(2): 04016057.

27. Highways England Asset Information Group. *Asset data management manual, part 2 - requirements and additional information*. Highways England, 2019.

28. Director General of the Ordnance Survey UK. *A brief description of the national grid and reference system*. His Majesty’s Stationery Office (HMSO), 1946.

29. Goldstein H. *Classical Mechanics*. 3 ed. Boston, MA: Addison-Wesley, 2002.

30. Sturm P. *Pinhole Camera Model*. Boston, MA: Springer US, 2014. pp. 610–613.

31. Zeiler M and Fergus R. Visualizing and understanding convolutional networks. *European Conference on Computer Vision (ECCV)* 2013; 8689: 818–833.

32. Krizhevsky A, Sutskever I and Hinton GE. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097–1105.

33. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*.

34. Bottou L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT’2010*, pp. 177–186.

35. Shin HC, Roth H, Gao M et al. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging* 2016; 35: 1285–1298.

36. Szegedy C, Wei Liu, Yangqing Jia et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9.

37. Russakovsky O, Deng J, Su H et al. Imagenet large scale visual recognition challenge. *International Journal of Computer
Figure Captions

Figure 1. Three existing asset management systems in the literature. Panel (a) shows two examples of the 3-dimensional image reconstruction and asset segmentation described in Golparvar-Fard et al.\textsuperscript{18}. Panel (b) shows the labelling exercise undertaken by Zhu et al.\textsuperscript{15}. The bounding box and class for each traffic sign is manually collected and used to train a system capable of detecting traffic signs in a whole scene image of the highway. The system proposed by Balali et al.\textsuperscript{6} is shown in panels (c) and (d). A traffic sign is detected from a Google Street View image, and its position is subsequently computed and mapped.

Figure 2. Samples of the data sources used to develop the asset monitoring decision support system. Panel (a) shows an image of a traffic sign and a reference marker on the highway taken by the forward-facing camera on the survey vehicle. The inventory entries for the two assets are shown in panel (c), and the vehicle position and heading provided by the IMU are shown in panel (d). Panel (b) shows an example of the aerial imagery of the highway. This data source provides a view of the A27 in the $(e, n)$ plane, and may be used to find the coordinates of assets and road markings on the surface of the highway. The position of the survey vehicle as the image shown in panel (a) is taken, and the two assets are marked on the aerial imagery.

Figure 3. An illustration of the IMU and camera on the vehicle. Panel (a) shows the vehicle in the $(x, z)$ plane. The IMU and camera are positioned on the vehicle at a height $h$ above the highway surface. The camera has a roll angle $\alpha$ and tilt angle $\beta$. The origin of the coordinate system relative to the vehicle, $o_v$, lies directly underneath the IMU on the highway surface. Panel (b) shows the offset of the camera relative to the IMU $(x_o, y_o)$ and the camera’s pan angle $\gamma$ in the $(x, y)$ plane.

Figure 4. Examples of the training data set for each class. Traffic signs, matrix signs, reference markers and the empty (random images) class are shown in panels (a) to (d) respectively.

Figure 5. Six examples of the system in operation. Panels (a) to (c) show assets that are localised and classified as the correct asset type given in the inventory. Consequently, we automatically verify the inventory entry for those assets. Contrarily, assets that are classified as a different asset type in the inventory, or random, are shown in panels (d) to (e). A reference marker that has been moved without updating the inventory, and a matrix sign for which the mounting height is recorded incorrectly, are shown in panels (d) and (e) respectively. Panel (f) shows a reference marker for which the view is impeded by another vehicle.