Retrieval of solid models based on assembly similarity

Rahul Renu and Gregory Mocko

Clemson University, USA

ABSTRACT

The research objective is to (re)use geometric part information to mine databases for process design information. Specifically, this research investigates an approach to use assembly solid model similarity to mine databases for assembly work instructions from the perspective of an automotive original equipment manufacturer. Results from this research will allow generation of assembly work instructions based on input of a solid model. This research presents an approach to determine similarity between query assembly solid model and database assembly solid models. In this method, similarity scores for the overall assembly solid model are used in conjunction with the similarity score of individual components. These similarity scores are obtained by computing histogram-based similarity scores, surface area differences and tessellation area distribution differences. A multi-index sort of the computed values for each (query-database) pair of assembly solid models results in a list of similar assembly solid models based on a query assembly solid model.

In addition, human designers were asked to identify similar geometric models in a controlled study. The results from the human study and the similarity algorithm are compared. It is found that using assembly model similarity, in conjunction with component model similarity, yields better correlation to survey results, as compared to the correlation between assembly model similarity alone and survey. Testing also shows that the proposed method has a statistically significant better correlation to survey results than traditional histogram based similarity approach. The approach to determine similarity of assembly solid models will be used to mine databases of solid models and retrieve their related assembly work instructions.

KEYWORDS

Solid model similarity; knowledge reuse; assembly work instructions; shape histogram; tessellation density

1. Frame of reference

This research attempts to help bridge the gap between design and production. With respect to this, the need to bridge the gap between design and production is first discussed. This is followed by a review of the existing approaches to bridge part geometry and production considerations.

1.1. Need to bridge the gap between assembly and design

Authoring assembly work instructions is a time consuming and subjective process. These work instructions are used during assessment of work content, redesign of assembly lines, training of assembly line associates and quality audits. Therefore, it is essential to have a consistent and correct set of assembly work instructions for the same product, across different assembly locations. In terms of the automotive development process, determining assembly work instructions at an early stage translates to a more efficient planning process.

The automotive development process [28] can be divided into two phases: the Development Phase and the Production Phase.

The Development phase can be further categorized into two sub phases: Design, and Prototyping & Evaluation. The cost of implementing design changes is least in the Design and increases subsequently. Understanding the impact of Design decisions on the Production phase can allow for early determination of downstream issues. This will result in proactive decision making as opposed to reactive decision making and ensure time- and cost-savings.

The Production phase can be further classified into two sub phases: Manufacturing; and Assembly. Manufacturing is where raw material is converted to the desired component. Assembly is where manufactured components are assembled together. Production specifications estimated in prior stages are modified and made more specific. The feedback loop from Production Phase to Development Phase needs to be closed. Knowledge of decisions made during the Automotive Development Process has to be reused during variant and new product
development. This is critical because the cost of implementing design changes increases ten-fold through every passing stage. Key outcomes of the Development phase and the Production Phase are solid models (part geometries) and assembly work instructions, respectively. It is recognized that product structure will have an impact on the subassemblies that an enterprise must assemble. However, original equipment manufacturers have adopted the concept of product modularity, and with this come similar product structures across various product variants. Linking these two elements will help in closing the feedback loop from manufacturing to design.

1.2. Existing approaches that link part geometry and assembly process considerations

Boothroyd and Dewhurst [3] have presented research to allow for consideration of manufacturing and assembly perspectives in the product design process. The Boothroyd and Dewhurst method [3] is a systematic method to review design of parts and the objective of this review is to reduce assembly time and cost. The Boothroyd and Dewhurst method [3] is based on a set of queries that are presented to the user in terms of tables. Navigating these tables requires the designer to be cognizant of part characteristics. These characteristics include: ease of handling, number of components, number of assembly directions and ease of inserting; some of which are subjective in nature. While this method is primarily used to analyze existing designs, other researchers have built upon this method and integrated it within the top-down design process. One such effort is from Warnecke and Bassler [27], who present the Assembly-Oriented Design (AOD) method to include consideration of assembly processes during the design process [21]. Their method aims to provide designers with perspectives of assembly process early in the design process to ensure that the number of iterations in the design process reduces. The AOD method has also been researched and adapted by DaimlerChrysler AG [26] in the automotive industry.

Zha and colleagues [29,30] build upon the AOD method and provide a knowledge-based approach to increase the level of automation within each stage of the AOD method. Zha and colleagues recognize that there are three levels at which assembly process perspectives need to be included: component level, product level and assembly-process level. At each level, the focus of the AOD method is to provide designers with guidelines for designing products that are less time consuming and less expensive. In the knowledge-based approach presented by Zha and colleagues [29,30], experienced design engineers and production experts explain complex concepts (pertaining to design geometries and assembly processes) to knowledge engineers. These knowledge engineers then generalize and codify the concepts, allowing for formalization of knowledge capture and reuse. This knowledge is supplemented with knowledge obtained from handbooks [3]. This method is subjective and depends on the ability of experts to comprehensively recall necessary information. It also depends on the ability of experts to explicate their knowledge.

It is important to note that consideration of design for assembly rules alone may not be efficient. Assembly is not the only downstream perspective that must be considered during design. Issues related to manufacturing, disassembly, environment and sustainability must also be considered. Studies have shown that use of different design for ‘X’ rules can provide contradictory opinions [17,19,23,24]. Kuo and colleagues [17] have presented a summary of several design for ‘X’ rules to help designers integrate all considerations. Design for assembly rules [3] are generic and lack enterprise-specific perspective. Design for assembly rules as an approach to link design and assembly has been used extensively and issues with this approach have been identified. This research aims to link solid models to assembly process descriptions by using a data mining approach. This research, specifically investigates reuse of solid model information and related assembly process knowledge. When engineers design new components, the solid model of these components can be assessed for similarity with respect existing component solid models. Based on the similarity of solid models, existing (related) process descriptions can be presented to engineers while they design the assembly process.

In particular, this research focuses on developing a method to link product design to assembly process design. This research explores the use of solid model similarity and text analysis approaches to develop a relationship between solid models and assembly work instructions.

The following three objectives are identified to formalize a linkage between assembly processes and solid model geometry. The research presented here focuses on the first objective.

1. Evaluate solid models for similarity in terms of their assembly processes
2. Investigate the natural language processing approaches required to analyze assembly work instructions
3. Use part geometry information to mine database of assembly work instructions and retrieve relevant work instructions

The need for a system that allows forecasting of assembly process information to the product development
phase is required. This will allow for cost- and time-efficient decisions regarding assembly processes to be made in the Development Phase. In this research, a method to evaluate solid model similarity of assemblies is presented. It is proposed that this method will be used to mine databases and retrieve assembly process information based on solid model similarity.

1.3. Approaches to determine similarity among solid models

Approaches to retrieve solid models from a database of models have been researched extensively [2,5–7,9–16,20,22]. Similarities of shapes have also been investigated from the perspective of image processing in multimedia applications [5,9,22]. The various approaches of shape retrieval have been reviewed by Bustos and colleagues [5]. In the literature reviewed by us, there have been minimal efforts that determine similarity of assembly models by considering individual component similarity in conjunction with assembly model similarity.

The shape retrieval method proposed by Osada and colleagues [20] uses random sampling of points on the two parts being compared and generates a histogram of the distribution of distances between the random sampling points. The generated histograms (one for each part being compared) are compared using Minkowski difference [11,20] (see Fig. 1).

![Figure 1. Illustration of histogram-based similarity computation.](image)

Ip and colleagues [11] have recognized that this approach of using shape histograms for comparison, provides a similarity at a coarse level-of-detail. The same findings have also been presented by Jayanti and colleagues [14]. Retrieval of similar shapes using the Osada histogram approach [20], may yield false negatives in a mechanical engineering setting [14]. This can be attributed to the fact that the histogram approach does not consider functional purpose of parts [14]. This method (histogram) is not viable if exact differences between two models are to be obtained. However, the advantages of this approach are the following:

1. Independent of modeling history
2. Sensitive to size
3. Can be applied to software-independent file formats (such as STL, STEP and IGES [4,8])
4. Independent of orientation of the part in the local coordinate system
5. Shape distributions of database parts can be stored, therefore offloading computation and providing for a run-time, computationally inexpensive shape retrieval method.

Histogram-based similarity approach has proven to be a successful method of solid model retrieval [6,10,11,20]. From an assembly process perspective, it is hypothesized that the histogram approach can be used to generate clusters of similar components. Additional techniques will be required to generate a ranked list of similar parts within each cluster and also to retrieve and consolidate assembly work instructions based on solid model similarity. This research will test this hypothesis.

2. Aims and significance

The approach developed in this research allows for determination of similarity of assembly solid models. The approach proposed in this research uses similarity of component models and their assembly model to compute similarity. This is tested against approaches that use components alone, and also against approaches that use assembly models alone. Assembly solid model similarity information will be used to retrieve assembly work instructions for reuse. This research supplements previous research that attempts to bridge the gap between design and production.

3. Approach to determine solid model similarity

The objective of determining similarity of assembly solid models presented is to retrieve assembly work instruction sets. 181 work instruction sets from an automotive Original Equipment Manufacturer are analyzed. It is found that there are, on average, four parts per work instruction set (p-value < 0.01). Therefore, the goal of this research will be to design a system that computes similarity of assembly solid models that contain at most four components. Fig. 2 shows an example of an assembly model with two components.

Determining solid model similarity is divided into four stages. In the first stage, Osada and colleagues’ [20] method is used to generate histogram-based similarity scores for all database solid models, with respect to query solid model. In the second stage, histogram-based similarity scores are used to generate clusters of similar solid models. In the third stage, solid models within each
cluster are analyzed for surface area differences and tessellation area distribution differences. In the final stage, a multi-index sort is performed to generate a ranked list of similar solid models (see Fig. 3).

Stage 1: Generating histogram-based similarity scores for solid models [adapted from [20]]:

This algorithm generates shape histograms for each of the two assembly solid models being compared. Next, the L1 Minkowski distance between the two shape histograms is calculated and this indicates the geometric similarity of the two parts. The detailed algorithm is presented below.

1. For every part (STL format)
   a. For every triangle
      i. Calculate area of each triangle
      ii. Store [area, cumulative area]
   b. Generate random number between 0 and total cumulative area
   c. Find corresponding triangle from [area, cumulative area]
   d. Generate random point on this triangle using the following formula:
   iii. \[ P = (1 - \sqrt{r_1}) A + r_1(1 - \sqrt{r_2})B + \sqrt{(r_1)(r_2)}C \]
   e. Repeat the previous three steps 1024 times
   f. Compute Euclidean pairwise distances for all generated points
   g. Generate histogram of distances
   h. Compute L1 Minkowski difference between the two histograms

2. Compute standard deviation of all scores
3. Normalize all scores over the standard deviation (O.S.)

Stage 2: Generating clusters of similar solid models

Histogram-based similarity provides similarity of overall shapes of solid models. This information is used to generate clusters of similar solid models. The solid models within clusters can then be investigated further for similarity at a level of finer resolution. The clusters are generated by creating five bins. The bin widths are determined by dividing the histogram-based score in fifths.

Stage 3: Computing surface area and tessellation area distribution differences

The difference between STL files is essentially the difference between the tessellations that the files are composed of (see Fig. 4). This forms the basis for the following pseudo-code that is used to recognize specific differences between CAD (STL) models within clusters from the previous stage:

1. For each STL file
   a. For each tessellation
      i. Calculate surface area
      ii. Store vector of tessellation coordinates and surface area

Figure 4. Illustration of tessellation differences in different solid models.
Table 1. Example of computing similarity score.

<table>
<thead>
<tr>
<th>Name</th>
<th>Cluster Number</th>
<th>Difference in surface area</th>
<th>N.T.</th>
<th>L.M.A.</th>
<th>O.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>15</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>A.1</td>
<td>1</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>A.2</td>
<td>3</td>
<td>12</td>
<td>6</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>37</td>
<td>18</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>22</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>B.1</td>
<td>2</td>
<td>35</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>B.2</td>
<td>5</td>
<td>56</td>
<td>12</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>63</td>
<td>30</td>
<td>21</td>
<td>42</td>
</tr>
</tbody>
</table>

2. Count the number of tessellations (N.T.) that are found in the complement of the intersection of the two surface area arrays
3. Generate histogram of all tessellation’s area
4. Compute L1 Minkowski difference of the tessellation areas (L.M.A)

Stage 4: Multi-index sort to generate ranked list of similar solid models

A multi-index sort is performed based on the following parameters, in the listed order:

1. Cluster number
2. Difference in surface area
3. N.T. – Number of unique tessellations
4. L.M.A. – L1 Minkowski score of tessellation areas
5. O.S. – Histogram-based similarity score [20]

This results in a ranked list of similar solid models based on an input model. This approach can be used to analyze individual component’s solid model geometry as well as solid model geometry of their assemblies. Similarity scores of individual components are summed with the scores of their respective assemblies to obtain an overall score.

As an example, let us consider two assembly models, ‘A’ and ‘B’. Their component models are ‘A.1’ & ‘A.2’; and ‘B.1’ & ‘B.2’. Table 1 illustrates how the similarity score for ‘A’ and ‘B’ will be computed with respect to a query model.

Total score for ‘A’ and ‘B’ will be used to perform a multi-index sort. The result of this sort will be a ranked list of solid models that are similar to the query model.

4. Experiment design

This section presents testing results for solid model retrieval. First, a discussion of the solid models used in this study is presented along with details of the survey conducted. This is followed by comparison of user study ranking results to the rankings obtained from the algorithms.

4.1. Description of solid models

All solid models used in this research are STL ASCII files. Several component models are obtained from the C-Design Lab at Purdue University (Source: https://engineering.purdue.edu/cdesign/wp/). All STL file pairs that are to be compared for similarity, are created with the same resolution of tessellations. The dimensions for these models are obtained by opening these files in SolidWorks © and using millimeters as the default unit of measurement. In terms of the classification of solid models presented by Jayanti and colleagues [14], the solid models used in this research are distributed as follows:

- Solids of revolution: 60%
  - Bolt like parts: 18%
  - Cylindrical parts: 18%
  - Long pins: 12%
  - Spoked wheels: 12%
- Flat-thin wall components: 22%
  - Bracket like parts: 12%
  - Curved housings: 12%
- Rectangular-cubic prism: 18%
  - Small machined blocks: 9%
  - Thin plates: 9%

The complexity of these assembly model files were assessed by counting the number of tessellations and comparing them to the number of tessellations from primitive shapes (cone, cube, cylinder, pyramid, sphere, torus and wedge) [25]. The results from this comparison are presented in Tab. 2.

Table 2. Complexity of assembly models used in survey.

<table>
<thead>
<tr>
<th>Primitive shape</th>
<th>Number of tessellations for primitive shape</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyramid</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Wedge</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Cube</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Cone</td>
<td>216</td>
<td>0</td>
</tr>
<tr>
<td>Cylinder</td>
<td>396</td>
<td>0</td>
</tr>
<tr>
<td>Torus</td>
<td>5940</td>
<td>43</td>
</tr>
<tr>
<td>Sphere</td>
<td>7056</td>
<td>43</td>
</tr>
<tr>
<td>Average</td>
<td>1956</td>
<td>2.85</td>
</tr>
</tbody>
</table>

The last column in Tab. 2 represents the percent of assembly models used in the survey that have lesser number of tessellations (lesser complexity) than corresponding primitive shape (P). The last row shows 2.85% of survey assembly models have lesser complexity than the average complexity of primitive shapes.

4.2. Results for solid model retrieval

The survey was administered via a website and consisted of seven questions. Each question had five associated
options. Each group of question and options from this point on will be referred to as a test set. All questions and options consisted solid model diagrams of two components and their related assembly (see Fig. 5).

Participants were asked to rate each option based on its assembly process similarity to that of the question. The participants were asked to rate using the scale shown in Tab. 3.

4.3. Algorithm testing

Fleiss’ Kappa was computed for each question to check for inter-rater agreement (see Tab. 4). It is found that test set 2 showed slight disagreement among participants. Therefore results from test set 2 were not considered for further analysis. For all other questions, Fleiss’ Kappa indicated fair agreement [18] at least.

The ranked list of solid models, for each of the seven test sets from the survey was compared to the ranked list obtained from each algorithm from the following experiments:

Experiment 1. Histogram-based scores of assembly models only

<table>
<thead>
<tr>
<th>Test Set</th>
<th>Fleiss Kappa</th>
<th>Level of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0.3431</td>
<td>Fair</td>
</tr>
<tr>
<td>Q2</td>
<td>-0.0208</td>
<td>Slight disagreement</td>
</tr>
<tr>
<td>Q3</td>
<td>0.3721</td>
<td>Fair</td>
</tr>
<tr>
<td>Q4</td>
<td>0.6226</td>
<td>Substantial</td>
</tr>
<tr>
<td>Q5</td>
<td>0.2313</td>
<td>Fair</td>
</tr>
<tr>
<td>Q6</td>
<td>0.2627</td>
<td>Fair</td>
</tr>
<tr>
<td>Q7</td>
<td>0.2384</td>
<td>Fair</td>
</tr>
</tbody>
</table>

Table 3. Likert scale used in survey.

1. Not at all similar
2. Somewhat similar
3. Similar
4. Very similar
5. Identical

Table 4. Fleiss’ Kappa results from the survey.

Figure 5. Excerpt of survey.
Experiment 2. Histogram-based scores of component models
Experiment 3. Histogram-based scores of assembly models and their respective component models
Experiment 4. Proposed algorithm scores of assembly models alone
Experiment 5. Proposed algorithm scores of component models
Experiment 6. Proposed algorithm scores of assembly models and their respective component models

The results of these comparisons are presented in Tab. 5. The comparison shows that the proposed algorithm performs better than the shape histogram approach in all cases except when assembly models alone are used. Tests for statistical significance are performed with a level of significance of 0.01. The null and alternative hypotheses are presented below.

**Null Hypothesis:** The correlation coefficient is less than or equal to zero between the entities being compared.

**Alternative Hypothesis:** There exists a positive correlation between the two entities being compared.

The p-value obtained in Experiment 4, was greater than 0.01. In this case we fail to reject the null hypothesis. This implies that there is sufficient evidence to suggest that there is disagreement between the rankings obtained from the survey and the ranking obtained from the algorithm. All other p-values were found to be less than 0.01 and we can reject the null hypothesis in all these cases. This implies that in all other cases there is sufficient evidence to suggest that there is agreement between rankings. More specifically, these results show that ranked list from Experiment 6 (the proposed approach using component and assembly models to compute similarity) best correlate to human interpretation of assembly model similarity.

Other experiments were conducted to test the performance of the proposed solid model similarity algorithm (Experiment 6). The repeatability of the clustering algorithm was tested. The algorithm was run on 300 solid models and the clusters were formed. This was repeated nine times keeping all parameters the same. For each of the nine trials, the solid models grouped into the first cluster were analyzed for repeatability. The number of differences found in the first cluster in each pair of trials was calculated and complied into a matrix. The sparsity of the matrix is found to be 0.58. This shows that the repeatability of the clustering algorithm is acceptable. The difference can be attributed to the fact the clustering algorithm operates on histogram-based scores; and these scores are probabilistic in nature.

When STL files are created, users typically have the choice of controlling the resolution of tessellations. As the resolution of the STL files is made finer, the number of tessellations increases (see Fig. 6).

The sensitivity of the proposed algorithm to the resolution of the tessellations is tested. The goal of this test was to test the hypothesis that the causation for correlation is tessellations. Also, the test will indicate the sensitivity of correlation values to tessellations. Four levels of resolution for tessellations are chosen. The experiment is performed on one query model and five database models from the survey (test set 6). The query and database models are varied through all four levels of resolution. Kendall’s Tau rank correlation coefficient is used to check correlation between survey results and algorithm results. Results from this experiment are presented in Tab. 6.

Results from this experiment indicate that as the resolution of tessellations decreases, the correlation to survey results reduces. Therefore, it is vital that maximum

---

**Table 5.** Results of Kendall’s Tau correlation computation for survey results and results from each experiment.

<table>
<thead>
<tr>
<th></th>
<th>Assembly models only</th>
<th>Component models only</th>
<th>Component and Assembly models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shape histogram</strong></td>
<td>0.0667</td>
<td>0.1333</td>
<td>0.5333</td>
</tr>
<tr>
<td>(Experiment 1)</td>
<td>(Experiment 2)</td>
<td>(Experiment 3)</td>
<td></td>
</tr>
<tr>
<td><strong>Proposed Algorithm</strong></td>
<td>-0.1333</td>
<td>0.5333</td>
<td>0.7000</td>
</tr>
<tr>
<td>(Experiment 4)</td>
<td>(Experiment 5)</td>
<td>(Experiment 6)</td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 6.** Example of change in number of tessellations as resolution changes.
number of tessellations be used and resolution of these
tessellations must remain consistent across query and
database models.

4.4. Conclusions

From the tests conducted, the following conclusions can be drawn:

- Use of proposed approach (multi-index sort along with component and assembly model similarity) provides best correlation to survey results and will be used for assembly solid model retrieval.
- As the tessellation resolution moves from coarse to fine, the correlation to survey results increases. Therefore, fine resolution tessellation models will be used for assembly solid model retrieval.

5. Closure and future work

A method to retrieve assembly solid models based on individual component and overall assembly similarity is presented and validated. The algorithm developed in this research yields better results for retrieval of solid model similarity when compared to five other approaches. Testing also shows that the proposed method has a statistically significant better correlation to survey results than traditional histogram based similarity approach.

To test the causation for the correlation, the tessellation resolution of assembly solid models was varied. It is found that as the tessellation resolution reduces (number of tessellations reduces), the correlation to survey results reduces.

By analyzing 181 assembly work instruction sets from an automotive Original Equipment Manufacturer, it has been determined that there are, on average, four parts per work instruction set. Since the goal of this research is to predict assembly work instructions based on assembly model similarity, the proposed assembly solid model similarity algorithm must be tested on assembly models that consist of three and four parts. A survey is being designed to perform this testing. For this survey, in addition to solid models obtained from C-Design Lab at Purdue University (https://engineering.purdue.edu/cdesign/wp/), solid models from GrabCAD© (https://grabcad.com/) and 3DContentCentral© (http://www.3dcontentcentral.com/) are being used. Assembly solid models from a graduate-level Computer-Aided Design course (from Clemson University) will also be incorporated in this survey.

Since there does not exist an assembly solid model database and benchmark, assembly solid models are being collected for this purpose. Complexity of these assembly solid models is being assessed by counting number of tessellations [25]. Existing shape similarity algorithms [1,7,10,13,14] will be compared to the proposed algorithm using precision-recall curves. These curves can then be reused by other researchers to assess their assembly model similarity algorithms. Comparison to approaches that use feature-based similarity [7,12] is required. It is hypothesized that determining specific assembly process related features will not be a non-issue. Assembly processes not only depend on the interface between two components, but it also depends on other features of both components as they relate to handling of the components.

Once tested and refined, the approach presented here will be used to mine databases of assembly work instructions. Natural Language Processing tools will be used to analyze assembly work instructions retrieved. These assembly work instructions will be consolidated and presented to the user, resulting in a system that takes input of assembly solid models and provides an output of assembly work instructions.

ORCID
Rahul Renu http://orcid.org/0000-0002-7069-9545
Gregory Mocko http://orcid.org/0000-0001-8845-4581

References


[23] Rayate, V.: A tool for selection of design for manufacturing and assembly rules during product design stage while considering end of life conditions, Clemson University.


