Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 119 (2023) 822–827

www.elsevier.com/locate/procedia

33rd CIRP Design Conference

Classifying Parts using Feature Extraction and Similarity Assessment

Louis Schäfer*,[a], Niklas Treml[a], Marvin Carl May[a], Gisela Lanza[a]

[a]*wbk Institute of Production Science, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany*

* Corresponding author. Tel:+49-1523-950-2570; Fax:+49-721-60845005. *E-mail address:* louis.schaefer@kit.edu

Abstract

In wake of customization and individualization of products, the compendium of designed products produced by each manufacturer is ever-increasing. With such an increase in numbers and variants of products complexity soars. However, with this increase an unprecedented rise in implicit knowledge stored in this historic product data comes hand in hand. This paper aims at reducing the complexity within product design and manufacturing by making this implicit knowledge available to engineers. To that end, a toolchain is introduced to support product data formalization which consists of a feature extractor to map CAD model features to a reference data base and similarity assessment to facilitate comparison, understanding and automatic reference model extension. That approach is validated with a car axle assembly in production planning to automate manufacturing sequence deduction from a product model and the implicit knowledge stored in its predecessors and corresponding manufacturing sequences.

Keywords: Similarity Determination; Feature Extraction; Product Part Classification; Production System Planing; Data Analytics.

1. Introduction

Customization and individualization nowadays lead to a larger product variance, shorter product life cycles and, thus, high frequency product development and production planning. Engineering competence and the product development digitization are decisive for today's success of manufacturing companies [7]. In multi-variant production, past product variants and their production system interrelations contain implicit knowledge that remains unused [4]. Therefore, there is the need to support people in dealing with product data in the production planning process [6].

Production systems are geared towards the product variant specifications. Even for non-configurable products, the similarity between production systems follows the similarity of requested and known variants. Especially for suppliers, the product design and specifications defined and provided by the *Original Equipment Manufacturer* (OEM) often define a variant-specific production system. By formalization and integration of product and production knowledge, interdependencies between product variants and their corresponding production systems can be detected and, thus, reused for production planning tasks [25]. Often there is a low involvement supplier-customer relationship in product development. Suppliers receive a request in the form of *Computer Aided Design* (CAD) data and ensure manufacturability through design for manufacturing. On this basis a variant specific production process is planned and commissioned.

This article proposes a methodical approach and corresponding toolchain to support the process of formalizing product data, identifying relevant product properties and using these to set up a modular reference model based on past product variants and similarities. When connecting the product reference model to production system properties, the approach is based on the overall framework, presented by Schaefer et al. [25]. This approach covers the automated classification of each part of an assembly regarding a modular reference system, fed from past product variants, to counter time constraints and complexity by suggesting possible production planning designs and ultimately derive manufacturing sequences from this product data.

Sec. 2 provides an overview over relevant fields of action and literature. The methodical approach for classifying parts by feature extraction and mapping them to a built reference system is outlined in Sec. 3. Sec. 4 shows the application in an industrial setting, using the example of a rear axel manufacturer. Sec. 5 gives an outlook and conclusion.

2212-8271 © 2023 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2023.03.127

2. Fields of Action

Automatically determining product similarities requires an in depth feature extraction, setting up a reference system finally classification, which are individually introduced.

2.1. Feature Extraction

Reusing existing product models and the knowledge contained in them shows a great deal of promise for improving time of development [13]. Due to the absence of data driven approaches and the difficulty to formalize implicit knowledge, this knowledge base is often not used systematically [3]. Thus, the input description for product data must be defined, for instance as 3D product models from CAD Software. Product models vary in their applicability to the object of study (feature extraction) depending on the underlying data. *Engineering-Bill of materials* (E-BOM) for example are structured regarding products in the design phase, consisting of part data like design visualizations, names and geometric parameters [18]. While *Manufacturing-Bill of materials* (M-BOM) are used for manufacturing and assembly and therefore containing meta-information about assembly groups, hierarchically structured product modules, material and part lists, and production-related information such as quantities and relationships between parts of the assembly properties [24]. In CAD models all pertinent geometrical parameters are included in a machine-readable format [2], which makes a data driven extraction and formalization of product data possible [22]. Features like dimensions, mass or volume can be used as machine- and human-readable data carriers for subsequent process steps. [2]. This data can be automatically identified and linked to its part or the corresponding product variant and stored in a database. [10]. Feature extraction as shown by [31] implements a *Machine Learning* (ML)-based method to identify relevant machining features in 3D models, while [15] tries to establish a design assistance system by utilizing an autoencoder to generate 3D point clouds from CAD models. The American Society of Mechanical Engineers developed a *model-based definition* (MBD) to embed digital *product manufacturing information* (PMI) by 3D annotations directly to the product data in 3D models. [5]

2.2. Reference System

In literature, reference systems are often used in product development, while the usage in the production planning process is uncommon [25]. The use of modeled reference systems for products and production systems enables the mapping and use of company-specific variants and their heterogeneous product characteristics [19]. While the general layout of reference structures may vary in practice [20, 26], focusing on product variants not product generations, modular product reference system, like "MQB" by Volkswagen, are required [25]. [9] presents a framework for product modelling that supports developers in the synthesis and analysis of modular products. During the early stage of product development, it demonstrates how data from previous product generations may be leveraged to generate product models using *Model-Based Systems Engineering* (MBSE). The *association of the German automotive industry* (VDA) recognizes the necessity of beyond geometry-related 3D focused data management concepts, inclduding alphanumeric and non-geometry-related metadata to link product data to its respective function in a digital form [28]. [12] presents an approach that aims at supplementing a product model with a process and production system model to map their interrelations. Here, [12] uses the digital product data from the product model as input for the production system design, based on the formalization of design knowledge.

2.3. Part Classification

Key literature covering retrieval and clustering of product part data [11] highlight the necessity to *preprocess data* for part classification and similarity assessment to categorize historic and available product data into their predefined parts and attach relevant production information as formalized meta-data [23]. This is crucial to enable knowledge transfer through successful product matching for similarity assessment. With feature-based part classification, the goal of data preprocessing is to identify the relevant features (e.g. length, width or weight).

Product variants and their corresponding production systems are distinguished by a high degree of similarity in parts and production processes used. This environment of variant production results in the possibility to perform accurate *similarity assessment* between new parts of a new variant and a reference data set of historic variants. The assessment of part similarities is carried out to match parts of new variants (unlabelled data) to the reference system (preprocessed and labelled data) and the corresponding meta information [25]. While linguistic methods like [14, 27] map text modules, [24] compares products' bill of materials for minimal cost of change. More complex approaches for representational learning use ML-based methods to compare similarities based on product point clouds like [1] or Generative Adversarial Networks [17, 30] to independently extract relevant information from the data. But most approaches are based on CAD feature recognition and extraction [21, 23, 29, 16]. [23] assesses similarities by extracting and comparing geometric features using e.g. k-means clustering. However, the focus is on extracting features rather than matching parts to apply the underlying implicit knowledge to the variants in comparison.

2.4. Summary of Related Work

In summary, for each sub-step of the covered approach, solutions with different limitations exist. Today the majority of models is relying on only one information input (i.e. semantic or geometric data). While the use of different feature classes for a two-stage similarity assessment is rare in practice, it increases its accuracy. For production planning with a low amount of (training) data cannot be achieved with currently existing approaches such as [16] and [31]. The approach should thus use a minimum number of product variants and should not require large training datasets or complex ML-based 3D extraction.

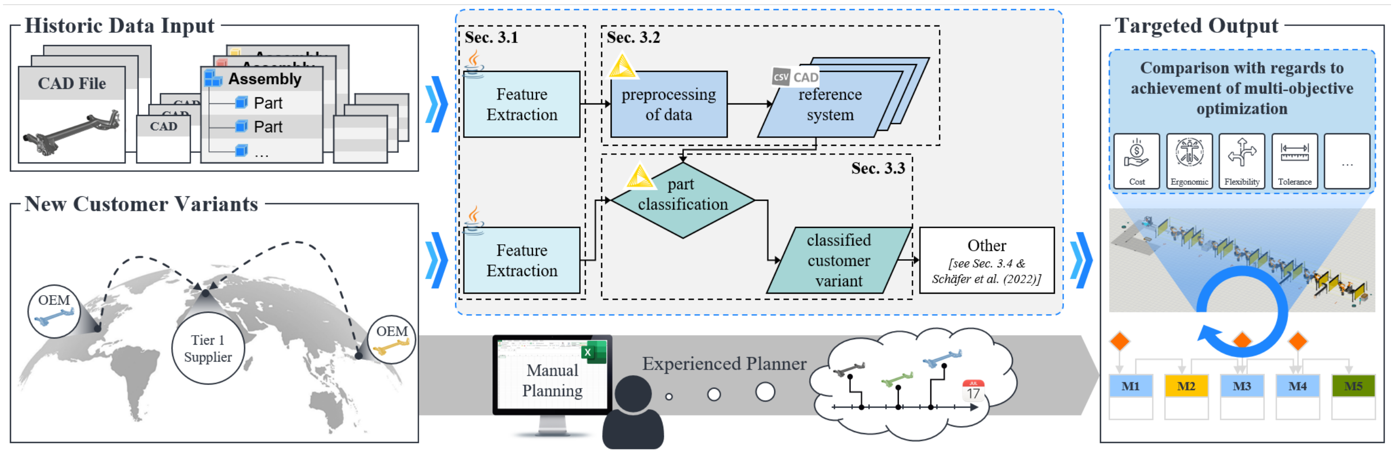

Fig. 1. Methodical Approach and the Scope of this Article embedded into the Big Picture of Assisted Production Planning according to [25].

3. Methodical Approach

Following the findings from the preceding section, the approach presented in this article revolves around product variant specific assembly system planning. This process represents repetitive activities, especially in the automotive supplier industry, due to the high production quantities of the individual customer variants. To support the production planner in this process, implicit knowledge is to be formalized and used (by means of using similarities between historic and current product variants) in the automated derivation of assembly priority graphs and the planning of the assembly system. Fig. 1 visualizes this overarching approach, highlighting the contribution that this article is intended to provide. It depicts the process as being separated into three major steps: First, a tool for extracting features from 3D product models is presented. This technique may be applied to both old and new products (Sec. 3.1). The data from historic variants is then analyzed and preprocessed to cluster the parts (Sec. 3.2). With every new variant, where a variant-specific production system is planned, similarities between their parts and parts of the historic data input are assessed and every part is classified by means of the reference system (Sec. 3.3). Further details are given in the following.

3.1. Feature Extraction

Feature extraction aims at abstracting the knowledge contained within CAD data and E-BOM and making it accessible for the subsequent similarity assessment of parts. Here, automating the process to efficiently support humans within repetitive planning activities is targeted. Customer variants within a specific product family share certain characteristics in their structure represented by their respective parts lists. By utilizing the distinct geometric and semantic imprint from all parts of a product variant represented by the extracted features, a reference system may be generated from the sum of all clustered parts. CAD models together with their E-BOM incorporate all pertinent export information, geometric characteristics (height, width, volume, etc.), geometric descriptive features (mass, material, quantity of predefined geometric shapes, such as cones, surfaces, etc.), and a semantic feature that denotes the identifier of each part. The data export is possible due to the machine-readable structure. To access the CAD application and export data, *Application Programming Interfaces* (APIs) are used. In order to further process the extracted features the data is stored within a *Comma-seperated Value* (CSV) export format.

3.2. Reference System

In the reference system, the properties of the historic customer variants (see also top workflow in Fig. 1) can be stored based on predefined characteristics. In order to organize the current database, for example, according to a geometric design, data must be divided into part classes. Clustering the data using the extracted features from Sec. 3.1 helps determining these classes. The subsequent similarity analysis of new product variants, which looks similar for similar parts across production history, is built on this foundation. To set-up a reference system usable for similarity assessment of new product variants a preprocessing of historic variant data aims at identifying features that are most suitable for differentiating parts within the assembly and therefore most suitable for determining similarities between parts across different customer variants.

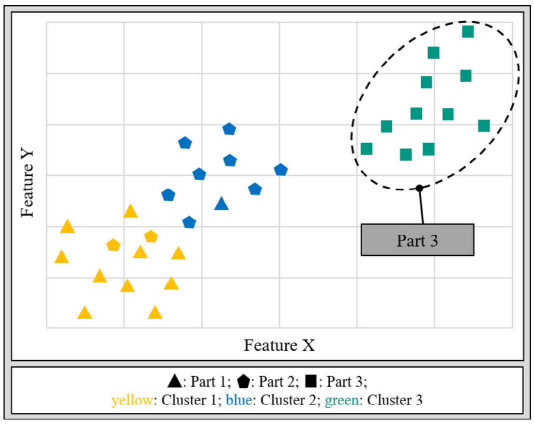

Fig. 2. Clustering of historic Product Data.

Fig. 2 depicts how historical product data is preprocessed. Through k-means clustering, it is possible to identify features that help to determine part classes, to derive differences between classes and to use those differences to classify new parts in Sec. 3.3. Fig. 2 shows how *Feature X & Y* are applicable for isolating *Part 3* as a cluster, however at the same time poorly differentiate *Part 1 & 2*. With suitable features and resulting part classes/labels being identified, all historic variant data is structured and stored forming a reference system, ready for use in part classification.

3.3. Part Classification

If an OEM now requests a new variant for production (see also bottom workflow in Fig. 1), in the very first step CAD files and E-BOM are the only information available to generate a production offer. In order to quickly determine production cost for the automotive supplier, here, similarities to historic product variants can be formally used. Therefore, the predefined features are extracted and can then be used to classify all parts in order to subsequently derive e.g. the assembly sequence.

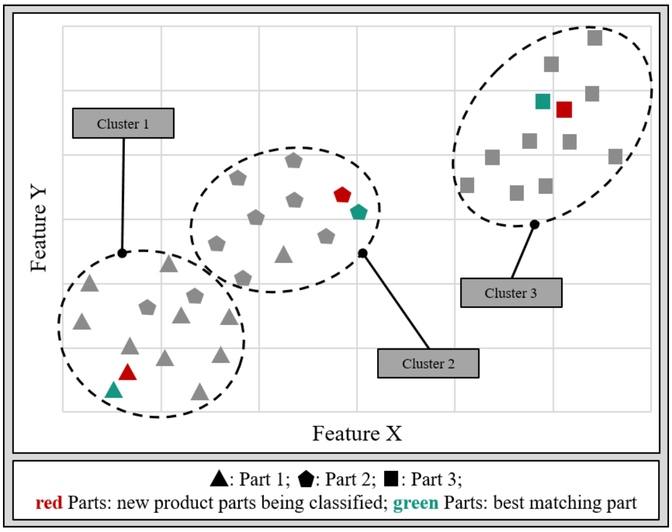

Fig. 3. Similarity Assessment of new Product Variant.

This data can be processed and compared to the historical variants in order to take advantage of the existing knowledge gathered in previous variants and their production processes.

Fig. 3 visualizes this process. Using a similarity analysis such as *k-Nearest Neighbors* (k-NN), unlabeled data from a new variant is assigned to the clustered part modules of the reference system. The nearest neighbor (regarding the geometric features) acts as a *best matching part* resulting in a *best matching product* (BMP) after the recurring algorithm has concluded the processing of all new variant's constituent parts. With this process completed, the newly labeled data from the new variant is automatically assigned to the already-existing part clusters, thereby adding a new variant to the established reference system and allowing to derive production relevant information by applying corresponding part-related planning rules from past variants to the new planning task.

4. Method Application

In the following, the proposed approach is exemplary applied to an industrial setting. After a brief introduction to the use case and the accompanying product peculiarities, the proposed approach from Sec. 3 is exemplarily applied to an industrial setting. The section concludes with a practical outlook on how MBSE enables the interlinkage between product reference system and production process information (Sec. 4.4).

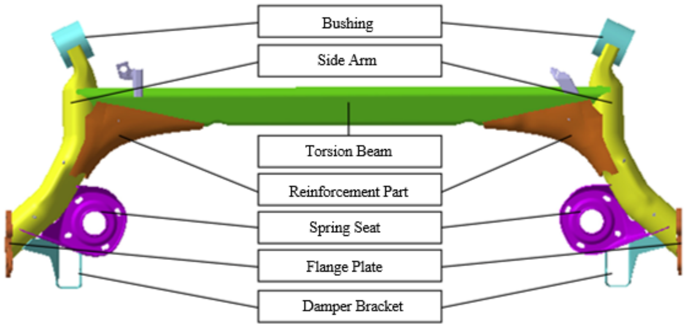

Fig. 4. Rear Twist Beam according to [32].

The product family of *Rear Twist Beam* (RTB) axles serves as an application example (see also Fig. 4). These products are usually developed by the OEM and manufactured and assembled by automotive suppliers in large quantities over many years. Here, the focus mainly lies on the customer variant-specific planning of welding assembly lines. This involves repetitive planning activities. The assemblies are similar enough in terms of their structure, but differ in the components with respect to individual variants. In the application case, CAD assemblies with a total of 60 parts are available. As a prerequisite for production system planning, the following steps are executed in order to formally make use of product similarities.

4.1. Feature Extraction

Regarding the data basis, all assemblies/parts are represented by CAD files and E-BOMs.The E-BOM contains a list of parts (and their respective names), however no meta-information about the assembly structure. A feature extractor has been developed, automatically extracting 18 features via the Java API from PTC CREO. In the following a subset of all (for this exemplary use case most fitting) geometric features (see Table 1) is being used. The API enables a parallel extraction from multiple CAD files and the developed tool automatically stores the data using a standard CSV format.

4.2. Reference System

The extracted features can now be used to cluster the parts. Easy application is possible by using a low-code data analytics software, such as KNIME [8]. Data processing by means of analyzing suitable features to determine similarities between different product variants (on part level) is performed using k-means clustering.

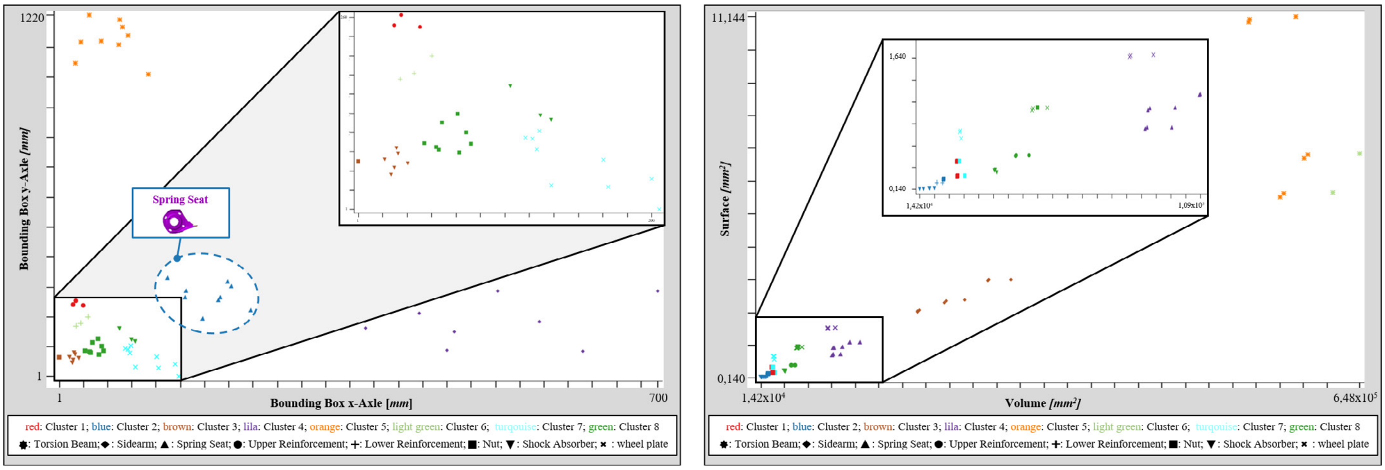

Fig. 5. (a) Suitable Features for Isolating Part Clusters; (b) Less suitable Feature Combination.

Table 1. An Excerpt of the directly extracted Features and their *Unit of Measurement* (UoM).

(Direct) Features	UoM	Features	Formula
Part Name	*text*	Ratio	V/S
Bounding Box x-Axle	*mm*	Density ρ	m/V
Bounding Box y-Axle	*mm*	Compactness	S^3/V^2
Bounding Box z-Axle	*mm*	Crinkliness	$(S^3/V^2)^3$
Volume V	mm^3		
Total Surface S	mm^2		
Mass m	*kg*		

By examining all plausible combinations of the extracted features, the results can be compared and the suitability for isolating correct clusters can be worked out. Fig.4 (a) and (b) provide an illustration of one positive and negative combination of features for clustering. Demonstrated in Fig.4 (a) the clustering of the features x, y, z reaches an accuracy of 93.3%. Parts such as Torsion Beams, Spring Seats, Side Arms, Reinforcement Parts and Flange Plates can all be distinguished effectively, however Bushings and Damper Brackets are more difficult to differentiate (using this selection of geometric features). These examples call for an additional semantic comparison of part identifiers. Although this semantic similarity search follows a distinct workflow, it is nonetheless comparable to the k-means clustering procedure: Using the information provided in the E-BOMs, a historical database of potential names for the part modules (Table 2 yellow row) is compiled and a semantic similarity analysis is carried out. An excerpt of the resulting reference system consisting of all relevant information about historic assemblies and their parts can be seen in Table 2. The structure provides for all parts to be clustered to there respective part module (class label from Fig. 4).

4.3. Part Classification

With the scenario of planning variant specific assembly systems for new customer variants, this step aims at assigning all parts to the classes within the built up reference system. In order to use feature-based classification techniques, all predefined features must be extracted from the product data of the new variant. Here, according to Table 3 several algorithms can be compared. It should be noted that within this use case, through a combination of geometric and semantic similarity search, all 60 parts from the product family can be correctly assigned. With different results from geometric and semantic analysis there are several options: Either a part-dependant default solution is adopted (e.g. default use of semantic analysis), a majority vote is conducted or a human control instance intervenes.

In addition, after all parts are classified, a *best matching part* (nearest neighbor) and the resulting BMP are determined. The BMP enables reuse of production-relevant knowledge in the subsequent planning of the welding assembly line (Sec. 4.4).

Table 2. An Excerpt from the Reference System.

	Part Class ***Torsion Beam***			Part Class ***Side Arm***		...
Features	Variant OEM_1	Variant OEM_2	V_n	Variant OEM_1	...	...
Name	"torsion-beam"	"BEAM"	...	"side_arm"	...	...
Bounding-Box-x	145,05 [mm]	105,23	...	578,58	...	...
Bounding-Box-y	1206,2	1131,5	...	198,85	...	...
Bounding-Box-z	95,71	83,02	...	103,76	...	...
Volume	1.414.328	593.943	...	400.312	...	...
...	...	...	...	...	...	...

Table 3. Results from the Feature-based Part Classification.

Clustering	**Features**	**Accuracy**	
k-means (k-NN)	x, y, z	93,3%	
PCA & k-means (k-NN)	x, y, z, V, S, m, ρ	66,6%	
Classificator	**Features**	**80/20 split**	**70/30 split**
Decision Tree		83,3%	94,4%
Random Forest	*all*	100%	83,3%
SVM		33,3%	72,2%

4.4. Next Steps

During planning, the assembly system planned for the BMP at that time can serve as a first orientation. In detail, with all parts assigned the goal now is to derive a product-specific assembly precedence graph. For this, the part modules from the reference system must be completed with their joint connections. These joints decisively determine the assembly sequence as well as the resulting assembly system. SCHAEFER ET AL. give an overview of how to derive the necessary information. Here, MBSE more precisely *Systems Modeling Language* (SysML) enables explicitly modeling classes (e.g. "part", "process" & "ressource") and elements (e.g. "Torsion Beam") as well as their attributes (e.g. "length" with parts or "duration" with processes) and interrelations (e.g. an allocation of "weld seam" to "welding process"). For details see also [25].

5. Summary and Outlook

This article describes a methodical approach and its exemplary application to use product similarities to improve production system planning and ultimately provide a contribution codesigning products and production system at an early stage. The approach clusters historic parts regarding geometric similarities and enables an automated classification of new parts according to the built up reference system modules. The approach supports humans in the repetitive, variant-specific assembly system planning based on historic data and already designed products. Hereby it reduces the time to market and provides a contribution to maximize efficiency and the reuse of implicit product design knowledge in production system planning. Here, the proposed feature-based approach imposes a minimal requirement on available data.

Further research should focus on validating the approach with other use cases and extending the feature-based method by integrating a geometrically more detailed similarity search such as point cloud representation of parts. This would enable a more specific determination of part variants and their implications for the production system planning.

Acknowledgement

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the *The Future of Value Creation – Research on Production, Services and Work* program (02J19B099) and managed by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the content of this publication.

References

[1] Achlioptas, P., Diamanti, O., Mitliagkas, I., Guibas, L., 2018. Learning representations and generative models for 3d point clouds, in: International conference on machine learning, PMLR. pp. 40–49.
[2] Agovic, A., Trautner, T., Bleicher, F., 2022. Digital transformation-implementation of drawingless manufacturing: A case study. Procedia CIRP 107, 1479–1484.
[3] Albers, A., Börsting, P., Turki, T., 2010. Application of design patterns for the development of primary shaped microsystems: a case study. Multi-Material Micro Manufacture M 4, 2010.
[4] Albers, A., Haug, F., Fahl, J., Hirschter, T., Reinemann, J., Rapp, S., 2018. Customer-oriented product development: supporting the development of the complete vehicle through the systematic use of engineering generations, in: 2018 IEEE ISSE, pp. 1–8.
[5] ASME, 2012. Digital Product Definition Data Practices: Engineering Drawing and Related Documentation Practices. American Society of Mechanical Engineers.
[6] Bauer, F., Gausemeier, J., 2015. Planungswerkzeug zur wissensbasierten Produktionssystemkonzipierung. Heinz Nixdorf Institut, Paderborn.
[7] Bauernhansl, T., Ten Hompel, M., Vogel-Heuser, B., 2014. Industrie 4.0 in Produktion, Automatisierung und Logistik. Springer.
[8] Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., Wiswedel, B., 2009. KNIME - the konstanz information miner. SIGKDD Explor. Newsl. 11, 26–31.
[9] Bursać, N., 2016. Model Based Systems Engineering zur Unterstützung der Baukastenentwicklung im Kontext der Frühen Phase der Produktgenerationsentwicklung. Ph.D. thesis. IPEK, Institut für Produktentwicklung.
[10] Chan, C., Tan, S., 2005. Volume decomposition of cad models for rapid prototyping technology. Rapid Prototyping Journal .
[11] Gezawa, A.S., Zhang, Y., Wang, Q., Yunqi, L., 2020. A review on deep learning approaches for 3d data representations in retrieval and classifications. IEEE access 8, 57566–57593.
[12] Hagemann, S., Sünnetcioglu, A., Stark, R., 2019. Hybrid artificial intelligence system for the design of highly-automated production systems. Procedia Manufacturing 28, 160–166.
[13] Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., Ramani, K., 2005. Three-dimensional shape searching: state-of-the-art review and future trends. Computer-Aided Design 37, 509–530.
[14] Jörg, M.A., 2005. Ein Beitrag zur ganzheitlichen Erfassung und Integration von Produktanforderungen mit Hilfe linguistischer Methoden. Shaker.
[15] Krahe, C., Bräunche, A., Jacob, A., Stricker, N., Lanza, G., 2020. Deep learning for automated product design. Procedia CIRP 91, 3–8.
[16] Krahe, C., Marinov, M., Schmutz, T., Hermann, Y., Bonny, M., May, M., Lanza, G., 2022. Ai based geometric similarity search supporting component reuse in engineering design. Procedia CIRP 109, 275–280.
[17] Li, C.L., Zaheer, M., Zhang, Y., Poczos, B., Salakhutdinov, R., 2018. Point cloud gan. arXiv preprint arXiv:1810.05795 .
[18] Liu, M., Lai, J., Shen, W., 2014. A method for transformation of engineering bill of materials to maintenance bill of materials. Robotics and Computer-Integrated Manufacturing 30, 142–149.
[19] Nurcahya, E., 2007. Configuration instead of new design using reference product structures, in: Future Product Development. Springer, pp. 125–134.
[20] Ponn, J., Lindemann, U., 2008. Konzeptentwicklung und Gestaltung technischer Produkte. Springer.
[21] Rea, H.J., Corney, J.R., Clark, D.E., Pritchard, J., Breaks, M.L., MacLeod, R.A., 2002. Part-sourcing in a global market. CE 10, 325–333.
[22] Röhner, S., Breitsprecher, T., Wartzack, S., 2011. Acquisition of design-relevant knowledge within the development of sheet-bulk metal forming, in: DS 68-6: Proceedings of the 18th ICED.
[23] Roj, R., Sommer, M., Woyand, H.B., Theiß, R., Dültgen, P., 2022. Classification of cad-models based on graph structures & machine learning .
[24] Romanowski, C., Nagi, R., 2005. On comparing bills of materials: a similarity/distance measure for unordered trees. IEEE Transactions on SMC 35, 249–260.
[25] Schäfer, L., Frank, A., May, M.C., Lanza, G., 2022. Automated derivation of optimal production sequences from product data. Procedia CIRP 107, 469–474.
[26] Schuh, G., 2014. Produktkomplexität managen: Strategien-Methoden-Tools. Carl Hanser Verlag GmbH Co KG.
[27] Tseng, T.L.B., Huang, C.C., 2007. Rough set-based approach to feature selection in customer relationship management. Omega 35, 365–383.
[28] VDA, 2014. VDA 4953-2 .
[29] Woo, Y., Wang, E., Kim, Y., Rho, H., 2005. A hybrid feature recognizer for machining process planning systems. CIRP annals 54, 397–400.
[30] Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J., 2016. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. Advances in neural information processing systems 29.
[31] Zhang, Z., Jaiswal, P., Rai, R., 2018. Featurenet: Machining feature recognition based on 3d convolution neural network. CAD 101, 12–22.
[32] Zhao, L.H., Zheng, S.L., Feng, J.Z., 2014. Failure mode analysis of torsion beam rear suspension under service conditions. Engineering Failure Analysis 36, 39–48.