UnScenE: Toward Unsupervised Scenario Extraction for Automated Driving Systems from Urban Naturalistic Road Traffic Data

Nico Weber\textsuperscript{1,2}, Christoph Thiem\textsuperscript{1}, and Ulrich Konigorski\textsuperscript{2}

Abstract—Scenario-based testing is a promising approach to solve the challenge of proving the safe behavior of vehicles equipped with automated driving systems (ADS). Since an infinite number of concrete scenarios can theoretically occur in real-world road traffic, the extraction of relevant scenarios that are sensitive regarding the safety-related behavior of ADS-equipped vehicles is a key aspect for the successful verification and validation of these systems. Therefore, this paper provides a method for extracting multimodal urban traffic scenarios from naturalistic road traffic data in an unsupervised manner for minimizing the amount of (potentially biased) prior expert knowledge needed. Rather than an (expensive) rule-based assignment by extracting concrete scenarios into predefined functional scenarios, the presented method deploys an unsupervised machine learning pipeline. It includes principal feature analysis, feature extraction with so-called scenario grids, dimensionality reduction by principal component analysis, scenario clustering as well as cluster validation. The approach allows exploring the unknown natures of the data and interpreting them as scenarios that experts could not have anticipated. The method is demonstrated and evaluated for naturalistic road traffic data at urban intersections from the inD and the Silicon Valley dataset. The findings encourage the use of this type of data as well as unsupervised machine learning approaches as important pillars for a systematic construction of a relevant scenario database with sufficient coverage for testing ADS.

I. INTRODUCTION

The maturity of technical implementations of automated driving systems (ADS)\cite{1} and resulting fields of application are continuously increasing. While ADS-equipped vehicles promise to contribute significantly to a safer, more efficient and more comfortable future mobility\cite{2}, the greatest challenge for a market launch of such systems arises from the need for prior proof of safety of the intended functionality\cite{3} for future operation in real-world road traffic\cite{4,5}. Regarding to an urban operational design domain (ODD), possible increases in traffic safety are particularly relevant, as nearly 70\% of accidents involving personal injury in Germany do occur in urban areas\cite{6} (data collection period: 2016-2020).

Since existing safety validation approaches would require billions of test kilometers under representative conditions before market launch\cite{5}, new methods for the release of ADS-equipped vehicles are currently under development. One of these methods is the scenario-based development and test approach, as proposed by project PEGASUS\cite{7}. Following the paradigm of this approach, the majority of conventional driving kilometers is not challenging enough and thus a reduction of the test effort is to be expected when testing exclusively relevant scenarios\cite{8}. In recent years, the focus of research into safety validation approaches for ADS has primarily been set on highway applications (e.g., a Highway Chauffeur)\cite{7}. With the approval of the first Level 3 ADS for traffic jam situations within highway environments by German Federal Motor Transport Authority (KBA) by end of 2021\cite{9}, series-production ADS-equipped vehicles are expected to be introduced on public roads soon.

However, both the objective determination what relevant means with respect to scenarios for testing ADS and a commonly accepted methodology for the construction of a representative scenario database of sufficient coverage are subject of ongoing research\cite{2,10}. The basic challenge here, independent of the ODD, is the enormously high number of traffic situations that the ADS has to cope with, referred to as open-context problem. The goal of a scenario extraction method can be summarized as mapping of this infinite-dimensional open context to a finite and manageable set of scenarios\cite{11} that reflects the nature of the traffic dynamics of interest in a sufficiently valid manner for subsequent testing. Real-world data represents a valuable data source for the construction of a comprehensive scenario database\cite{2,12}. While the construction of such a database solely by expert knowledge already seems extremely challenging for the highway environments, even when applying sophisticated statistical approaches\cite{13}, it appears to be virtually impossible for urban ODDs. This can be traced back to significant changes regarding the PEGASUS six-layer model for structuring scenarios\cite{14} for the first four layers when transitioning to urban ODDs. The main reason for this is a substantially increasing variability in terms of both traffic spaces and traffic dynamics. With regard to traffic dynamics, the less rule-based behavior and multimodal interaction of various road user (RU) types is to be considered as crucial aspect.

Existing publications concerning data-driven scenario extraction can be divided in rule-based approaches, machine learning based approaches or a combination of both (e.g.,\cite{15,16,17}). Rule-based approaches have the advantage of not relying on large amounts of annotated (labeled) data. However, the complexity of the rules to be implemented and thus the required effort increases with the complexity of the reality of interest\cite{18} to be captured. On the one hand, this leads to potentially undetected known, unsafe scenarios. On the other hand, it is impossible to explore previously unknown, unsafe scenarios, which, however, are of great importance for a reliable safety argumentation.
Supervised machine learning approaches, in turn, require a sufficient amount of labeled data. As these labels are partly unknown in advance, discovering them must be an intrinsic part of a comprehensive scenario extraction method before respective supervised approaches can be deployed.

Therefore, this paper proposes a method for extracting multimodal urban traffic scenarios from naturalistic road traffic data in an unsupervised manner, minimizing the amount of prior expert knowledge required, as shown in Fig. 1. Based on knowledge gained from first rule-based investigations [19], we follow a generic clustering procedure [20] for exploring the unknown natures of the data, which are interpretable as scenarios [21]. The method utilizes principal feature analysis [22] for feature selection and proposes the use of scenario grids for a scenario representation compatible with the application of static data based clustering algorithms. We are emphasizing a scenario representation compatible with established approaches in the field of computer-vision, like convolutional neural networks, for a future deployment of the method within a semi-supervised machine learning pipeline. For the evaluation of the method, hierarchical agglomerative clustering (HAC) is applied after reducing the dimensionality by principal component analysis. The method is evaluated using the inD dataset [23] and the Silicon Valley Intersections dataset [24].

The main contributions of this paper are as follows: provision of an overview of existing time series clustering methods and data-driven scenario extraction approaches (Sec. II), description of a method for Unsupervised Scenario Extraction (UnScenE) from naturalistic road traffic data (Sec. III), exemplary implementation and evaluation of the scenario extraction method (Sec. IV), and, finally, an elaboration on the results showing future research directions (Sec. V).

II. BACKGROUND

A. Requirements Specification

The specification of requirements for the scenario extraction method should be accompanied by a prioritization of them in order to develop a common understanding and to achieve a focus on the most important aspects of the development task at hand. We utilize the MoSCoW prioritization, where each requirement is marked as must (M), should (S), could (C) or wont (W) according to its importance. Note, that requirements marked as W are potentially as important as the ones marked with M, but can be left for a future release of the implementation [25]. The following requirements for the scenario extraction method were identified:

- The scenario extraction method must allow a clustering of multimodal urban traffic scenarios of varying number of road users (M2).
- The scenario extraction method must be applicable to different urban traffic spaces with as little adaptation effort as possible (M3).
- The scenario extraction method must include a feature representation that is compatible with supervised machine learning approaches (M4).
- The scenario extraction method must allow a slicing of time series data in meaningful sub-sequences (M5).
- The scenario extraction method should have a feature representation that includes multiple road user dynamics simultaneously at the scenario level (S1).
- The scenario extraction method should minimize the amount of prior expert knowledge required for the application (S2).
- The scenario extraction method is intended to enable scenario clustering without including road network information (C1).
- The scenario extraction method is intended to enable scenario clustering based on different data sources in future releases (W1).

B. Clustering of Time Series Data

The major strategies for clustering time series data can be divided into raw-data based, feature-based, and model-based approaches [26]. Raw-data based (direct) approaches typically compare different time series using established clustering algorithms by replacing the distance or similarity measure for static data with an appropriate one for time series, e.g., dynamic time warping [20].
Model-based approaches assume that the time series under investigation are based on an underlying model or that they are determined by a combination of probability distributions and are evaluated by means of a similarity measure between fitted models [27]. Both, model-based and feature-based approaches can be categorized as indirect approaches, as they first convert raw time series data into a feature vector of lower dimension and perform the clustering on the model parameters or the feature vector, respectively, rather than on the raw time series data themselves [26]. Feature-based (representation-based [28]) approaches can be further divided into strategies deploying manual or automated feature extraction (Representation Learning [29]).

Despite the enormous number of existing approaches within the major strategies described, we are not aware of any implementation that meets all of our requirements specified. This is mainly caused by three characteristics within the nature of our problem at hand: First, the scenario extraction method must be able to deal with time series of different lengths, as different road users have different velocities while passing through the observation area (cf. Req. M1). Second, the method has to cope with multivariate time series, since a relevant scenario is composed of at least two road users potentially involving multiple time series per road user for a meaningful scenario representation (cf. Req. M2). Third, the extraction method has to cope with a varying amount of time series to be considered, because semantically similar scenarios can include different numbers of road users (cf. Req. M2 and S1). Therefore, we deployed a novel approach for scenario representation that is tailored to the problem at hand allowing the use of established clustering algorithms for static data.

C. Data-Driven Scenario Extraction

Most of the approaches dealing with data-driven scenario extraction are developed for highway use cases (e.g., [4], [16], [30], [31], and [32]). Elpsas et al. [17] propose a method, where rule-based detections are used to train fully convolutional networks for extracting lane change and cut-in trajectories belonging to a concrete scenario, we want to compare the dynamics of multiple road users simultaneously (cf. Req. M4) and include a slicing of the time series data, since one ego-vehicle can encounter multiple concrete scenarios during its lifetime (cf. Req. M5) [37]. Furthermore, instead of a sequential and pairwise similarity estimation between trajectories belonging to a concrete scenario, we want to compare the dynamics of multiple road users simultaneously (cf. Req. S1). Summarizing, to the best of our knowledge, there is no commonly accepted method for extracting relevant multimodal urban traffic scenarios for testing ADS that meets all the requirements specified.

III. SCENARIO EXTRACTION METHOD

The UnScenE method for extracting multimodal urban traffic scenarios developed with respect to the specified requirements (cf. Sec. II) is shown in Fig. 1. It follows the generic clustering procedure according to Xu and Wunsch [20] and contains novel instantiations within the different steps to solve the task at hand. While the amount of data decreases with the passing of the different steps, the knowledge of the inherent patterns regarding the reality of interest (i.e., relevant scenarios concerning specified urban traffic spaces) increases. Since the method in the first stage of development must be able to process naturalistic road traffic data particularly, exemplary explanations of the application refer to this type of data source.

A. Spatiotemporal Filtering

Naturalistic road traffic data contain partly irrelevant parts for subsequent scenario extraction (e.g., vehicles driving straight through the observation area without other RUs present). Given a non-empty set of input patterns in form of multivariate time series data \( X = \{x_1, \ldots, x_j, \ldots, x_N\} \), where \( x_j = \{x_{j1}, x_{j2}, \ldots, x_{jd}\} \in \mathbb{R}^d \), with a \( d \)-dimensional feature space, the goal of this process step is to determine the relevant subset of samples \( j \) out of \( N \), where \( j < N \).
Hence, this process step aims at reducing noise in the data by removing the irrelevant data proportions. Within the proposed method, this is achieved by a search algorithm based on the post-encroachment time (PET) \cite{39} of all possible combinations of an ego-vehicle (ego) trajectory and those of encountered road users (challengers) while passing through the observation area (ego-lifetime). As an ego can encounter multiple concrete scenarios within its lifetime (cf. Req. M5), it has to be accounted for a slicing of the corresponding trajectories in such cases, additionally.

In detail, this step is performed by a two-layered procedure for each car included in the dataset, as our goal is to extract relevant test scenarios for ADS-equipped vehicles. In the first layer, a distance matrix is generated for each combination of ego and challenger, containing the Euclidean distances of the involved trajectories for all common time steps within the observation area. By user-adaptive parametrization of the threshold for the distance between trajectories $d_{\text{min}}$, it is possible to define how close the involved trajectories have to come to each other, so that an interaction between the involved road users is conceivable. The value of $d_{\text{min}}$ can be interpreted as radius of a moving circle surrounding the ego, whose area must overlap the challenger trajectory. Within the second layer, the minimum value of the distance matrix values that satisfy this criterion is used to calculate the PET of the trajectories involved. By setting a PET threshold $t_{\text{PET}}$, the temporal evolvement of the scenario is taken into account. If the two layers of the algorithm are passed through several times during the lifetime of an ego-vehicle, the resulting ego-challenger combinations can be used to slice the respective time series.

It should be noted that, at the current state of implementation, this algorithm is accompanied by two main assumptions: First, an intersection area (rather than an intersection point) between trajectories involved in a scenario is considered sufficient to maintain it as potentially relevant. Second, the PET between an ego-vehicle and the nearest challenger is crucial for the following classification of the concrete scenario to be kept. In case of an empty set after spatiotemporal filtering, the extraction method is aborted and, e.g., new data containing relevant scenarios must be recorded for successful application. For further discussion, the entirety of the remaining dataset is referred to as intersecting trajectories.

### B. Principal Feature Analysis (Feature Selection)

While the spatiotemporal filtering involves a reduction in the number of samples, i.e., reduces the $N$, this process step aims at reducing the number of columns, i.e., reducing the $d$. Thus, the task can be described as choosing $n$ distinguishing features out of $d$, whereas $n < d$. An elegant selection of salient features can greatly decrease the storage requirements, simplify the subsequent design process, and facilitate the understanding of the data \cite{20, 22}.

With respect to the problem at hand, an approach has to be chosen that is able to determine the importance of the features of the intersecting trajectories in order to subsequently select a meaningful subset of them. While there are various methods to reduce the dimensionality of a feature set, e.g., principal component analysis (PCA), these approaches are characterized by resulting in a lower-dimensional representation, where the features are not physically interpretable anymore \cite{20, 22}. While this is unproblematic for other tasks, these approaches are not effective in the context of subsequent manual feature extraction, where physical interpretability is of great importance.

Hence, the proposed method deploys feature selection by principal feature analysis (PFA) \cite{22}, which makes it possible to exploit the structure of the principal components of a feature set to find a subset of the original feature vector. The method includes five steps, with the first step calculating the covariance matrix of the standardized intersecting trajectories dataset $X_{\text{std}}$. This is followed by the calculation of the principal components and the eigenvalues of the covariance matrix $C$. By choosing the explained variance ratio $\text{var}_{\text{PFA}} \in [0, 1]$, $s$ columns of the matrix $A$, representing the orthonormal eigenvectors of $C$, are kept, constructing the subspace matrix $A_s$. The parametrization of $\text{var}_{\text{PFA}}$ decides how much of the variability of the data is desired to be retained. In the third step, the rows $v_i \in \mathbb{R}^s$ of $A_s$ are clustered using $k$-means algorithm. Since each vector $v_i$ represents the projection of the $i$th feature of $X_{\text{std}}$ in the lower-dimensional space, the $s$ elements of $v_i$ correspond to the weights of the $i$th feature on each axis of the subspace $s$. As the amount of mutual information increases with the similarity of the absolute values of these vectors, the clusters derived by $k$-means can be used to choose one feature of a subset of highly correlated features, respectively \cite{22}. According to \cite{22}, these features represent each group optimally by means of high spread in the lower dimension, reconstruction and insensitivity to noise. It is of note, that the number of clusters should be chosen slightly higher than $s$.

Finally, using this method it is possible to systematically reduce the intersecting trajectories dataset regarding its dimensionality. This paves the way for a well-founded extraction of features for the construction of the scenario grids, as described in the following process step.

### C. Feature Extraction

With the spatiotemporal filtered and dimensionally reduced intersecting trajectories dataset available, the following process step consists of a feature extraction suitable for the subsequent clustering of multimodal urban traffic scenarios. Ideally, such a feature representation should be of use in distinguishing patterns belonging to different clusters, immune to noise, and easy to obtain and interpret \cite{20}. Considering these generic requirements as well as the task-specific requirements (cf. Sec. [I]), we propose a two-step process to finally construct a matrix feasible for serving as input for the subsequent process step of clustering.

We utilize the principle of a grid-based representation, within which a certain spatial area around the ego is discretized and defined as potentially relevant in terms of the ego behavior. The basic principle has already been proposed
for use within the robotic domain [40] and various modules of an ADS, such as decision making or motion planning [41]. Furthermore, Gruner et al. [42] propose the use of such grid-based representations for scenario classification based on object-list data and confirm the usability of this type of representation for training artificial neural networks. In contrast to [42], we propose the construction of such a discrete, multi-channel grid structure per scenario and not per scene. The reasons for this are on the one hand the offline character of the scenario extraction use case and the knowledge of the evolution of an entire scenario based on naturalistic road traffic data. On the other hand, our approach aims at maximizing the degree of automation with respect to scenario labeling through unsupervised machine learning approaches, while [42] implements a rule-based approach for a semi-automated scene labeling. In addition, the previous spatiotemporal filtering already accounts for the temporal evolution of a scenario and scenes are assigned to a respective concrete scenario, even if the label of the corresponding scenario is still unknown. Finally, our investigations showed that the scenario-level representation requires significantly less computational effort.

1) **Key Frame Calculation:** This step comprises the determination of the point in time or frame, respectively, of the corresponding concrete scenario (key frame), which is used for the subsequent construction of the different grid channels defining a scenario tensor. On the one hand, the determination of this key frame should be applicable as generically as possible to all relevant ego-challenger combinations identified. On the other hand, the snapshot of the scenario created on the basis of the key frame should capture the distinguishing characteristics of the specific scenario category as accurately as possible.

Our investigations have shown that a computation based on the maximum of the yaw rate of the ego or challenger within the specific concrete scenario entails the best compromise on robustness and computational effort. In detail, first the maximum yaw rates of the RU involved in a concrete scenario are calculated for all ego-challenger combinations within the intersecting scenarios. Then, the maximum of the respective yaw rate value set is calculated. In case of \( \varphi_{e,\text{max}} \geq \varphi_{c,\text{max}} \), the frame associated with \( \varphi_{e,\text{max}} \) is used to construct the scenario tensor. Since the construction of the scenario tensor is always done with respect to the ego-vehicle state, in case of \( \varphi_{e,\text{max}} < \varphi_{c,\text{max}} \), a shift of time to the associated challenger frame has to take place. In Fig. 2 the resulting key frames are implicitly shown by the exemplary occupancy grid channels for two concrete scenarios.

2) **Scenario Tensor Construction:** The key frame for each concrete scenario defines which point in time or frame, respectively, is to be used to construct the corresponding scenario tensor \( \Phi = \{G_1, \ldots, G_j, \ldots, G_l\} \), where \( G_i \) is the \( i \)th grid channel matrix of size \( (a_{gr} \cdot r_{gr,1}) \times (a_{gr} \cdot r_{gr,2}) \). All feature values within the \( l \) grid channels are calculated with respect to the ego coordinate system at the key frame. Thus, the scenario tensor consists of a discrete, multi-channel grid structure, representing the scenarios spatiotemporal characteristics from a bird’s-eye view. A feasible number of grid channels \( l_{gr} \) is determined by the result of previous PFA, with \( l_{gr} \) less or equal to the dimension of the reduced intersecting trajectories dataset. Both the region of interest \( a_{gr} \) of the rectangular grids and the resolution vector \( r_{gr} = (r_{gr,1}, r_{gr,2}) \), containing the longitudinal and lateral grid resolution, respectively, can be adapted to the specific application. While \( a_{gr} \) determines how far the ego looks into space and time, \( r_{gr} \) determines how fine the grid resolves the spatiotemporal evolvement of the respective scenario.

In the lower part of Fig. 2, the occupancy grid \( G_1 \), containing both the trajectories of the ego and the nearest challenger as well as grid cells occupied by other surrounding RUs for the corresponding key frame, is visualized for an exemplary ego-to-vehicle and ego-to-pedestrian scenario. In agreement with [42], this scenario representation is basically applicable to different ODDs, adaptive by means of the region of interest and resolution. Furthermore, other grid forms are conceivable. It should be noted that this type of scenario representation addresses the fulfillment of essential requirements stated (e.g., Req. M1, M2, M3, M4, and S1).

To obtain a representation suitable for clustering, each scenario tensor is flattened column-wise into a scenario grid matrix of size \( (a_{gr} \cdot r_{gr,1}) \times (a_{gr} \cdot r_{gr,2} \cdot l) \). After standardization as well as dimensionality reduction by applying PCA, the entirety of resulting scenario grids can finally be forwarded for the following process step as cluster input matrix \( M_e \).

D. **Clustering**

The following process step of the UnScenE method consists of the actual application of a clustering procedure. The structure of the cluster input matrix \( M_e \) basically allows the application of any clustering algorithm based on static data. Since the scenario extraction method is in general applicable to different data sources and traffic spaces, the user can and
must select the most appropriate clustering approach for the dataset at hand.

The evaluation in the scope of this paper shows the most promising clustering results using hierarchical agglomerative clustering (cf. Sec. IV). This is in accordance with literature, in which HAC is described as suitable for use cases with many clusters and many samples [21]. Furthermore, this is in line with other publications in the context of real-data based scenario extraction, both based on similarly structured data sources [33] as well as for other types of data sources [34].

E. Cluster Validation and Result Interpretation

Clustering is a subjective process in nature due to the absence of a ground truth [20], since the main goal of the clustering and the subsequent result interpretation itself is to find these before unknown labels as good as possible within the given data. Clustering results heavily depend on the choice of the clustering approach, and even for the same algorithm, the selection of related parameters [20]. Furthermore, clustering results are a matter of view and the definition of similarity highly depends on the problem.

In literature, three different testing criteria categories, namely external, internal and relative indices, are defined to be able to estimate the quality of clustering results [20]. In accordance with present work in the field of unsupervised scenario extraction (e.g., [33], [34], and [38]), one branch of our evaluation approach can be assigned to the external indices category, where external information is used as standard to validate the clustering results. For this purpose, we use the map information available within the datasets in the form of images of the corresponding traffic spaces. In detail, we compare the trajectories involved in different concrete scenarios belonging to a respective cluster by visual validation. In addition, we compare the clustering results for the Bendplatz traffic space with the results of a rule-based baseline approach, since for the latter the elaborate rule-based implementation was intentionally accepted to get a better reference to the presented extraction method. Finally, we compare different clustering structures in order to obtain a reference for deciding which one may best reveal the characteristics of the data. Since cluster analysis is not a one-shot process [20] the method includes feedback loops depending on the validation results for trials and repetitions with an adjusted parametrization at different process steps. In case that the validation criteria are met, the extracted relevant scenarios can be used for subsequent applications, e.g., scenario-based testing of an ADS.

IV. Evaluation

In this chapter, we evaluate the UnScenE method for extracting relevant urban traffic scenarios using the inD dataset [23] as well as the Silicon Valley Intersections dataset (sv dataset) [24]. Some process steps are evaluated for the entire datasets to get an impression of the generic character of the method. Other process steps are demonstrated for exemplary traffic spaces in order to illustrate specific advantages and drawbacks of the method in depth.

A. Datasets and Parametrization

There are various reasons for choosing the inD dataset for the main part of the evaluation of the scenario extraction method. The main advantages of the inD dataset compared to other similar datasets are its size, representativeness and accuracy [23]. The dataset consists of more than 11,500 naturalistic road user trajectories including cars, trucks and busses as well as about 5,000 pedestrian and bicyclist trajectories recorded at four unsignalized intersections in Aachen, Germany. In particular, the high proportion of vulnerable road users (VRU) makes this dataset an interesting challenger to the method (cf. Tab. I), as it can be assumed that it is especially difficult to detect recurring patterns within the less rule-based behavior of VRU. In addition, we use the sv dataset [24] to evaluate individual process steps. This dataset includes naturalistic road traffic data for seven traffic spaces in the Silicon Valley area. Due to the diversity of the traffic spaces as well as changed boundary conditions (especially layer 1-4 of the six-layer model) compared to the inD dataset, more comprehensive conclusions regarding the feasibility of the extraction method can be drawn.

As shown in Tab. I, Bendplatz traffic space and Frankenburg traffic space of the inD dataset both have a high total number of trajectories and a high percentage of VRUs. This makes these two traffic spaces particularly interesting for in-depth studies, since a higher share of multimodal interaction can be assumed.

For the subsequent studies, the following parametrization was performed: $t_{PET} = 3$ s, $d_{traj} = 1$ m, $\text{var}_{\text{PFA}} = 0.95$, $a_{gr} = 30$ m, $s_{gr,1} = s_{gr,2} = 1$ px/m, and $\text{var}_{\text{PCA}} = 0.99$. Even if the numerical values have been chosen on the basis of expert judgement and, if possible, literature references [42], [43], they should not be given too much weight, since the focus is on demonstrating the applicability of the method.

B. Spatiotemporal Filtering and Feature Selection

While the application of the search algorithm for spatiotemporal filtering is trivial, interesting conclusions can already be drawn on the basis of the proportion of the remaining relevant intersecting trajectories dataset. Since the datasets provide a classification regarding the respective RU types (cars, trucks, pedestrians, and bicyclists), high-level scenario categories relevant from the viewpoint of proving the safety of an ADS can be defined as ego-to-vehicle (e-to-v) scenarios, ego-to-pedestrian (e-to-p) scenarios, and ego-to-bicyclist (e-to-b) scenarios. For the example of the

| Table I
| IN D DATASET OVERVIEW |
|-----------------------|
| **Traffic Space**     | **Share in total rec. RU** | **VRU/Vehicles** |
| Heckstraße            | 0.09                        | 0.07              |
| Aseag                 | 0.17                        | 0.12              |
| Bendplatz             | 0.28                        | 0.5               |
| Frankenburg           | 0.46                        | 1.6               |
Bendplatz traffic space, it is found that on average about 4% of pedestrians, 14% of vehicles, and 15% of bicyclists remain within the dataset for the subsequent process steps. These numerical values may serve as an indication of the combination of external risk to which the various RU types are exposed and internal RU type specific risk tolerance. In addition, this gives a hint on the required amount of recordings, even with the conservative parametrization performed, most of the original amount of data is classified as not relevant for the task at hand.

Fig. 3 shows exemplary results of the principal feature analysis in terms of the cumulative explained variance $\text{var}_{\text{PFA}}$ over the number of features for different traffic spaces and RU types. Note, that even if this is not a continuous function, the continuous representation of the graphs supports the understanding of the respective results. Fig. 3 shows the curves for the object type car for the traffic spaces with the minimum and maximum number of features required for the defined cumulative explained variance for the corresponding dataset. It is noticeable that the extrema of the sv dataset wrap around those of the inD dataset. This corresponds to the intuitive expectations, since the traffic space $\text{sv}_{\text{san_jose}}$ (lat.: 37.3627, long.: -121.8759) is located in direct proximity to a highway entrance and only vehicles are present. It is also plausible that traffic space $\text{sv}_{\text{santa_clara}}$ (lat.: 37.3252, long.: -121.9490) requires more features for capturing the same cumulative explained variance, since it includes a complex parking lot with two 4-way stops and six normal stop entries.

While pedestrians in general need more features for the same value of cumulative explained variance, the analysis of the inD dataset shows that the influence of the traffic space is marginal for this RU type, which is why only one curve is shown here. On average, the 11 features considered ($x\text{Center}$, $y\text{Center}$, $\text{heading}$, $x\text{Velocity}$, $y\text{Velocity}$, $x\text{Acceleration}$, $y\text{Acceleration}$, $\text{lonVelocity}$, $\text{latVelocity}$, $\text{lonAcceleration}$, $\text{latAcceleration}$) [23] can be reduced to about the half, with pedestrians defining the lower threshold. While this seems intuitively logical for the inD dataset, since some features differ only by the reference coordinate system, this process step can be an even more valuable tool for objective feature selection for differently structured datasets.

C. Feature Extraction

The evaluation of the following process steps is conducted exemplarily for the Bendplatz traffic space of the inD dataset (cf. Fig. 2). The PFA results in seven features for $\text{var}_{\text{PFA}} = 0.95$, with the pedestrian class forming the feature superset of all RU classes. On this basis, the features $x\text{Center}$, $y\text{Center}$, $\text{heading}$, $x\text{Velocity}$, $y\text{Velocity}$, $x\text{Acceleration}$, and $y\text{Acceleration}$ are used for the construction of the scenario tensor. Since the information of the features $x\text{Center}$, $y\text{Center}$, and $\text{heading}$ can be combined in the occupancy grid channel, this results in a total of five grid channels consisting of the occupancy grid and velocity and acceleration grids in both spatial directions with regard to the ego state at the respective key frame. While Fig. 2 shows the deviation of the feature representation for the case of semantically dissimilar scenarios, Fig. 4 shows the similarity of the feature representation for semantically similar scenarios using the example of the $x$-velocity grids for two e-to-v scenarios (same scale). The underlying scenarios were chosen randomly from the group of left turning vehicles, respectively, out of the cluster visualized in Fig. 6.

D. Clustering

The reasons for choosing HAC as clustering method are described in Sec. III. HAC follows a bottom-up approach, starting with $N$ clusters, each of which includes exactly one sample. A series of merge operations follows based on a linkage criterion until a threshold is reached or all samples are forced to the same cluster [20], [33]. The results of HAC can be visualized by dendrograms, which are characterized by an inverted tree-like structure [43]. While the root node of a dendrogram represents the entire dataset, each leaf can be regarded as one sample. Therefore, the intermediate nodes describe the degree of similarity between samples and the height of each branch represents the distance between a pair of samples or clusters, or a sample and a respective cluster [20].

Fig. 5 shows the dendrogram for the e-to-v scenarios at Bendplatz, created based on all available recordings. Our
experiments have shown the best results for this scenario category with ward linkage criterion. On the one hand, this shows that the different process steps of the method are consistent up to this point and that the developed feature representation is compatible with static data based clustering methods. On the other hand, however, clustering algorithms can always produce a partitioning given a dataset, regardless of whether or not there exists a particular structure in the data [20]. Thus, it is only the subsequent cluster validation and result interpretation that provides the actual gain in knowledge and therefore confidence in the clustering results.

E. Cluster Validation and Result Interpretation

Our approach to validate the cluster results involves three branches (cf. Sec. II). The first branch consists of a comparison of different clustering methods and their respective parametrizations. Thereby, we present the parametrizations that produced the best results in the scope of this paper. The second branch covers a visual validation approach, as described by Fig. 6 and Fig. 7.

In detail, this approach entails the comparison of RU trajectories in different concrete scenarios of one specific cluster, randomly chosen for a defined distance threshold. The trajectories shown in Fig. 6 belong to all concrete e-to-v scenarios within the green, red, and purple clusters of the left side of the dendrogram, all of which combine into a single cluster as the distance threshold value increases up to 240. This result is promising since only semantically similar scenarios are included within the cluster, which can all be assigned to a left-turning ego-vehicle in the presence of oncoming traffic (cf. Fig. 6). In addition, this example illustrates the direction independence of the feature representation, as semantically similar scenarios are clustered regardless of their specific position within the traffic space.

While 369 concrete e-to-v scenarios were available for the clustering at the Bendplatz traffic space, a significantly lower number of 33 e-to-p scenarios remained as cluster input due to their lower probability of occurrence. In analogy to the procedure for the e-to-v scenarios, Fig. 7 shows such e-to-p scenarios, all of which originate from a specific

- **Fig. 5.** Dendrogram depicting the clustering structure of ego-to-vehicle scenarios at Bendplatz (color-coded for distance threshold value of 180).

- **Fig. 6.** Semantically similar ego-to-vehicle scenarios.

- **Fig. 7.** Semantically similar ego-to-pedestrian scenarios.
cluster. The impression that the extraction method is able to cluster semantically similar scenarios is strengthened. This impression also applies to the case of multimodal interaction in Fig. 7 including less rule-based pedestrian behavior.

To get a broader impression about the performance of the scenario extraction method, the third branch of the validation procedure includes the attempt to get a statement about the accuracy of the overall clustering. For this purpose, we compare the results of the clustering process with respect to the e-to-v scenarios for the Bendplatz traffic space with a rule-based baseline approach particularly designed for this traffic space [19]. The methodology for determining overall accuracy is based on [43] and has been extended for this use case. In detail, this involves assigning to each sample one of the nine associated scenario labels originating from the rule-based approach [19]. Subsequently, the most frequently occurring labels per cluster are summed up over all clusters and are divided by the total number of samples. Fig. 8 shows the results of this procedure for a varying distance threshold. As the distance threshold increases, the number of clusters decreases, which is accompanied by a decreasing accuracy. With respect to the dendrogram, increasing the distance threshold corresponds to cutting the tree-like cluster structure at a higher point. As can be seen from Fig. 8 there is a large drop in overall accuracy from 0.75 to 0.56 in the range of the distance threshold between 180 and 190. This corresponds analogously to a transition from 21 to 16 clusters. With a distance threshold of 180 and a cluster number of 21, it can be observed that many clusters have a very high accuracy, while a few larger clusters show low accuracy. In detail, 18 of the 21 clusters have an accuracy above 0.9, while one cluster with an accuracy of 0.32 is the lower outlier. Finding the reasons for this result is subject of future work.

It should be noted that the labels of the rule-based approach used for determining the accuracy do not represent the ground truth either. Thus, false positive as well as false negative ratings of the accuracy can occur. Nevertheless, this approach offers a building block for the determination of a sufficient number of scenario clusters with respect to a reasonable balance between scenario coverage and effort during validation and testing of an ADS.

V. CONCLUSION AND OUTLOOK

This paper proposes a method for extracting multimodal urban traffic scenarios from naturalistic road traffic data in an unsupervised manner. The method comprises five process steps, namely spatiotemporal filtering, principal feature analysis, feature extraction, clustering and cluster validation. The results of the principal feature analysis show a dependence of the cumulative explained variance within the data on both traffic space and road user type. In addition, the required features for the subsequent steps can be reduced by a factor of about two within the exemplary evaluation. For feature extraction, a discrete, multi-channel grid structure for scenario modeling is proposed, resulting in a scenario tensor with various grids defined by the previously selected features. This feature representation particularly addresses the requirements M3, M4, M5, S1, S2, and C1 (cf. Sec. II).

Based on the developed feature representation it is possible to apply clustering methods which rely on static data. To evaluate the method, hierarchical agglomerative clustering is applied to an urban traffic space and the corresponding results are discussed. Thereby, both the results of visual validation for selected multimodal urban traffic scenarios and the comparison with a rule-based based baseline approach are promising. These results confirm in particular the fulfillment of the requirements M1 as well as M2 (cf. Sec. II).

Based on the results, future research should push for a broader evaluation of the method by investigating more traffic spaces as well as more clustering approaches. Furthermore, answering the question for the few, large outliers in the cluster-individual overall accuracy represents a relevant aspect. In addition, addressing requirement W1 by incorporating other data sources is of interest. Moreover, the incorporation of the method into the scenario-based simulation platform presented in [19] represents a future use case, with the extracted scenarios representing one main input channel for the adaptive replay-to-sim approach (ARTS), in addition to an agent-based simulation and the ADS-under-test. Finally, the method will be integrated into a semi-supervised machine learning pipeline to achieve the medium-term goal of a robust and generic scenario classifier. The presented method contributes to the quantitative real-data based extraction of relevant traffic scenarios. This method can be seen as building block toward a systematic, data-driven construction of a relevant scenario database with sufficient coverage in a fully automated manner for validating the safe behavior of ADS.

ACKNOWLEDGMENT

We want to thank Stefan Berger and Ulrich Eberle (Opel Automobile GmbH, Stellantis NV) for the productive discussions and the feedback on our approaches as well as for peer review of this publication prior to its submission.
REFERENCES

[1] Society of Automotive Engineers, Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, Std. J3016.
[2] M. Wood et al., “Safety first for automated driving.” [Online]. Available: https://www.daimler.com/dokumente/innovation/sontiges/safety-first-for-automated-driving.pdf
[3] International Organization for Standardization, Road Vehicles - Safety of the Intended Functionality, Std. ISO/PAS 21 448.
[4] S. Hallerbach, Y. Xia, U. Eberle, and F. Koester, “Simulation-based identification of critical scenarios for cooperative and automated vehicles,” SAE International Journal of Connected and Automated Vehicles, vol. 1, no. 2018-01-0966, pp. 93–106, 2018.
[5] W. Wachenfeld and H. Winner, “The release of autonomous vehicles,” in Autonomous Driving: Technical, Legal and Social Aspects, M. Maurer, J. C. Gerdes, B. Lenz, and H. Winner, Eds. Springer, 2016, pp. 425–449.
[6] Statistisches Bundesamt, “Verkehrsunfälle,” 2021, Accessed: Feb. 12, 2021. [Online]. Available: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/jmh/ html
[7] PEGASUS Project Consortium, “Pegasus method: An overview,” 2019, Accessed: Jan. 20, 2022. [Online]. Available: https://www.pegasusprojekt.de/files/mpf/ Pegasus-Abschlussveranstaltung/PEGASUS-Gesamtmethode.pdf
[8] S. Ulbrich, T. Menzel, A. Reschka, F. Schultid, and M. Maurer, “Defining and substantiating the terms scene, situation, and scenario for automated driving,” in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 982–988.
[9] Daimler AG, “First internationally valid system approval for conditionally automated driving,” 2021, Accessed: Jan. 17, 2022. [Online]. Available: https://www. daimler.com/innovation/product-innovation/autonomous-driving/ system-approval-for-conditionally-automated-driving.html
[10] S. Riedmaier, F. D. Ladwig, B. Schick, and F. Diermeyer, “Survey on scenario-based safety assessment of automated vehicles,” IEEE Access, vol. 8, pp. 87 456–87 477, 2020.
[11] C. Neurohr, L. Westhofen, M. Butz, M. H. Bollmann, U. Eberle, and R. Galbas, “Criticality analysis for the verification and validation of automated vehicles,” IEEE Access, vol. 9, pp. 18 016–18 061, 2021.
[12] A. Putz, A. Zlocki, J. Bock, and L. Eckstein, “System validation of highly automated vehicles with a database of relevant traffic scenarios,” in 12th ITS Eur. Congr, 2017.
[13] N. Weber, D. F. Frerichs, and U. Eberle, “A simulation-based, statistical approach for the derivation of concrete cases for the release of highly automated driving functions,” in AutomE 2020 - Automotive meets Electronics; 11th GMM-Symposium on Automotive Electronics; 11th GMM-Symposium, vol. 1, no. 2018-01-1066, pp. 93–106, 2018.
[14] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, “The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections,” in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 1929–1934.
[15] T. W. Liao, “Comparison of requirement prioritization techniques to find best prioritization technique,” International Journal of Modern Education & Computer Science, vol. 7, no. 11, 2015.
[16] T. W. Liao, “Clustering of time series data—a survey,” Pattern recognition, vol. 38, no. 11, pp. 1857–1874, 2005.
[17] E. A. Maharaj, P. D’Urso, and J. Caiado, Time series clustering and classification. Chapman and Hall/CRC, 2019.
[18] S. Rani and G. Sikka, “Recent techniques of clustering of time series data: a survey,” International Journal of Computer Applications, vol. 52, no. 15, 2012.
[19] Wang Wenshuo, Ramesh Aadiya, Zhu Jiacheng, Li Jie, and Zhao Ding, “Clustering of driving encounter scenarios using connected vehicle trajectories,” IEEE Transactions on Intelligent Vehicles, vol. 5, no. 3, pp. 485–496, 2020.
[20] E. de Gelder and J. P. Paardekooper, “Assessment of automated driving systems using real-life scenarios,” in 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 589–594.
[21] A. Erdogan, B. Ugranli, E. Adali, A. Sentas, E. Mungan, E. Kaplan, and A. Leitner, “Real-world maneuver extraction for autonomous vehicle validation: A comparative study,” in 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 267–272.
[22] H. Watanabe, T. Maly, J. Wallner, T. Dirndorfer, M. Mai, and G. Prokop, “Methodology of scenario clustering for predictive safety functions,” in 9. Tagung Automatisiertes Fahren, 2019.
[23] J. Kerber, S. Wagner, K. Groh, D. Notz, T. Kühbeck, D. Watzening, and A. Knoll, “Clustering of the scenario space for the assessment of automated driving,” in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 578–583.
[24] F. Montanari, R. German, and A. Djanatliev, “Pattern recognition for driving scenario detection in real driving data,” in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 590–597.
[25] L. Hartjen, R. Philipp, F. Schuldt, B. Friedrich, and F. Howar, “Classification of driving maneuvers in urban traffic for parametrization of test scenarios,” in 9. Tagung Automatisiertes Fahren, 2019.
[26] M. Barbier, C. Laugier, O. Simonin, and J. Ibáñez-Guzmán, “Classification of drivers manoeuvre for road intersection crossing with synthetic and real data,” in 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 224–230.
[27] C. King, T. Braun, C. Braess, J. Langner, and E. Sax, “Capturing the variety of urban logical scenarios from bird-view trajectories,” in VEHITS, 2021, pp. 471–480.
[28] L. Ries, P. Rigoll, T. Braun, T. Schulik, J. Daube, and E. Sax, “Trajectory-based clustering of real-world urban driving sequences with multiple traffic objects,” in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021, pp. 1251–1258.
[29] P. J. Cooper, “Experience with traffic conflicts in canada with emphasis on “post encroachment time” techniques,” in International calibration study of traffic conflict techniques. Springer, 1984, pp. 75–96.
[30] T. W. Liao, “Clustering of driving encounter scenarios using connected vehicle trajectories,” in Proceedings of the 15th ACM International Conference on Multimedia, ser. MM ’07. New York, NY, USA: Association for Computing Machinery, 2007, pp. 301–304.
[31] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, “The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections;” in 2020 IEEE Intelligent Vehicles Symposium (IV), 2020, pp. 1929–1934.
[32] levelXdata, “Silicon valley intersections dataset,” 2020, Accessed: Jan. 28, 2022. [Online]. Available: https://levelXdata.com/
[33] J. A. Khan, I. U. Rehman, Y. H. Khan, T. J. Khan, and S. Rashid, “Comparison of requirement prioritization techniques to find best prioritization technique,” International Journal of Modern Education & Computer Science, vol. 7, no. 11, 2015.
[34] A. A. Patel, “Experience with traffic conflicts in canada with emphasis on “post encroachment time” techniques,” in Post encroachment time techniques, 2007, pp. 246–2470.
[35] M. Schuldes, L. Westhofen, L. R. Taylor, K. Lott, M. Schuldes, H. Winner, N. Wagener, C. Neurohr, M. H. Bollmann, F. Körtke, J. Miller, M. Hess, J. Bock, and L. Eckstein, “6-layer model for a structured description and categorization of urban traffic and environment,” IEEE Access, vol. 9, pp. 59 131–59 147, 2021.
[36] T. Ponn, M. Breitfuß, C. Braess, J. Langner, and E. Sax, “Clustering of relevant scenarios for testing automated driving systems in urban areas;” in 30th Aachen Colloq. Sustainable Mobility 2021, 2021, pp. 1697–1734, arXiv:2109.03 648.
[37] R. Xu and D. Wunsch, Clustering, 10th ed. Hoboken, New Jersey: Wiley & Sons, Inc, 2008.