

Article

Methods for Magnetic Signature Comparison Evaluation in Vehicle Re-Identification Context

Juozas Balamutas ¹, Dangirutis Navikas ¹, Vytautas Markevicius ¹, Mindaugas Cepenas ¹, Algimantas Valinevicius ¹, Mindaugas Zilys ¹, Michal Prauzek ², Jaromir Konecny ², Michal Frivaldsky ³, Zhixiong Li ⁴ and Darius Andriukaitis ^{1,*}

¹ Department of Electronics Engineering, Faculty of Electrical and Electronics Engineering, Kaunas University of Technology, Studentu St. 50, LT-51368 Kaunas, Lithuania; juozas.balamutas@ktu.lt (J.B.); dangirutis.navikas@ktu.lt (D.N.); vytautas.markevicius@ktu.lt (V.M.); mindaugas.cepenas@ktu.lt (M.C.); algimantas.valinevicius@ktu.lt (A.V.); mindaugas.zilys@ktu.lt (M.Z.)

² Department of Cybernetics and Biomedical Engineering, VSB—Technical University of Ostrava, 708 00 Ostrava, Czech Republic; michal.prauzek@vsb.cz (M.P.); jaromir.konecny@vsb.cz (J.K.)

³ Department of Electronics and Mechatronics, Faculty of Electrical Engineering and Information Technologies, University of Žilina, 010 26 Žilina, Slovakia; michal.frivaldsky@uniza.sk

⁴ Department of Manufacturing Engineering and Automation Products, Opole University of Technology, 45-758 Opole, Poland; z.li@po.edu.pl

* Correspondence: darius.andriukaitis@ktu.lt; Tel.: +370-37-300-519

Abstract: Intelligent transportation systems represent innovative solutions for traffic congestion minimization, mobility improvements and safety enhancement. These systems require various inputs about vehicles and traffic state. Vehicle re-identification systems based on video cameras are most popular; however, more strict privacy policy necessitates depersonalized vehicle re-identification systems. Promising research for depersonalized vehicle re-identification systems involves leveraging the captured unique distortions induced in the Earth’s magnetic field by passing vehicles. Employing anisotropic magneto-resistive sensors embedded in the road surface system captures vehicle magnetic signatures for similarity evaluation. A novel vehicle re-identification algorithm utilizing Euclidean distances and Pearson correlation coefficients is analyzed, and performance is evaluated. Initial processing is applied on registered magnetic signatures, useful features for decision making are extracted, different classification algorithms are applied and prediction accuracy is checked. The results demonstrate the effectiveness of our approach, achieving 97% accuracy in vehicle re-identification for a subset of 300 different vehicles passing the sensor a few times.

Keywords: intelligent transportation systems; vehicle re-identification; magnetic signature; signal classification



Citation: Balamutas, J.; Navikas, D.; Markevicius, V.; Cepenas, M.; Valinevicius, A.; Zilys, M.; Prauzek, M.; Konecny, J.; Frivaldsky, M.; Li, Z.; et al. Methods for Magnetic Signature Comparison Evaluation in Vehicle Re-Identification Context. *Electronics* **2024**, *13*, 2722. <https://doi.org/10.3390/electronics13142722>

Academic Editor: Felipe Jiménez

Received: 22 June 2024

Revised: 5 July 2024

Accepted: 10 July 2024

Published: 11 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction and Related Work

Usage of magnetic sensing in intelligent transportation systems (ITS) is rapidly gaining popularity [1–3] compared to more traditional video processing [4,5]. Various systems for road traffic surveillance largely utilize MEMS (micro-electromechanical systems) magnetometers for the purpose of vehicle detection and classification [6,7]. Examples of such applications include vehicle speed calculation [8,9], length estimation [10], classification into defined categories [11], smart parking [12] and autonomous driving [13]. Wireless magnetic sensor networks are a popular choice for information acquisition. This solution has the advantages of low cost, low power consumption and miniaturization. Study [14] highlights the drawback of magnetic sensing for a parking application, which is the weak magnetic region under the vehicle. It is difficult to separate this magnetic signal from ambient noise since the magnetometers are mounted in the middle of the parking lane. To solve this problem, a UWB radio channel was proposed. The authors of [15] investigated anomaly detection for a moving magnetic target in controlled laboratory conditions. The

velocity–frequency relationship, direction angle and closest point of approach distance were highlighted in this study. The mathematical relationship between the detection platform and magnetic target was established to account for velocity differences [16].

Machine learning methods for magnetic signal analysis are gaining popularity. In [11], a method for vehicle detection and recognition using machine learning, input vectors composed from cepstrum energy and Mel frequency cepstral coefficients was utilized. Vehicles were classified into five defined categories with high accuracy, although the sample size for training and testing was not very large. The authors of [17] defined vehicle classification as the process of assigning each vehicle to a pre-defined vehicle class based on some features extracted from its magnetic signature. The task of vehicle re-identification [18] was applied in determining segment travel times. Using an inductive loop detector, the authors implemented a decision tree using statistical features like lane speed and travel duration estimation. Classification into four different classes was performed in [19]. The authors extracted 10 different statistical features from the z axis of the triaxial magnetometer—such as mean, variance and local minimums and maximums. Classification was performed using different machine learning algorithms (BP NN, SVM, Random Forest) on a dataset containing more than 4500 vehicle entries. The classification result reached 80%. Real-time vehicle classification into nine classes was performed in [20] utilizing three-layer feedforward artificial neural networks (ANN). The authors established that using only time-domain features requires low computation and few memory resources, which enables easier implementation and real-time operation. A frequent problem in vehicle classification is an imbalanced dataset—certain types of vehicles like buses and motorcycles occur less often compared to sedans and hatchbacks. A solution called SMOTE was proposed by the authors of [21]. According to certain rules, the SMOTE algorithm randomly generates new minority sample points, which will be merged into the original. Utilizing this algorithm with a K-nearest neighbor (KNN) classifier, they managed to classify vehicles by magnetic signatures into four classes with 95% accuracy. Balid, in [22], used various statistical measurements to classify vehicles according to an FHWA scheme. Different classifiers were tested, all reaching almost 98% accuracy.

In [23], researchers explored the potential of estimating vehicle location using a magnetic sensor array. Although the array consisted of only two sensors, the mathematical principles can be extended to larger arrays. The vehicle, traveling parallel to the sensor array, is modeled as a magnetic dipole, generating an ideal dipole magnetic signature. While the overall accuracy was not specified, the study concluded that it is feasible to detect vehicle location using a magnetic sensor array. Similarly, Kwong [24] developed a system for the real-time estimation of travel time across the arterial segment of 1.5 km by identifying the same vehicles. The matching procedure is based on a statistical model of signature distance. The model can be used to predict rates of correct, incorrect and missed matches. As stated by [25], vehicle re-identification (identifying the same exact vehicle in traffic) can give essential information for traffic management about travel time and origin–destination matrices. Real-time traffic data are essential for optimizing traffic flow and reducing travel time and greenhouse gas emissions. The authors collected 261 temporal magnetic signatures from 25 different vehicles in the parking lot. A total of 10 sensors with 25 cm spacing and 200 Hz sampling frequency were used. Their findings included that the Euclidean distance and DTW (dynamic time warping) algorithms are suitable for same vehicle identification; however, performance is linked to the lateral position of the vehicle on the road.

Fourteen wireless magnetic sensors were installed in an array [26], and vehicle signature features consisting of a collection of peak values were collected. By comparing the upstream and downstream detectors' registered signatures and calculating different statistical features (mean, standard deviation), arrays were compared using distance metrics. The authors suggested that using their developed algorithm, the mismatch rate is around 8%, and, for congested traffic conditions, it reaches 14%. The authors of [27] utilized six magnetic sensors with 20 cm spacing for AVMR (automatic vehicle model recognition)—given

an unknown vehicle's signature and set of a database vehicle signatures, the goal is to associate the correct vehicle model. Re-identification using DTW and machine learning models (k-NN, LMD, Gaussian classifier) has shown promising results; however, there are no experiments indicating how the algorithms would perform with new, unseen data.

The advantages of wireless sensor networks utilizing magnetic sensing are summarized in [28]—they do not require an external power supply, they have a compact size, they minimally damage the road surface and they have resistance to interference from the urban environment. The authors emphasize as a disadvantage that magnetic sensors do not provide information about the direction of the vehicles—available systems cannot detect which gate the car entered the area through or left through. With our work, we try to challenge this statement and demonstrate that it is possible to determine specific vehicle traveling patterns in a specified area with great accuracy.

The available research is concentrated on single-magnetometer feature extraction for vehicle classification to strictly defined categories. Not much work is available on the vehicle re-identification problem—finding the same vehicle passing over sensors by comparing its magnetic signature with previously recorded signatures. Articles revolving around vehicle re-identification methods involved testing in restricted conditions or with a small data subset. In this work, a feature extraction method for a magnetometer array is proposed, and different machine learning methods are compared for same vehicle re-identification in real traffic conditions. Focus is shifted to define better input features and compare different classification algorithms.

2. Materials and Methods

Vehicles are registered using our system, presented in a previous work [29], updated to 15 sensors with 10 cm spacing between sensors. The system constantly measures Earth's magnetic field with tri-axis magnetic sensors, each sampled at 1 kHz frequency. Measured values are constantly detrended and register only change in the magnetic field. The magnetometer's measuring range is set to $\pm 400 \mu\text{T}$, and practical experiments showed that the measured ambient noise is $1.5 \mu\text{T}$ when no vehicles or other metallic objects are near the sensors. The system starts to register magnetic signatures, then the center sensor of the array's calculated module exceeds $5 \mu\text{T}$ and waits for a decrease in amplitude or timeout condition. Collected signatures for all sensors (matrix consisting of 15 $x/y/z$ arrays) are transmitted to a central server. Additionally, images from a temporarily installed camera with license plate recognition are collected and saved together with a .csv file.

In the preprocessing stage, raw collected $x/y/z$ values undergo low-pass filtering with a 100 Hz cutoff frequency in order to smooth the signal. Modules for every sensor are calculated according to the following formula:

$$mod_n = \sqrt{x_n^2 + y_n^2 + z_n^2} \quad (1)$$

where mod_n —calculated module, $x/y/z$ —raw registered values, n —sensor number.

Signature length depends on vehicle speed and vehicle length. Many feature extraction methods and machine learning algorithms expect input data to be of uniform length. Signals with a different number of points might lead to inaccuracies and inconsistencies in the analysis. In the deployed signature collection system, vehicles travel with speeds ranging from 30 km/h to 60 km/h. Most of the time, they are light passenger cars with an average length of 4.5 m. The collected signatures sample count ranges from 600 to 1300 points. Based on these observations, it was decided to resample all signals to a fixed number of 500 points. This process involves interpolating or resampling the signals so they all have the same length regardless of the original number of data points. The vehicle signature sampling process is initiated by a single magnetometer located in the center of the array; other sensors values might be shifted with reference to this sensor depending on the angle at which the vehicle is passing through the sensors. To address these issues, all signature

values are cropped to a specified threshold—5 μ T. The cropping and resampling process is illustrated in Figure 1.

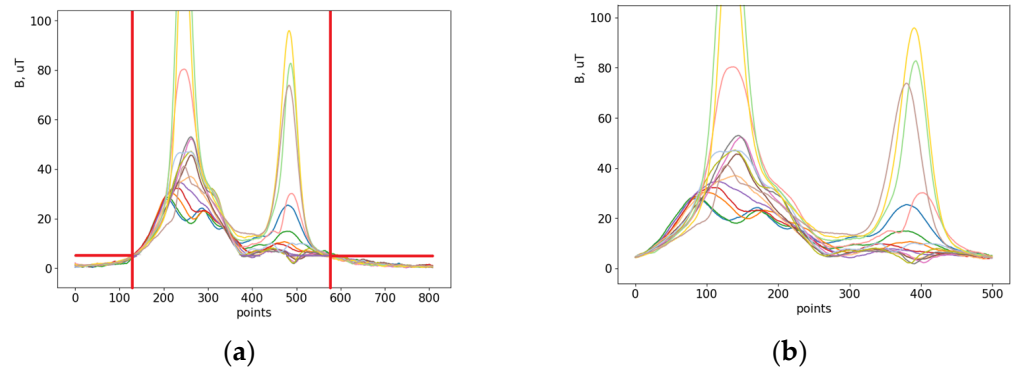


Figure 1. Registered signature cropping and resampling process. Original calculated modules (a) and cropped and resampled modules (b).

For each calculated module index, we identify points where the amplitude exceeds a pre-defined threshold. This search is performed from the beginning and the end of the signal. A new array is then formed based on these identified indices. The middle of the signal is not checked as the amplitude might drop below the threshold, but it still represents an important part of the signal.

The pipeline for the experiments is outlined in Figure 2. A record is defined as one unique vehicle passing once through the signature collection system and contains 15 $x/y/z$ signal arrays of registered magnetic distortions. The collected database contains numerous different vehicles with multiple records. Information about vehicle make, model and driving trajectory is taken from the registered image. For experimentation, we select subsets from the database with the required number of vehicles n , each containing m passing entries from passing through the system. Constraints are added so that no vehicles of the same model should exist, and driving direction must be the same. One subset is selected for training samples and another subset for testing. Record comparison, pair generation, feature extraction, model training and evaluation are performed in parallel, applying the same actions.

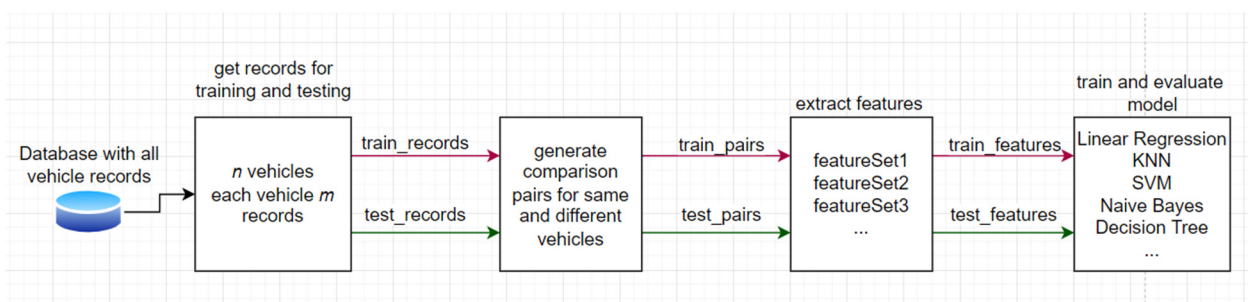


Figure 2. Pipeline for experiments.

Record pairs labeled “same” are defined as pairs of records that refer to the signatures from the same unique vehicle, while “different” records refer to pairs of records that refer to signatures from different vehicles. In our experiment, we used these definitions to evaluate the performance of our algorithm in distinguishing between matching (same) and non-matching (different) records. An example of the records is shown in Figure 3.

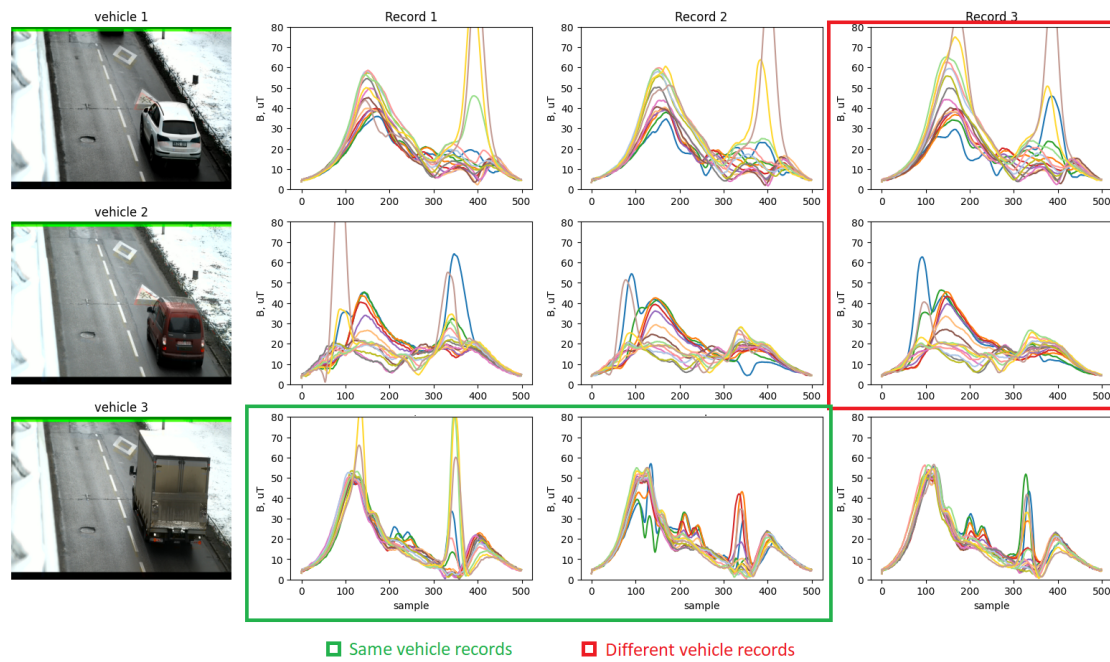


Figure 3. Explanation of same and different vehicle records with one “same” and one “different” pair marked.

The comparison pair generation algorithm (Algorithm 1) involves two steps—generating pairs for the same vehicles and, after that, pairs for different vehicles. Empty lists are initialized to store references to records. Each vehicle is iterated, and all record pairs are added to the list. Same records are not compared to themselves; in this way, the maximum number of “same” labels are created. In order to have a balanced dataset, an equal number of different record pairs should exist. For different record comparison pairs, two vehicles are randomly selected and checked to see whether they are different, and then a random record is selected for each vehicle. This process is repeated until different vehicle comparison pairs have an equal number to the same vehicle comparison pairs. The condition exists that randomly selected records cannot repeat.

Algorithm 1. Generate Comparison Pairs

```

1: Input: Database of  $n$  vehicles, each with  $m$  records
2: Output: Comparison pairs for each vehicle and between different vehicles
3: Initialize empty list sameVehiclePairs
4: Initialize empty list differentVehiclePairs
                                     >Generate same-vehicle comparison pairs
5: for each vehicle  $v \in \{1, 2, \dots, n\}$  do
6:     for each record  $r_i \in \{1, 2, \dots, m - 1\}$  do
7:         for each subsequent record  $r_j \in \{r_i + 1, r_i + 2, \dots, m\}$  do
8:             Add  $(v, r_i, r_j)$  to sameVehiclePairs
9:         end for
10:    end for
11: end for
                                     >Generate different-vehicle comparison pairs
12: countSamePairs  $\leftarrow$  length of sameVehiclePairs
13: while length of differentVehiclePairs  $\neq$  countSamePairs do
14:     Randomly select two different vehicles  $v_1$  and  $v_2$  from  $\{1, 2, \dots, n\}$ 
15:     if  $v_1 \neq v_2$  then
16:         Randomly select one record  $r_i$  from  $v_1$  and one record  $r_j$  from  $v_2$ 
17:         Add  $(v_1, r_i, v_2, r_j)$  to differentVehiclePairs
18:     end if
19: end while

```

After extensive analysis of the magnetometers axis values and modules comparison, the decision was made to use the calculated signature modules processed as explained earlier. An example of the registered magnetic signature modules for the same vehicle driving through the system is shown in Figure 4. Based on the collected signal analysis, it is evident that at least one side of the vehicle’s wheels is always presented in the single passing vehicle record, as shown in figure by the red dotted line. Sensors which are affected by the wheels register a distorted signature, usually with a much higher amplitude or unpredictable shape. Data of such sensors should be omitted from direct signal comparison. An example of different vehicles’ signatures is presented in Figure 5. It is evident that the signatures’ shape and amplitudes are significantly different between these two records.

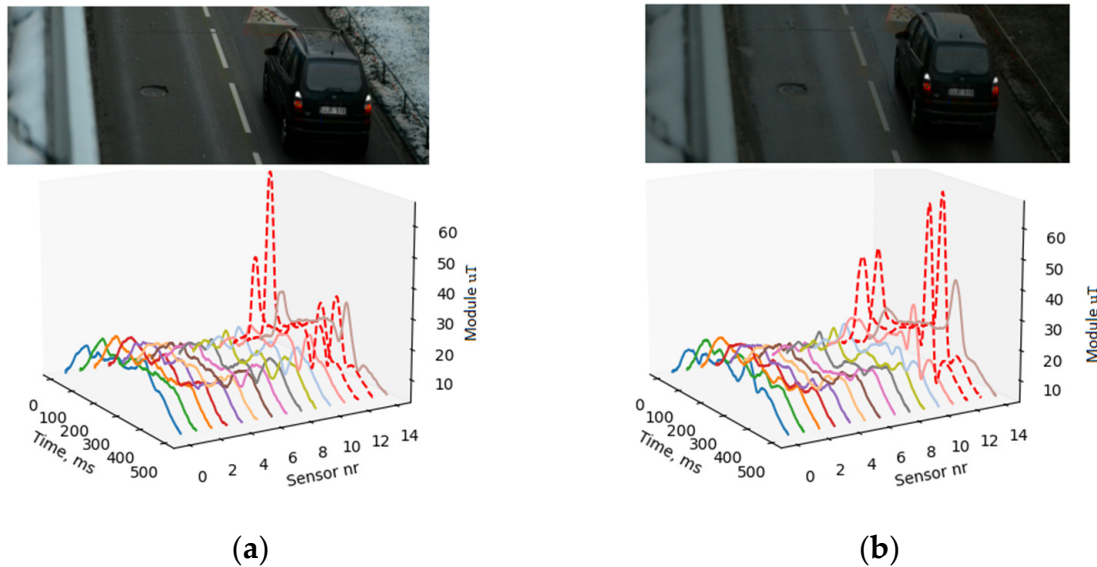


Figure 4. Two records (a,b) of the same vehicle driving through the registration system almost in the same trajectory. The red dotted line marks the wheel influence in the recorded signature.

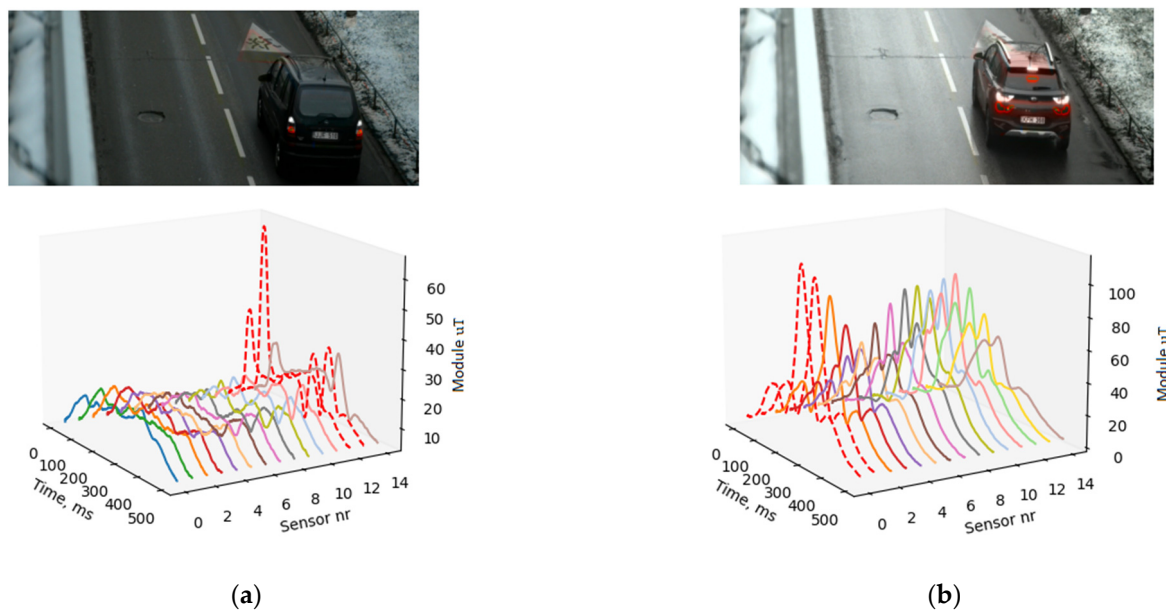


Figure 5. Record 1 registered signatures (a) and record 2 registered signatures (b) for different vehicles.

Vehicle re-identification is performed by searching for similarities between the registered magnetic signatures modules. The challenge lies in discovering how to effectively

evaluate these similarities. Different feature extraction ideas were tested. Every vehicle has a unique signature, although certain common attributes which are repeating between different vehicles can be observed. The shape of the module initially is always rising and reaches the maximum signal value or intermediate maximum. These ascents and descents can look like a parabola or an inverted parabola. Different vehicles have different numbers of intermediate maximum values in different locations. These amplitude values and location indices could be used as feature values. Signature shape approximation and single sensor values' relations to neighboring sensors' relations were investigated. These strategies were evaluated using stem plot diagrams but yielded unsatisfactory results that are not presented in this study. For vehicle re-identification, the following feature vectors were composed which showed less overlapping between same and different vehicle records:

1. Module amplitude maximum location differences;
2. Five maximum corr-coefficients from correlation matrix;
3. Ten corr-coefficients from best diagonal;
4. Five corr-coefficients calculated using vehicle center sensor;
5. Ten corr-coefficients from best diagonal and ten Euclidean distance values.

All feature vectors were calculated with normalized signature modules. Normalization was performed record-wise by taking all signatures of one record, finding the maximum amplitude and dividing all signatures by this value. This is the most reasonable approach because, if normalization were performed for every sensor separately, important information about amplitude differences between neighboring sensors would be lost. Without normalization, absolute amplitude difference information is presented which can better distinguish the same vehicles; however, in this case, the dynamic range can be high and unpredictable. To tackle this problem, it is feasible to perform value constraint in the feature extraction step so the machine learning algorithm obtains values defined in a specific range.

Feature vector 1 consists of 15 differences between module maximum indices. The idea is that same vehicle records will have smaller distance values compared to different vehicles. The vector is calculated using the following formula:

$$mind_k = loc(\max(sig1_k)) - loc(\max(sig2_k)) \quad (2)$$

where $mind_k$ is the calculated distance value between indices for sensor k ; $sig1$ corresponds to the signature module for record 1 and $sig2$ for record 2; function $\max()$ returns the maximum value of the array; function $loc()$ returns the array index.

Other feature vectors are calculated using correlation and Euclidean distance matrices. The matrices result in a 15×15 array while calculating values between all sensor pairs according to the following formula:

$$\rho_{XY}(k, j) = \frac{cov(X_k, Y_j)}{\sigma_{X_k} \sigma_{Y_j}} \quad (3)$$

where $cov(X_k, Y_j)$ is the covariance between two time series—it measures the degree to which two variables change together—and $\sigma_{X_k} \sigma_{Y_j}$ are standard deviations of the time series for every sensor $k \in \{1, 2, \dots, 15\}$, $j \in \{1, 2, \dots, 15\}$. Analogically, the Euclidean distance matrix is calculated using the following formula:

$$d_{XY}(k, j) = \sqrt{(X_k - Y_j)^2} \quad (4)$$

where d is the Euclidean distance value, and X_k , Y_j are signature modules for every sensor. An example of calculated matrices for same and different vehicles is shown in Figure 6.

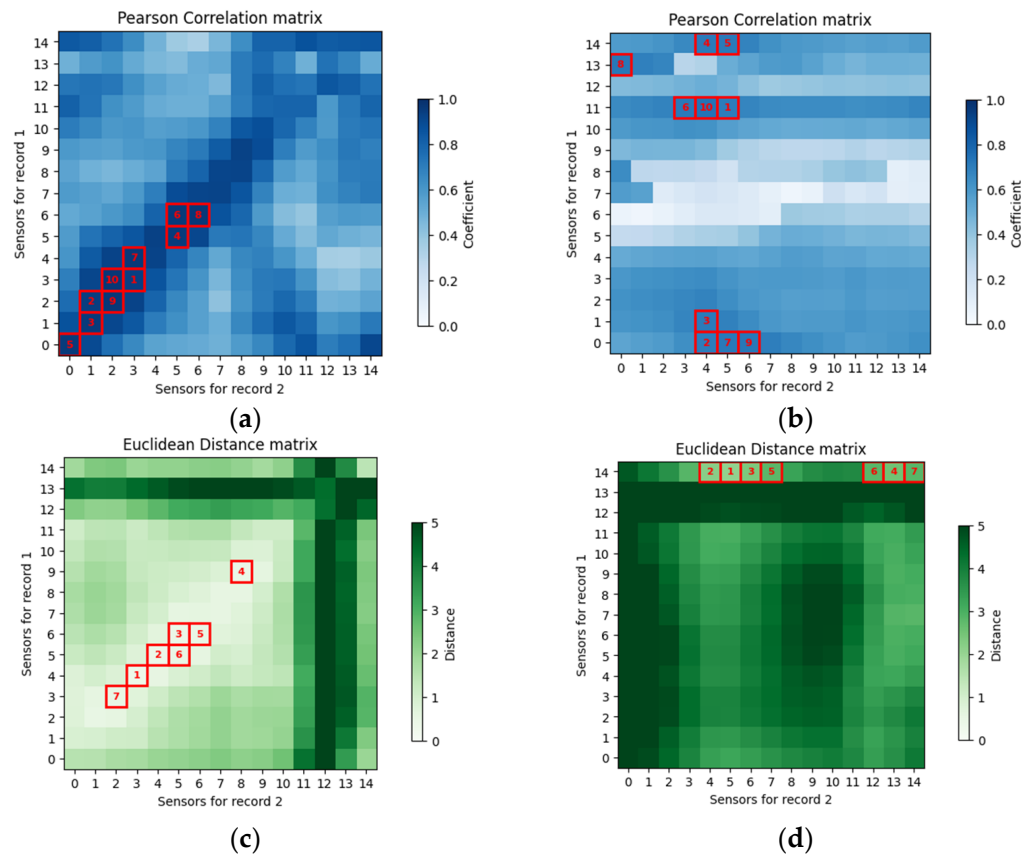


Figure 6. Matrices calculated from two records. Correlation coefficient matrix for same vehicle records (a) and different vehicle records (b). The maximum 10 values are marked in red boxes. Euclidean distance for same vehicle records (c), Euclidean distance for different vehicle records (d). The minimum 10 values are marked in red boxes.

The assumption is made that same vehicle records will have higher correlation coefficients and smaller Euclidean distance values compared to different vehicles’ records. Feature vector 2 is based on this assumption and uses the five highest correlation matrix values as a feature. Additionally, as can be seen from Figure 6, for same vehicles, it is possible to distinguish a single diagonal for the largest correlation coefficients and smallest Euclidean distances just by looking at the first 10 maximum (minimum for Euclidean distance) values. In an ideal driving trajectory, when the same vehicle ideally passes sensors, this diagonal should appear between the same sensor pairs—(0, 0), (1, 1), (2, 2) and so on. However, depending on the vehicle trajectory, this diagonal could be shifted. For feature vector 3, ten correlation coefficients taken from the diagonal are used, so the diagonal can be offset by -5 to $+5$ indices from the center position. The algorithm for best diagonal selection (Algorithm 2) is as follows:

The algorithm identifies the most significant diagonal in a given matrix by evaluating the average values of diagonals shifted within a range from -5 to $+5$. For each possible shift, the resulting diagonal average is calculated. For correlation values, the diagonal with the highest average is selected, and, for Euclidean distance values, the diagonal with the lowest average is selected. The resulting diagonal is an array with 10 values extracted from the matrix. Feature vector 5 is similar; only values are taken from both matrices (correlation and Euclidean).

Feature vector 4 focuses around center sensor and two neighboring sensors from both sides. As explained in our previous work [25], it is complicated to decide which sensor is in the middle of the vehicle. The purpose of focusing on the center sensor is to mitigate the influence of the vehicle wheels on the magnetic signature, as described earlier. If the center sensor is correctly labeled, there is no need to calculate the matrix for all sensor pairs,

but only for corresponding sensors. This simplifies calculations and directly returns five element feature vectors. Records used for analysis were manually inspected, and the center sensor position was labeled. An example of two different vehicles' signatures revolving around the center sensor and without wheel influence is shown in Figure 7. Vehicle 1 is an Opel hatchback, and vehicle 2 is a VW panel van; however, the processed and resampled signature modules visually are similar because of the first rising edge (samples 0–130). A similar signature shape raises the correlation coefficients, which are around 0.94. Usually this high a correlation value suggests that the compared records originate from the same vehicles. Although the signature shape follows a similar pattern, the amplitude differs, and the Euclidean distance matrix can help separate it into different vehicles because maximum values are not concentrated on the diagonal but are chaotic. Based on this observation, feature vector 5 was composed containing ten maximum correlation values and ten minimum Euclidean distance values extracted from the best diagonal.

Algorithm 2. Find Best Diagonal in Matrix

```

1: procedure FindBestDiagonal(matrix)
2:   avg arr  $\leftarrow$  []
3:   for shift in [-5, 5] do
4:     diagonal  $\leftarrow$  GetDiagonal(matrix, shift)
5:     avg  $\leftarrow$  Mean(diagonal)
6:     avg arr.append(avg)
7:   end for
8:   ind  $\leftarrow$  ARGMAX(avg_arr)
9:   best_shift  $\leftarrow$  ind - 5
10:  best diagonal  $\leftarrow$  GetDiagonal(matrix, best shift)
11:  return best diagonal
12: end procedure
13:
14: procedure GetDiagonal(matrix, shift)
15:  diagonal  $\leftarrow$  []
16:  for i in [0, 15 - |shift|] do
17:    if shift > 0 then
18:      diagonal.append(matrix[i, i + shift])
19:    else
20:      diagonal.append(matrix[i + |shift|, i])
21:    end if
22:  end for
23:  return diagonal
24: end procedure

```

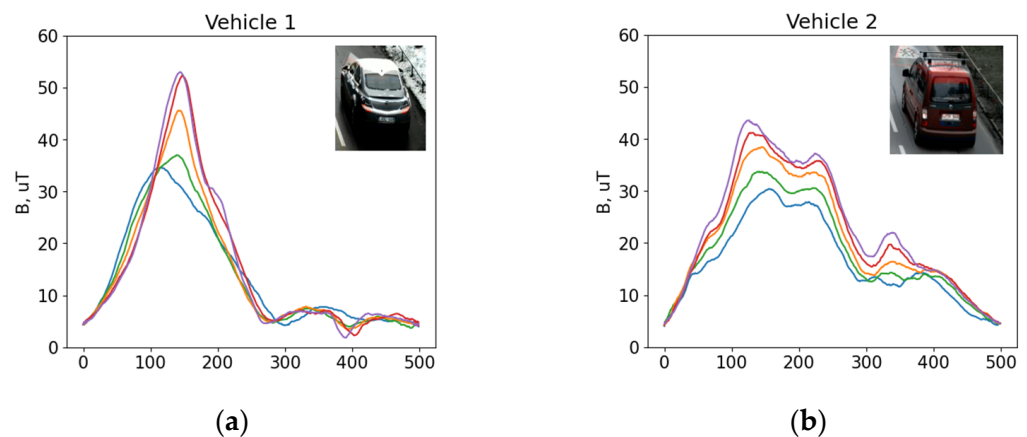


Figure 7. Two different vehicles' signatures around center sensor—vehicle 1 (a); vehicle 2 (b).

During the experiments, different methods for feature extraction were analyzed, and it was noticeable that the Pearson correlation coefficient and Euclidean distance emerged as the optimal parameters based on their computational efficiency, which allowed for the accurate comparison of records while minimizing processing time. Feature sets are composed of calculated feature vectors for generated comparison pairs.

In the results section, there is a comparison of accuracy, and calculation time is evaluated for the presented feature sets. Accuracy is evaluated by calculating the confusion matrix and accuracy metrics.

3. Results

The used dataset consists of 300 different vehicles passing the sensor in the same direction three times in real traffic conditions. Comparison pairs were generated, as explained earlier, resulting in 900 entries for same vehicle record pairs and 900 entries for different vehicles. Of all the pairs, 70% were used for training and 30% for testing. A feature boxplot of correlation coefficient features using the diagonal with the preselected center sensor (feature vector 4) is presented in Figure 8. For different vehicle records, correlation values had a wide range but were concentrated below 0.95. For the same vehicle records, the median was above 0.97; however, outliers existed. After inspecting the comparison pairs for the same vehicles with low correlation values, it was noted that incorrect center sensor selection generates more outliers in same vehicle comparison pair features. This suggests that, in order to reduce outliers, it is necessary to perform relabeling for the selected center sensor, or use a magnetic sensor array with smaller spacing or a different similarity evaluation method which is not affected by wheel-distorted signatures.

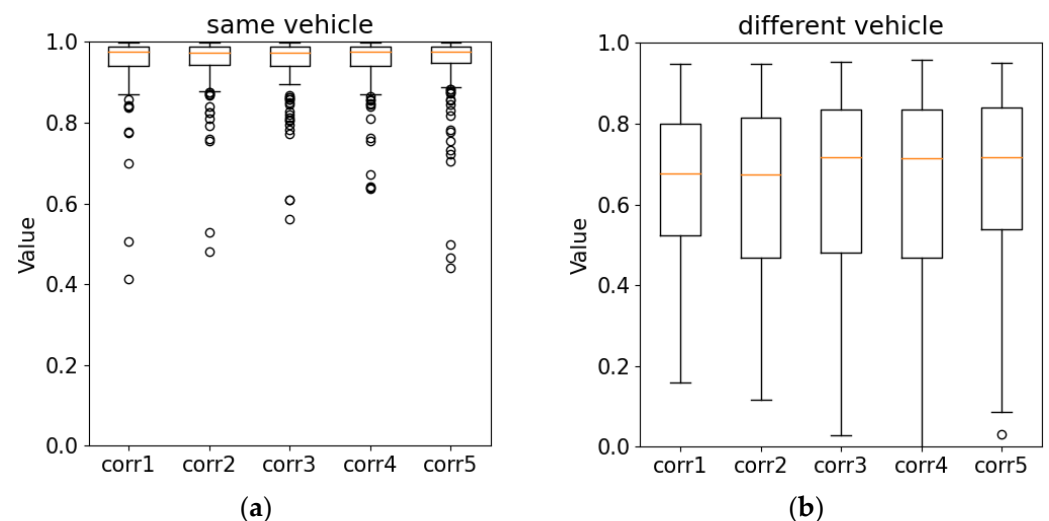


Figure 8. Box diagram showing feature set of five maximum correlations' distribution for same vehicle records (a) and different vehicle records (b). Round circles indicate outliers, orange line indicate median.

Seven different machine learning models for the classification problem were tested using the earlier-defined input feature sets. Linear regression was the simplest model. After training, model weights were calculated, which were used on input data, and summed together, and the activation function was applied. After training the model with simple feature sets (one value for one sensor pair) for correlation and amplitude differences and sorting calculated weights by importance, no significant sensor was noted—every time, a different sensor had the highest weight.

The nearest neighbor algorithm complexity is determined by value k . A higher value can give more accurate classification; however, the model might become overfitted and poorly generalize new unseen data. Value $k = 3$ was empirically selected after exploring small training/testing subsets. The KNN algorithm computes distances between input

features and trained samples and, based on majority vote, finds the closest neighbor sample. Intuitively, features for the same vehicle are closer together than for different vehicles.

The support vector machine model creates a hyperplane separating input features into two categories for same and different vehicles. Although the model is more aimed at high-dimensional feature space where the feature count exceeds sample count, it worked correctly in the vehicle re-identification task. Different kernel functions can be used to capture complex feature relationships. A linear kernel was selected.

The Naïve Bayes classifier works with probabilities based on Bayes' Theorem. The Gaussian Naïve Bayes classifier was used. The model was fitted by finding the mean and standard deviation of the two vehicle classes' input features.

A decision tree aims to create a tree-like model for decisions based on input features. Values are best split into homogeneous subsets with respect to the target variable. These models are transparent and intuitive, allowing us to understand and explore relationships within the data. By constructing multiple instances of decision trees during training time, a Random Forest classifier is achieved.

The Gradient Boosting Machine (GBM) is a powerful ensemble machine learning technique suitable for classification tasks. The boosting method converts weak learners into strong learners.

Models were compared using accuracy, precision and recall metrics, which are summarized in Table 1. It is notable that different methods yielded varying levels of accuracy across different feature sets. For instance, the Random Forest and Gradient Boosting Machine consistently demonstrated higher accuracy across most features compared to other methods. In this re-identification task, it was important to capture all positive indices (recall)—vehicle records which correspond to the same vehicle. While some methods exhibited high precision but lower recall (e.g., Naïve Bayes), others achieved a balance between precision and recall (e.g., Gradient Boosting Machine). The performance of each method varied across different features, indicating the importance of feature selection.

Table 1. Used features and machine learning method evaluation.

		Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Linear Regression	Accuracy	0.817	0.856	0.908	0.869	0.972
	Precision	0.774	0.776	0.845	0.812	0.983
	Recall	0.894	1.0	1.00	0.961	0.961
KNN	Accuracy	0.844	0.942	0.972	0.903	0.978
	Precision	0.888	0.904	0.947	0.876	0.989
	Recall	0.789	0.989	1.00	0.939	0.967
SVM	Accuracy	0.822	0.872	0.933	0.900	0.953
	Precision	0.776	0.796	0.882	0.860	0.96
	Recall	0.906	1.00	1.00	0.956	0.944
Naïve Bayes	Accuracy	0.819	0.947	0.958	0.869	0.969
	Precision	0.767	0.917	0.923	0.812	0.988
	Recall	0.917	0.983	1.00	0.961	0.950
Decision Tree	Accuracy	0.769	0.933	0.897	0.889	0.972
	Precision	0.765	0.882	0.839	0.857	0.967
	Recall	0.778	1.00	0.983	0.933	0.978
Random Forest	Accuracy	0.897	0.942	0.953	0.911	0.986
	Precision	0.874	0.896	0.914	0.889	0.973
	Recall	0.928	1.00	1.00	0.939	1.00
Gradient Boosting Machine	Accuracy	0.883	0.936	0.928	0.908	0.978
	Precision	0.856	0.887	0.874	0.889	0.967
	Recall	0.922	1.00	1.00	0.933	0.989

Different machine learning models were compared using a Receiver Operating Characteristic (ROC) curve. It illustrates the performance of a binary classification model across

different discrimination thresholds. The true-positive rate is plotted against false-positive rates at various thresholds. ROC curves for different analyzed feature sets are shown in Figure 9. Feature sets that only include correlation coefficients or amplitude evaluations tend to perform worse compared to feature sets that combine both types of features. This highlights the benefit of mixing signature amplitude and shape evaluation. These findings underscore the significance of feature engineering in signature re-identification tasks, emphasizing the critical role played by feature selection and combination strategies.

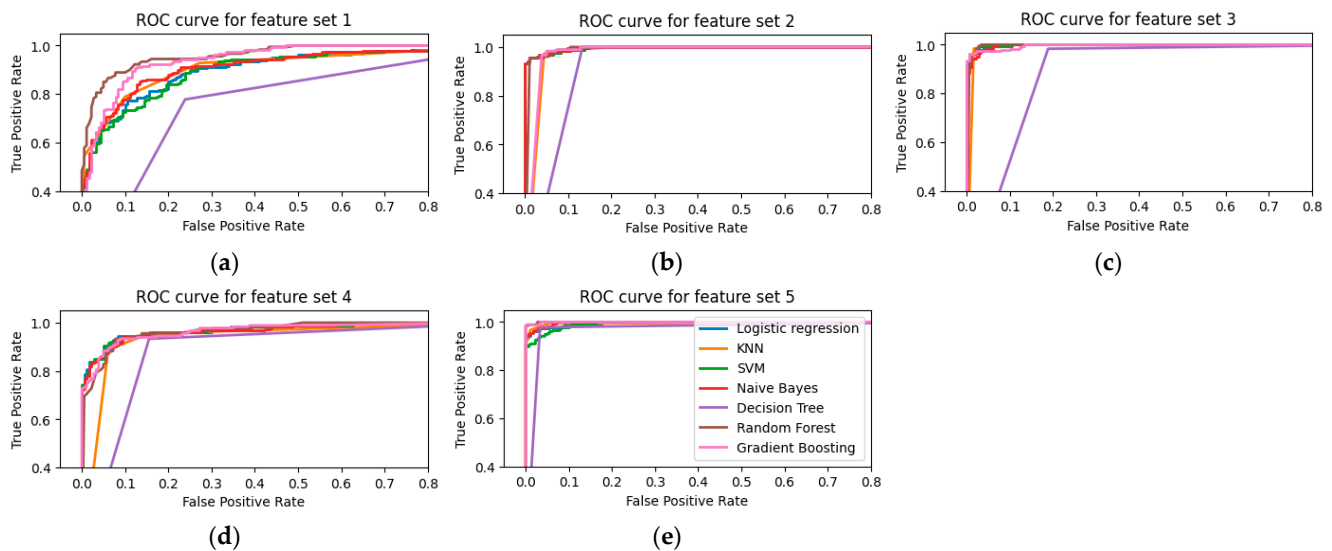


Figure 9. ROC curves for comparing different machine learning methods with different feature sets. (a) feature set1, (b) feature set2, (c) feature set3, (d) feature set4, (e) feature set5.

4. Discussion

In this paper, a novel methodology for feature extraction and signature comparison using a magnetometer array is presented. Raw registered signals were preprocessed and different feature sets extracted.

Using only correlation features and ignoring amplitudes gives bad re-identification accuracy because different vehicle signatures can have a similar shape. Alternatively, when using only amplitude difference, same vehicle records might appear as different because registered amplitude values fluctuate based on driving pattern.

Combining correlation coefficients with amplitudes for specific sensor pairs gives the best re-identification accuracy, which reaches 98% with the Random Forest algorithm.

Choosing a machine learning algorithm is not critical. With a good feature set, all tested algorithms perform well, with accuracy reaching 97%, while, with a bad feature set, the accuracy gap is more visible—achieving accuracy values of 76–93%. The best accuracy is reached with the Random Forest algorithm and KNN. In all cases, the traditional decision tree’s performance is the worst.

The proposed method does not require the intensive computation associated with deep learning models, making it more suitable for real-time applications and deployment in resource-constrained environments. Unlike deep-learning-based methods, our approach ensures anonymity by not relying on image or video data, which often contain personally identifiable information.

5. Conclusions

Our findings can be summarized in the following conclusions:

1. Using only correlation coefficients or amplitude features for specific sensor pairs, the re-identification accuracy is about 93%.
2. Feature selection has a higher impact than the machine learning method used.

3. Same vehicle identification accuracy using correlation coefficients combined with Euclidean distances can reach 97%.

Author Contributions: Conceptualization: D.N., J.B. and D.A.; Methodology: V.M., A.V. and M.Z.; Software: J.B. and M.C.; Validation: D.A., A.V., M.Z., M.P., J.K., Z.L. and M.F.; Visualization: Z.L., J.B., M.F. and M.C.; Investigation: V.M., D.N., J.B., A.V., J.K., M.P. and M.Z.; Resources: D.N., M.Z., M.P., M.F. and M.C.; Data Curation: D.N., V.M., Z.L. and J.K.; Writing—Original Draft Preparation: D.N., J.B., D.A. and M.C.; Writing—Review and Editing: J.B., D.N., D.A. and A.V.; Supervision: D.N., V.M. and D.A.; Project Administration: D.A. and M.Z.; Funding Acquisition: D.A., A.V. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from the Research Council of Lithuania (LMTLT), agreement no. S-A-UEI-23-1 (22 December 2023).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bikku, T.; Narayana, V.; Arepalli, G.; Khadherbhi, S. Sensors Systems for Traffic Congestion Reduction Methodologies. In Proceedings of the 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 12–14 December 2019; pp. 452–455. [\[CrossRef\]](#)
2. Spandonidis, C.; Giannopoulos, F.; Elias, S.; Reppas, D.; Theodoropoulos, P. Development of a MEMS-Based IoV System for Augmenting Road Traffic Survey. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–8. [\[CrossRef\]](#)
3. Yue, W.; Li, C.; Mao, G.; Cheng, N.; Zhou, D. Evolution of road traffic congestion control: A survey from erspective of sensing, communication, and computation. *China Commun.* **2021**, *18*, 151–177. [\[CrossRef\]](#)
4. Komolovaite, D.; Krisciunas, A.; Lagzdinyte-Budnike, I.; Budnikas, A.; Rentelis, D. Vehicle Make Detection Using the Transfer Learning Approach. *Elektron. Elektrotechnika* **2022**, *28*, 55–64. [\[CrossRef\]](#)
5. Surgailis, T.; Valinevicius, A.; Markevicius, V.; Navikas, D.; Andriukaitis, D. Avoiding Forward Car Collision using Stereo Vision System. *Elektron. Elektrotechnika* **2012**, *18*, 37–40. [\[CrossRef\]](#)
6. Hodoň, M.; Karpis, O.; Sevcík, P.; Kociánová, A. Which Digital-Output MEMS Magnetometer Meets the Requirements of Modern Road Traffic Survey? *Sensors* **2021**, *21*, 266. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Alsayfi, M.S.; Dahab, M.Y.; Eassa, F.E.; Salama, R.; Haridii, S.; Al-Ghamdi, A.S. Big Data in Vehicular Cloud Computing: Review, Taxonomy, and Security Challenges. *Elektron. Elektrotechnika* **2022**, *28*, 59–71. [\[CrossRef\]](#)
8. Sanchez, R.O.; Christopher, F.; Rajagopal, R.; Varaiya, P. Arterial travel time estimation based on vehicle re-identification using magnetic sensors: Performance analysis. In Proceedings of the Conference Record-IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 5–7 October 2011; Volume 10, pp. 997–1002. [\[CrossRef\]](#)
9. Balid, W.; Tafish, H.; Refai, H. Versatile real-time traffic monitoring system using wireless smart sensors networks. In Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Doha, Qatar, 3–6 April 2016; Volume 201, pp. 1–6. [\[CrossRef\]](#)
10. Miklusis, D.; Markevicius, V.; Navikas, D.; Cepenas, M.; Valinevicius, A.; Zilys, M.; Cuinas, I.; Klimenta, D.; Andriukaitis, D. Research of Distorted Vehicle Magnetic Signatures Recognitions, for Length Estimation in Real Traffic Conditions. *Sensors* **2021**, *21*, 7872. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Chen, X.; Kong, X.; Xu, M.; Sandrasegaran, K.; Zheng, J. Road Vehicle Detection and Classification Using Magnetic Field Measurement. *IEEE Access* **2019**, *7*, 52622–52633. [\[CrossRef\]](#)
12. Arab, M.; Nadeem, T. MagnoPark-Locating On-Street Parking Spaces Using Magnetometer-Based Pedestrians' Smartphones. In Proceedings of the 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), San Diego, CA, USA, 12–14 June 2017; pp. 1–9. [\[CrossRef\]](#)
13. Bakirci, M. Simulation of Autonomous Driving for a Line-Following Robotic Vehicle: Determining the Optimal Manoeuvring Mode. *Elektron. Elektrotechnika* **2023**, *29*, 4–11. [\[CrossRef\]](#)
14. Zhang, Z.; He, X.; Yuan, H. An Antiinterference Traffic Speed Estimation System With Wireless Magnetic Sensor Network. *IEEE Trans. Ind. Inform.* **2020**, *16*, 2458–2468. [\[CrossRef\]](#)
15. Ou, J.; Qiu, J.; Dong, X.; Wang, Z.; Du, J.; Chang, Q. Research on the Moving Magnetic Object Recognition Method Based on Magnetic Signature Waveform. *IEEE Trans. Magn.* **2022**, *58*, 1–8. [\[CrossRef\]](#)
16. Wang, Z.; Zheng, E.; Liu, J.; Guo, T. Adaptive Orthogonal Basis Function Detection Method for Unknown Magnetic Target Motion State. *Appl. Sci.* **2024**, *14*, 902. [\[CrossRef\]](#)
17. Prateek, G.V.; Nijil, K.; Hari, K.V.S. Classification of Vehicles Using Magnetic Field Angle Model. In Proceedings of the 2013 4th International Conference on Intelligent Systems, Modelling and Simulation, Bangkok, Thailand, 29–31 January 2013; pp. 214–219. [\[CrossRef\]](#)

18. Tawfik, A.Y.; Abdulhai, B.; Peng, A.; Tabib, S.M. Using Decision Trees to Improve the Accuracy of Vehicle Signature Reidentification. *Transp. Res. Rec.* **2004**, *1886*, 24–33. [[CrossRef](#)]
19. Dong, H.; Wang, X.; Zhang, C.; He, R.; Jia, L.; Qin, Y. Improved Robust Vehicle Detection and Identification Based on Single Magnetic Sensor. *IEEE Access* **2018**, *6*, 5247–5255. [[CrossRef](#)]
20. Sarcevic, P.; Pletl, S.; Odry, A. Real-Time Vehicle Classification System Using a Single Magnetometer. *Sensors* **2022**, *22*, 9299. [[CrossRef](#)]
21. Xu, C.; Wang, Y.; Bao, X.; Li, F. Vehicle Classification Using an Imbalanced Dataset Based on a Single Magnetic Sensor. *Sensors* **2018**, *18*, 1690. [[CrossRef](#)] [[PubMed](#)]
22. Balid, W.; Tafish, H.; Refai, H.H. Intelligent Vehicle Counting and Classification Sensor for Real-Time Traffic Surveillance. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1784–1794. [[CrossRef](#)]
23. Zhou, X. Vehicle location estimation based on a magnetic sensor array. In Proceedings of the 2013 IEEE Sensors Applications Symposium Proceedings, Galveston, TX, USA, 19–21 February 2013; pp. 80–83. [[CrossRef](#)]
24. Kwong, K.; Kavalier, R.; Rajagopal, R.; Varaiya, P. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transp. Res. Part C Emerg. Technol.* **2009**, *17*, 586–606. [[CrossRef](#)]
25. Pitton, A.C.; Vassilev, A.; Charbonnier, S. Vehicle Re-Identification with Several Magnetic Sensors. In *Advanced Microsystems for Automotive Applications 2012: Smart Systems for Safe, Sustainable and Networked Vehicles*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 281–290.
26. Sanchez, R.O.; Flores, C.; Horowitz, R.; Rajagopal, R.; Varaiya, P. Vehicle re-identification using wireless magnetic sensors: Algorithm revision, modifications and performance analysis. In Proceedings of the 2011 IEEE International Conference on Vehicular Electronics and Safety, Beijing, China, 10–12 July 2011; pp. 226–231. [[CrossRef](#)]
27. Amodio, A.; Ermidoro, M.; Savaresi, S.M.; Previdi, F. Automatic Vehicle Model Recognition and Lateral Position Estimation Based on Magnetic Sensors. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2775–2785. [[CrossRef](#)]
28. Čulík, K.; Štefancová, V.; Hrudkay, K. Application of Wireless Magnetic Sensors in the Urban Environment and Their Accuracy Verification. *Sensors* **2023**, *23*, 5740. [[CrossRef](#)]
29. Balamutas, J.; Navikas, D.; Markevicius, V.; Cepenias, M.; Valinevicius, A.; Zilyus, M.; Frivaldsky, M.; Li, Z.; Andriukaitis, D. Passing Vehicle Road Occupancy Detection Using the Magnetic Sensor Array. *IEEE Access* **2023**, *11*, 50984–50993. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.