Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 96 (2021) 133–138

www.elsevier.com/locate/procedia

CIRPe 2020 – 8th CIRP Global Web Conference – Flexible Mass Customisation

# Comparing CAD part models for geometrical similarity: A concept using machine learning algorithms

S. Bickel[a,*], C. Sauer[a], B. Schleich[a], S. Wartzack[a]

[a] *Engineering Design, Friedrich-Alexander-University Erlangen-Nuremberg (FAU), Martensstrasse 9, 91058 Erlangen, Germany*

**Abstract**

The efficient execution of process planning activities requires the knowledge from several distinct domains. However, in this context, a common issue is the existence of collective knowledge in the one domain with a lack of networked expertise with other domains. Motivated by this, the paper proposes an approach to support process planning by comparing the generated part design with older, validated products. The utilization of this earlier CAD part models has the potential to reduce development costs, shorten the production start-up time and improve the product quality. The new concept provides the manufacturing personnel with a method for comparing the newly designed part with a pool of validated models to identify the most similar one. Each previous CAD part model is linked with the necessary manufacturing information, so that the initial values for the manufacturing process are available without a time consuming testing phase. The method is divided in three main steps: the global similarity comparison, the segmentation of the part and the local similarity comparison. In the first step, the geometry is projected onto a sphere and then transformed to a matrix. Afterwards these matrices are compared and clustered into corresponding groups. In the following step, a Machine Learning algorithm segments the objects into specific, manufacturing relevant groups. In every cluster, the segmented geometries are again compared for similarity. The combination of the first and the second ranking results in a global similarity hierarchy for the newly designed part. In this paper, the entire procedure is shown with the example of sheet-bulk metal formed parts. This new manufacturing process particularly benefits from this method, as the amount of data is still limited and therefore little expert knowledge exists.



*Keywords:* Digital Engineering; similarity; comparison; Machine Learning; data-driven

## 1. Introduction

The change in manufacturing paradigms can be divided in roughly four segments: The first being the industrial application of steam engines in the late $18^{th}$ century. The production process shifted from hand production methods to machine supported work. This enabled the second change in manufacturing through the integration of the assembly line production by using conveyor belts. Then followed the application of computers and electronics which resulted in partial automation of work steps. The further development of this stage leads to the current change, where smart manufacturing marks the beginning of a new chapter. A typical procedure for smart manufacturing is shown in Figure 1. For this new era in manufacturing, many definitions have been developed. The National Institute of Standard and Technology, for example, describes smart manufacturing as a system which is fully-integrated and collabrative. Furthermore it can react in real time to the changing demands and conditions in a factory, supply network and customer needs [1]. Other definitions can be found in [2, 3, 4, 5]. They all share the main theme of smart manufacturing, which is the automation and connectivity in the manufacturing process to achieve higher quality and a shorter production time.

With an ever-increasing amount of data, a new technology has been established to realize these goals, which uses data-driven methods and algorithms in the field of smart manufacturing [6]. These methods mainly focus on one step of the smart manufacturing process, e.g. smart design, process planning, material distribution or manufacturing monitoring. The application of data-driven tools in the intersection between these different domains currently remains an issue for which the concept presented in this paper aims to provide a solution. The goal is to use the geometry design data to support the subsequent processes, especially the process planning. A concept for a similarity search method is stated in this paper, which is capable of

* Corresponding author. Tel.: +49-(0)9131-85-23658 ; fax: +49-(0)9131-85-23223. e-mail address: bickel@mfk.fau.de

2212-8271 © 2021 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2021.01.065

finding similar, already manufactured components on the basis of the initial geometry and therefore their respective production data to support the new process.

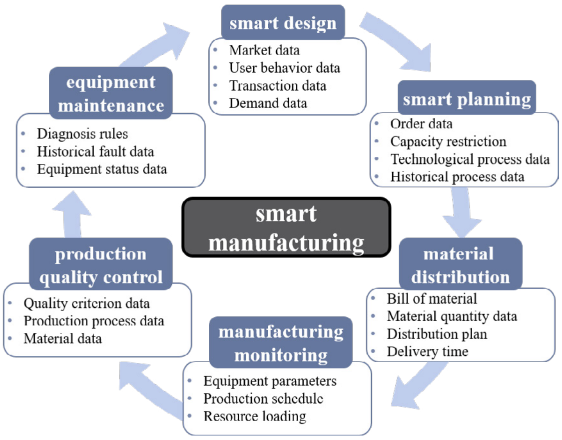


Fig. 1. Smart manufacturing cycle according to [6]

## 2. Methodical background

### 2.1. Similarity comparison

The idea of similarity search is to find the already constructed 3D-object that is most similar to a given search input. This task is achieved by converting a 3D geometry into a unique signature, which is then compared to various other signatures in order to find the one most similar to the initial search query. Similarity search is in general interdisciplinary since it is applied in many different areas such as computer vision, CAD data, computer graphics or multimedia. Therefore, the calculation process of this unique shape descriptor is based on different principles. Frequently deployed methods are based on: graphs, features, images, histograms or grouping technology (GT). This categorization is related to the work of [7, 8, 9, 10].

Graph based algorithms extract the necessary information about the geometry of an object from the connections of different shape components. Examples for this technique can be found in [11] and [12]. Other methods rely on features that are calculated from the individual shape. These features can be applied to the entire component or local areas only. Simple examples of features are the volume or the Fourier transformation of the shape boundary. A collection of methods is listed in [8] and [10]. Another way to compare geometries is to use images. This method offers the advantage that not only 3D geometry can be compared with each other but also a search for 3D models can be carried out on the basis of 2D drawings. The detailed explanation of these methods can be found in [13, 14]. Histogram based methods describe different geometric features through histograms. An example for this technique is the AD (angle and distance) histogram, according to [15].

The similarity search using the group technology has its roots in manufacturing technology. GT-codes are used for organizing production materials according to their part family. Popular and well established code-systems are the Opitz code [16], the DCLASS code and the KK3 system [17]. These codes are computed automatically from the geometry and are then used by the similarity search system. An example for an approach based on the Opitz code is stated in [7] and for the DCLASS code in [18]. Because of the large number of possible applications, many of the additional methods use different types of signature generation, such as text search, shape slicing or deformations. More information can be found in [7, 8, 9, 10].

However, not only the type of signature generation is distinguished, also the area of application to the geometry can be used as a distinctive factor. If the whole geometry is compared with another total geometry, it is considered to be a global similarity search. The search for partial areas of a geometry in an overall geometry is called partial shape retrieval and is based on the idea of each geometry being composed of several subgeometries. If the human body is referred to as a whole part, the hands are an example for a partial geometry. However, one hand can further be divided into various fingers. This exact definition of subcategories highlights the problem of partial shape retrieval. [19]

The specific challenges of the manufacturing process require a combination of these two methods. First, the system must find the global similarity between the components since the global shape has the greatest influence on the manufacturing parameters. Second, detailed areas of the geometry influence specific manufacturing parameters. These geometry areas have to be defined by the responsible process engineer who holds the necessary domain know-how.

This publication represents a new approach using global as well as local similarities to support the process engineer in finding similar parts. This procedure uses different methods of machine learning and a novel way of transferring geometry into a shape signature to achieve the stated goal.

### 2.2. Machine learning methods

The research field of machine learning is dedicated to the development and automation of computer-based learning processes. One of the first mentions of this topics explains machine learning as the ability of computers to learn without being explicitly programmed [20]. Many of the subsequent descriptions of this field can be summarized with these properties [21, 22, 23]: computer learning, automated processes, pattern recognition and adaptation to new data.

Machine learning methods can generally be divided by various characteristics. The kind of training is often used to seperate different learning methods among themselves. According to [24] these are: supervised learning, unsupervised learning, semi-monitored learning and active learning.

#### 2.2.1. Classification

Classification is a typical method of supervised learning because the different classes must be defined before the training process itself begins. The goal of classification is to assign data to previously defined classes. Due to the wide field of application a multitude of procedures and methods for carrying out a classification exists. According to [25, 26], examples for known classification methods are artificial neural networks, k-nearest neighbors, decision trees, support vector machines or naive bayes classifiers.

*2.2.2. Clustering*

The classification problem is closely related to cluster analysis, which combines values into a number of groups or classes unknown at the beginning of the analysis. Clustering is therefore an unsupervised learning method. During cluster analysis, groups or clusters are identified in the data. In order to carry out such an analysis, a problem definition must be given. This can, for example, be the division of customers according to target groups in order to improve the effectiveness of the advertising. There are many methods for carrying out cluster analysis. In [25, 27] mentioned example methods are: the k-means algorithm, the fuzzy variant or self organizing maps.

## 3. General approach

In this section the general approach is presented and explained in more detail. The individual modules of the overall system are: the projection method, the similarity comparison and the shape segmentation. These three methods are described in sequence, starting with a general overview of the procedure.

### *3.1. Overview*

To realize the idea of differentiation between local and global geometric similarity, the procedure must have a two-step structure. Therefore, the approach is modular and contains as main components: the similarity comparison based on the projection method, the clustering of results and the shape segmentation for further investigation. An overview of the whole process is depicted in Figure 2.

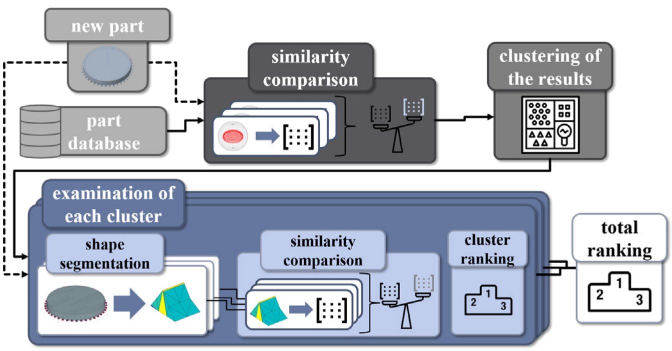


Fig. 2. Overall procedure of the similarity search

At first a new search component is compared with the available database. For this purpose the part must be transformed into a shape descriptor. To start this process, the geometry needs to be randomly sampled with points on the surface. This point cloud is then projected onto a detector sphere, which is transformed into a matrix. Through this method, any geometry can be converted into a matrix which is more clarified in section 3.2. In the next step the search-matrix is compared to each database-matrix via the correlation coefficient. The calculated scalar value can then be used for clustering the different results. This leads to parts being sorted in various clusters, respective to their correlation coefficient.

Subsequently, the clusters are examined individually and ranked internally based on their local geometric similarity. This similarity value is calculated with the special geometric features of the searched part, which has been defined by the user. The component is automatically separated by the system into different segments and the relevant geometry parts are again compared for similarity. This analysis employs nearly the same method for the comparison but instead of projecting the part from the center onto the detector sphere, the local similarity value uses the normal of each point as projection direction. The two types of projection are presented in more depth in the following section.

At the end of the process each local cluster is ranked and finally linked to the clusters of global similarity clusters. In the last step an overall ranking is determined and the corresponding components can be displayed to the employee.

### *3.2. Projection method*

The procedure for transforming 3D geometry to a matrix is based on the projection of points onto a surface. The projection method is based on the work of [28], where the process classifies finite element simulations. For the similarity comparison the projection receiver surface is a segmented sphere, as seen in Figure 3. Theses segments are called pixels and are spanned in a grid pattern of 224x224 around the sphere. They count the points of intersection with the projected lines and their final numbers are equal to the value of the corresponding matrix. The way theses point nodes are projected onto the sphere, de-

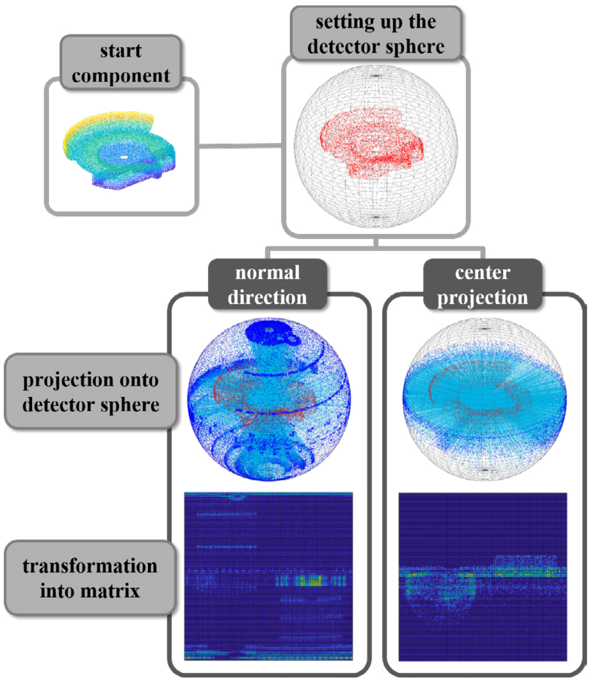


Fig. 3. Overview of the two different projection methods

pends on the similarity measure. A comparison of the two ways of projection is shown in Figure 3. For global similarity, the points are projected from the center of the object through each point onto the detector sphere. In contrast to the local similarity, the points are projected in normal direction. These directional values are calculated from the neighboring points. The difference between the two methods can be seen in the illustration. It has been observed that the projection from the center represents the general shape of a component (round, flat, cubic) better than the normal projection. This method offers advantages for smaller elements of a geometry since more information can be transferred. It should also be noted that the normal projection is more prone to errors due to the calculation of the normal direction, where no information about the inside or outside of the geometry can be determined.

After the points are successfully counted through every pixel, the sphere is converted into a matrix. This process is similar to the transition from a globe to a map.

### 3.3. Similarity comparison

After the geometry has been successfully transformed into a matrix, the next step is to determine the similarity of the matrices. The result of the comparison is a scalar value which serves as the basis for the clustering and ranking. To compare the similarity of two matrices, the 2D correlation coefficient is used. This metric can be applied for comparing images and matrices. The result is always a scalar number between -1 and 1. This value represents the similarity of the two matrices. When forming clusters, the search for groups starts at the best result. If the next component is within -0.01 of the calculated value, both components are combined into a cluster. If this is not the case, a new cluster is formed and again a comparison is conducted based on the value. These groups are then examined in more depth with shape segmentation, which is the preliminary work for the local similarity comparison.

### 3.4. Shape segmentation

In this chapter the process of shape segmentation, which is necessary for the determination of the local similarity, is presented in greater depth. The procedure is based on the work of [29], which uses supervised learning algorithms to divide the input geometry. This feature is necessary for the implementation of the method since the geometry elements have to be specified by the user before they can be examined in more detail. This kind of input can only be reproduced by supervised learning methods. An adapted process of the method established by [29] can be seen in Figure 4. The changes made include the adjustment of the feature vector generation and the application of different machine learning classification algorithms. A range of machine learning algorithms have been tested and the two most suitable algorithms for this task were neural networks and decision trees.

After the geometry has been successfully classified, it must be clustered to recognize identical elements. For this purpose, a cluster method is used that does not require a specific group

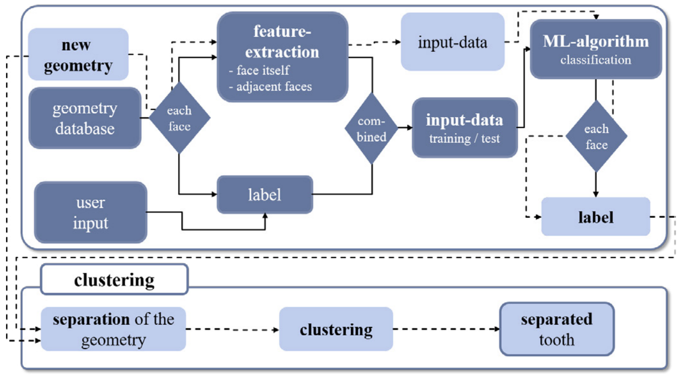


Fig. 4. Process of supervised shape segmentation based on [29]

quantity since this is unknown at the start of the search query. The db-scan algorithm (density-based spatial clustering of applications with noise) is applied to this clustering task because it is able to cluster data into an unknown number of groups and to detect noise in the input data. This last feature separates the db-scan method from many other clustering algorithms (e.g. k-means) and is important when the classification result is not completely perfect. In this case, the clustering method is able to distinguish between the noise (bad classified faces) and the good results (correctly labeled faces).

The outcome of the shape segmentation with the training and unknown test-dataset including the clustering of the separated geometry is displayed in Figure 5. The training dataset consists of five different geometries (Fig. 6 part *a-d, f*) and the test dataset contains one unknown geometry and one scaled version of a part (Fig. 6 part *e* and *c*). For the classification a neural network with two hiddenlayers (layout: 543-200-50-1) was trained and for clustering the db-scan algorithm was applied. The exact parameters and results for this analysis are listed in table 1.

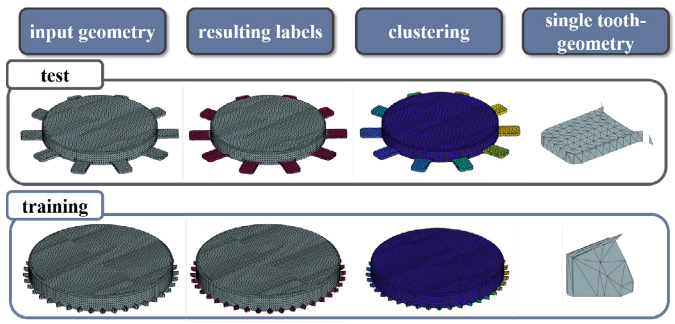


Fig. 5. Result of the different steps in the shape segmentation process

Table 1. Classification parameters and results for training and test dataset

| parameters | training | test |
|---|---|---|
| dataset | 146132 x 543 | 48114 x 543 |
| label (No. label 1 / No. label 2) | 120831 / 25301 | 37337 / 10777 |
| accuracy | 100,00 % | 94,80 % |

In this illustration the input geometry is shown on the left side, next to the segmented geometry. The colors gray and red visually separate the geometry by their corresponding labels. On the right side of the segmented part the outcome of the clustering can be seen. Each cluster group is represented by a different colour, resulting in the rainbow pattern for the tooth geometries. The last picture displays the final result of the shape segmentation: the isolated shape of a tooth. This geometry can now be utilized as a starting point for a local similarity analysis by applying the projection method in the normal direction, as explained in section 3.2.

## 4. Case Study

The previously described methods will be examined in this section with the help of a case study. Before the analysis can be started, a data set must be determined. This is composed of six different sheet-bulk metal formed parts, each of which differs in the shape of its teeth. Depending on their basic tooth geometry the components can be divided into three groups. All parts and their corresponding groups are shown in Figure 6.

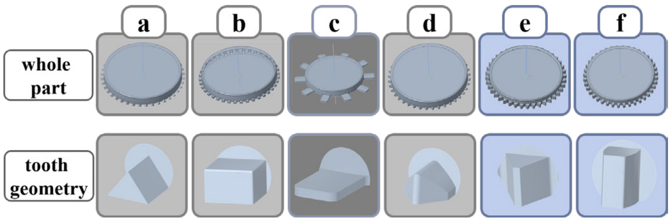


Fig. 6. Sheet-bulk metal formed parts and their tooth geometry for the case study

This investigation aims to represent these groups, implying that the first results of the similarity search must be from the same group as the search component. These parts can either be available as results in the same cluster or in clusters next to each other. Other parts should not be placed between components from equal groups.

The corresponding similarity values for the center projection method are shown in Figure 7. On the y-axis of the table, the search query part is listed, and on the x-axis all parts with their according result are shown. Therefore, on the diagonal all results are *1,000* because every part matches itself perfectly.

| part number | a | b | c | d | e | f |
|---|---|---|---|---|---|---|
| a | 1,0000 | 0,8817 | 0,6558 | 0,9074 | 0,8541 | 0,8409 |
| b | 0,8817 | 1,0000 | 0,6722 | 0,8863 | 0,8574 | 0,852 |
| c | 0,6558 | 0,6722 | 1,0000 | 0,6622 | 0,6485 | 0,6417 |
| d | 0,9074 | 0,8863 | 0,6622 | 1,0000 | 0,8570 | 0,8409 |
| e | 0,8541 | 0,8574 | 0,6485 | 0,8570 | 1,0000 | 0,8999 |
| f | 0,8409 | 0,8520 | 0,6417 | 0,8409 | 0,8999 | 1,0000 |

legende

similarity ranking

1. 2. 3. 4. 5. self

Fig. 7. Results of the similarity comparison with the case study parts

All result values have been clustered and colored according to the color scheme in the legend. Based on this analysis it can be concluded that the similarity search with the projection methods worked very well for this data set. In all search queries the components from the own class were first displayed as results. Furthermore part *c* is always the least similar component. An ideal example is the result for the geometry *e*. Here the most similar component is from the same group as the search part and the following three are all in one cluster. The work piece *c* is the most dissimilar compared to the search query.

An exemplary search run for the entire system is shown in the Figure 8. The illustration shows the search query for geometry *b* in more detail. First the global similarity comparison is carried out and afterwards the clusters are generated. In this case, three clusters are generated: the first for geometry *a* and *d*, the second cluster for geometry *e* and *f* and the last cluster with only geometry *c*. Afterwards, each cluster is analyzed by itself. The teeth are separated from the main body and compared with the other extracted teeth using decision trees as a classifier. This leads to a new ranking within each cluster. The procedure is carried out for each remaining cluster and an overall ranking is generated. In this particular analysis, geometry *a* is the most similar, followed by geometry *d*, *f*, *e* and *c*.

This result can then be used by the process engineer to apply the linked information from geometry *a* to the new, searched geometry. Especially for sheet-bulk metal formed parts, the information gain can be data for prototype production, e.g. the forming pressure. Furthermore, the process engineer can be provided with evaluation metrics for the process simulations, such as contact ratio or total equivalent plastic strain.

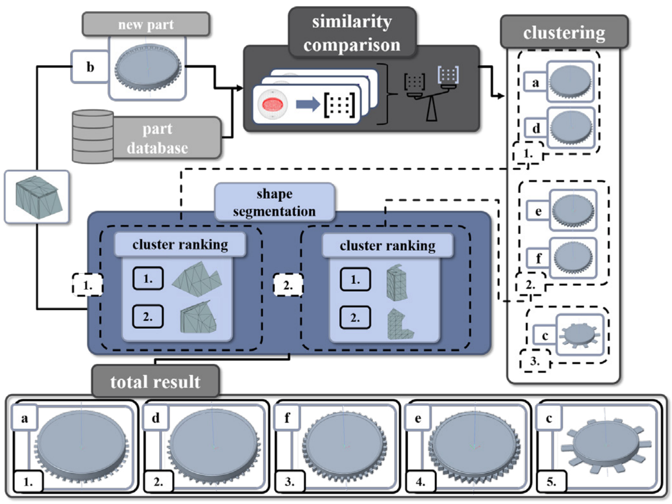


Fig. 8. Example run for part *b*

## 5. Conclusion

In summary, this paper shows a new approach for applying similarity search in the context of smart manufacturing. To adapt the search as accurately as possible to the application, the process uses a two step approach. First, searching for global similarity and subsequently splitting the geometry according to the user input. Then this separated geometry is again compared for similarity but only on the local level. For the similarity comparisons a new procedure is shown, based on the projection method that is able to adapt by changing the way of projection. Finally, the potential of this method was illustrated through applying it to a case study with sheet-bulk metal formed parts. These parts differ in their tooth geometry, a characteristic that was used for the segmentation of the geoemtry. To end the case study, an exemplary run-through was executed, including all intermediate steps. A possible boundary in the application could be the definition of the local geometry feature, which needs to be known and must have enough faces to train a machine learning model.

The field of flexible mass production can also benefit greatly from the similarity search because of the high output of different geometries. Therefore, for a new component variant the corresponding older, validated parts can be found automatically, including their manufacturing data. This results in faster production times and higher overall efficiency.

## 6. Outlook

For further research, many aspects could be considered. At first different machine learning algorithm should be taken into the analysis. The opportunities for newly available, powerful classifiers are increasing rapidly, so further investigations should take these methods into account. Especially ensemble learners should be observed for the classification task since decision trees show good results, which should increase by applying ensemble methods.

Finally, different similarity comparison algorithms for the matrices should be analyzed. Other ways of comparing matrices are available, for example peak signal-to-noise ratio (psnr) or structural similarity index (ssim) for measuring image quality. Perhaps these metrics can be combined with the correlation coefficient to result in an even better retrieval outcome.

## Acknowledgements

This research work is part of the FAU Advanced Analytics for Production Optimization project (E|ASY-Opt) and funded by the Bavarian program for the Investment for growth and jobs objective finance by the European Regional Development Fund (ERDF), 2014-2020. It is managed by the Bavarian Ministry of Economic Affairs and Media, Energy and Technology. The authors are responsible for the content of this publication.

We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan V GPU used for this research.

## References

[1] National Institute of Standard and Technology., 2014. Smart Manufacturing Operations Planning and Control.
[2] Wang, S., Wan, J., Li, D., and Zhang, C., 2015. Implementing Smart Factory of Industrie 4.0: An Outlook. International Journal of Distributed Sensor Networks
[3] Lucke, D., Constantinescu, C., and Westkaemper, E., 2008. Smart Factory: A Step towards the Next Generation of Manufacturing. Manufacturing Systems and Technologies for the New Frontier, pp. 115-118
[4] Brettel, M., Friederichsen, N., Keller, M., Rosenberg, M., 2014. How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. International Journal of Science, Engineering and Technology, Vol. 8, No. 1, pp. 37-44
[5] Zuehlke, D., 2010. Smartfactory-towards a Factory-of-Things. Annual Reviews in Control, Vol. 34, No. 1, pp. 129-138
[6] Tao, F., Qi, Q., Liu, A., Kusiak, A., 2018. Data-driven smart manufacturing. In: Journal of Manufacturing Systems 48, S. 114
[7] Zehtaban, L., Elazhary, O., Roller, D., 2016. A framework for similarity recognition of CAD models. In: Journal of Computational Design and Engineering 3 (3), S. 274285
[8] Tangelder, J.W.H., Veltkamp, R.C., 2008. A survey of content based 3D shape retrieval methods. Multimed Tools Appl 39, 441
[9] Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. In: Pattern Recognition 37 (1), S. 119
[10] Zheng, Q., Ji, J., Jun, Q., 2008. Content based 3D model retrieval: A survey, International Workshop on Content-Based Multimedia Indexing, London, pp. 249-256
[11] Mehalawia, M., Miller, R.A., 2001. A database system of mechanical components based on geometric and topological similarity. Part I: representation. In: Computer-Aided Design
[12] McWherter, D., Peabody, M., Regli, W.C., Shokoufandeh, A., 2001. Transformation invariant shape similarity comparison of solid models. In DETC01
[13] Chen, D., Xiao-Pei, T., Shen, Y.T., Ouhyoung, M., 2003. On Visual Similarity Based 3D Model Retrieval. Computer Graphics Forum 22: 223-232.
[14] Qin, F., Gao, S., Yang, X., Bai, J., Zhao, Q., 2017. A sketch-based semantic retrieval approach for 3D CAD models. Applied Mathematics-A Journal of Chinese Universities 32, 2752
[15] Ohbuchi R., Minamitani T., Takei T., 2003. Shape-similarity search of 3D models by using enhanced shape functions. In Theory and Practice of Computer Graphics p. 97-104.
[16] Opitz, H. 1970. A Classification System to Describe Workpieces. New York: Pergamon Press; Translated by Acton Taylor
[17] Salvendy, G. 2001. Handbook of Industrial Engineering Technology and Operations Management, 18. New York, John Wiley Sons, Inc
[18] Henderson, M., Musti, S., 1988. Automated group technology part coding from a three-dimensional CAD database. Journal of Manufacturing Science and Engineering, Transactions of the ASME, 110(3), 278-287
[19] Liu, Z., Bu, S., Zhou, K., Gao, S., Han, J., Wu, J., 2013. A Survey on Partial Retrieval of 3D Shapes. Journal of Computer Science and Technology 28, pp. 836-851
[20] Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. In: IBM Journal of Research and Development, Vol. 3, No. 3
[21] Michalski, R.S., Kodratoff, Y., 1990. Research in machine learning; recent progress, classication of methods, and future directions. In Machine Learning: An Articial Intelligence Approach, volume III. Morgan Kaufmann
[22] Carbonell, J.G., Michalski, R.S., Mitchell, T.M., 1983. An overview of machine learning. In Machine Learning: An Articial Intelligence Approach. Palo Alto: Tioga Publishing Company
[23] Kelleher, J., Tierney, B., 2018. Data Science. Boston: The MIT Press Essential Knowledge Series
[24] Han, J., Kamber, M., Pei, J., 2012. Data Mining. Concepts and Techniques. 3. Aufl. Waltham: Morgen Kaufmann
[25] Runkler, T., 2010, Data-Mining. Methoden und Algorithmen intelligenter Datenanalyse, 1. Aufl. Wiesbaden: Vieweg + Teubner
[26] Tan, P.-N., Steinbach, M., Kumar, V., 2010. Introduction to Data Mining. Boston: Pearson/ Addison-Wesley
[27] Petersohn, H., 2005. Data mining. Verfahren, Prozesse, Anwendungsarchitektur. 1. Aufl. Munich: Oldenbourg
[28] Spruegel, T.C., Schroeppel, T., Wartzack, S., 2017. Generic Approach to Plausibility Checks for Structural Mechanics with Deep Learning. In: Proceedings of the ICED17, Design Society
[29] Kalogerakis, E., Hertzmann, A., Singh, K., 2010. Learning 3D Mesh Segmentation and Labeling, ACM Transactions on Graphics, Vol. 29, No. 3