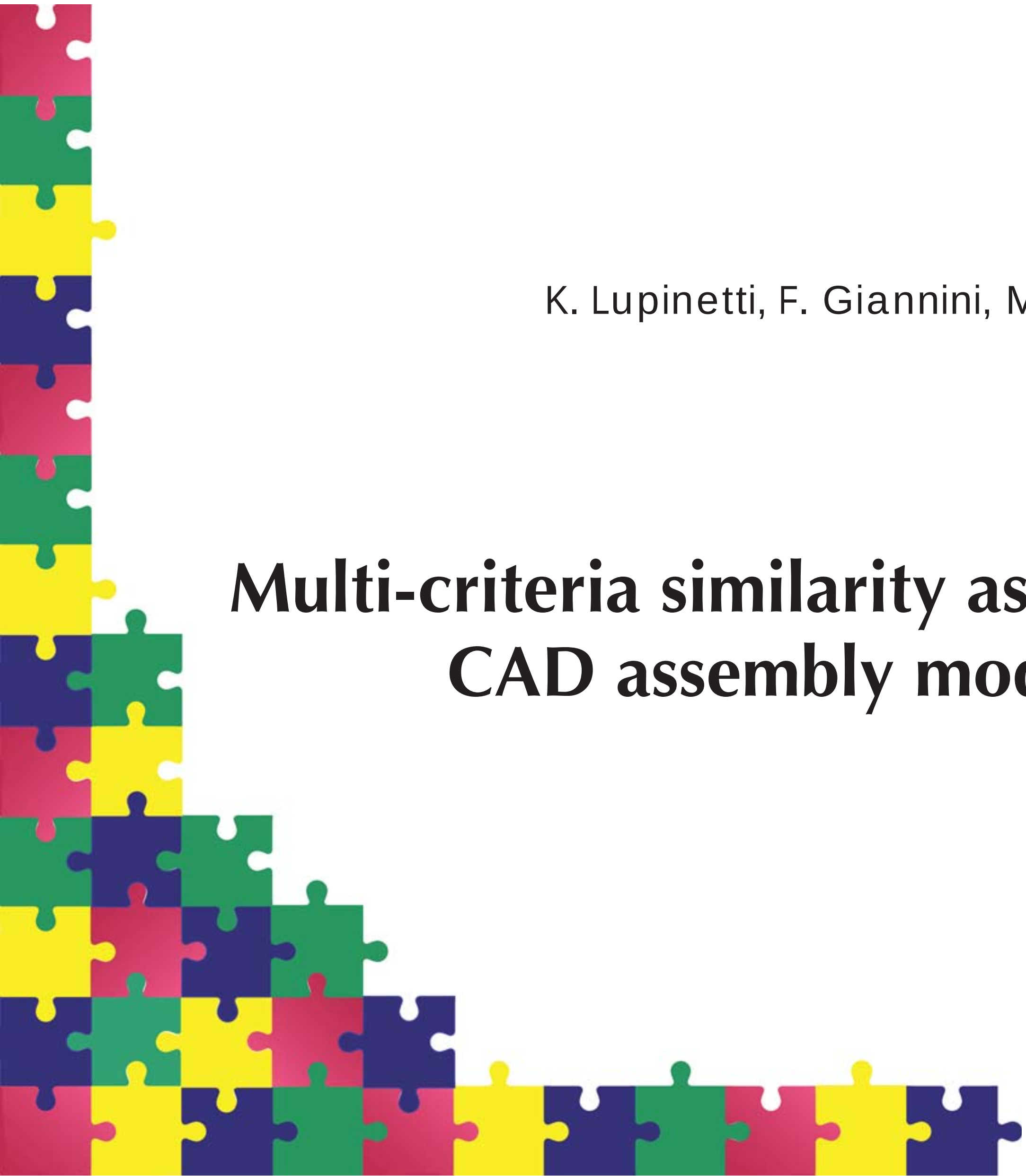
Consiglio Nazionale delle Ricerche
Istituto di Matematica Applicata
e Tecnologie Informatiche
“Enrico Magenes”

REPORT SERIES

K. Lupinetti, F. Giannini, M. Monti, J.-P. Pernot

Multi-criteria similarity assessment for CAD assembly models retrieval



18-07

IMATI REPORT Series

Nr. 18-07
August 2018

IMATI Report Series Nr. 18-07
August 9, 2018

Multi-criteria similarity assessment for CAD assembly models retrieval

Katia Lupinetti, Franca Giannini, Marina Monti, Jean-Philippe Pernot

Corresponding author:

Katia Lupinetti
Istituto di Matematica Applicata e Tecnologie Informatiche "*E. Magenes*"
Consiglio Nazionale delle Ricerche
Via de Marini, 6 (Torre di Francia) 16149 Genova - Italy
E-mail: katia.lupinetti@ge.imati.cnr.it

Franca Giannini
Istituto di Matematica Applicata e Tecnologie Informatiche "*E. Magenes*"
Consiglio Nazionale delle Ricerche
Via de Marini, 6 (Torre di Francia) 16149 Genova - Italy
E-mail: franca.giannini@ge.imati.cnr.it

Marina Monti
Istituto di Matematica Applicata e Tecnologie Informatiche "*E. Magenes*"
Consiglio Nazionale delle Ricerche
Via de Marini, 6 (Torre di Francia) 16149 Genova - Italy
E-mail: marina.monti@ge.imati.cnr.it

Jean-Philippe Pernot
LISPEN
Arts et Métiers, Campus d'Aix-en-Provence
2, cours des Arts et Métiers, 13617 AIX EN PROVENCE - France
E-mail: jean-philippe.pernot@ensam.eu

Abstract.

The use of Digital Mock-Up (DMU) has become mainstream to support the engineering activities all along the Product Design Process. Over the years, companies generated large databases containing digital models and documents related to their products. For instance, considering complex products, the DMU can be composed of several hundred thousand parts assembled together in assembly trees containing tens of sub-assemblies, and representing several terabytes of data. The ability to retrieve existing models is crucial for the competitiveness of companies, as it can help to leverage existing solutions, results and knowledge associated with previous products. To speed up the access to this large amount of reusable information, CAD models search approaches have been proposed, and notably the so-called content-based search techniques which help designers exploiting the implicit knowledge embedded in the models. As part of a system for the retrieval of CAD assembly models, this paper introduces a set of four measures to evaluate assembly similarities according to multiple criteria. These measures are combined to assess three different levels of similarity (local, partial and global). The local measure only considers the contribution of similar parts, while partial and global measures take also into account the number of similar parts compared to the total number of parts in the query and in the target model. Moreover, an ad-hoc visualization interface has been designed to clearly highlight the different similarities and to allow a fast identification of the target models. The validation of the proposed measures is discussed through several academic and industrial examples run on the developed prototype software.

Keywords: *Assembly similarity evaluation, Multiple similarity criteria, 3D assembly model retrieval, Partial and local similarity*

[page left intentionally blank]

Multi-criteria similarity assessment for CAD assembly models retrieval

Katia Lupinetti[a,b], Franca Giannini[a], Marina Monti[a], Jean-Philippe Pernot[b]

[a] *Istituto di Matematica Applicata e Tecnologie Informatiche ”Enrico Magenes”, CNR Via De Marini 6, 16149 Genova, Italy*
[b] *Arts et Métiers, LISPEN, HeSam, France*

ABSTRACT

The use of Digital Mock-Up (DMU) has become mainstream to support the engineering activities all along the Product Design Process. Over the years, companies generated large databases containing digital models and documents related to their products. For instance, considering complex products, the DMU can be composed of several hundred thousand parts assembled together in assembly trees containing tens of sub-assemblies, and representing several terabytes of data. The ability to retrieve existing models is crucial for the competitiveness of companies, as it can help to leverage existing solutions, results and knowledge associated with previous products. To speed up the access to this large amount of reusable information, CAD models search approaches have been proposed, and notably the so-called content-based search techniques which help designers exploiting the implicit knowledge embedded in the models. As part of a system for the retrieval of CAD assembly models, this paper introduces a set of four measures to evaluate assembly similarities according to multiple criteria. These measures are combined to assess three different levels of similarity (local, partial and global). The local measure only considers the contribution of similar parts, while partial and global measures take also into account the number of similar parts compared to the total number of parts in the query and in the target model. Moreover, an ad-hoc visualization interface has been designed to clearly highlight the different similarities and to allow a fast identification of the target models. The validation of the proposed measures is discussed through several academic and industrial examples run on the developed prototype software.

1. Introduction

Today, being able to efficiently retrieve CAD assembly models from large databases has become a paramount issue. Indeed, over the years, companies accumulate a huge amount of Digital Mock-Up (DMU) of their products. For instance, considering complex products, the size of the DMU can reach several hundred thousand parts assembled together in assembly trees containing tens of subassembly levels, and requiring several terabytes for the storing. To be more competitive on the market, new tools have to be developed to better support the phases of the Product Development Process (PDP), such as the retrieval of CAD assembly models. Actually, the retrieval of existing similar models provides a way to access to knowledge gained from previous designed solutions [1]. For example, it is possible to reuse existing digital models and exploit their associated information. Moreover, the similarity evaluation provides benefits in other engineering activities, as for the identification of interchangeable parts from separate projects to reduce management and manufacturing costs [2]. The ability to evaluate the similarity between two assembly models plays a central role in the development of a retrieval approach. However, assessing the similarity is not an easy task as it involves several issues.

A first issue concerns the plurality of the *similarity criteria* according to which two models can be considered similar. This strongly relies on user's purpose. Deshmukh et al. describe different application scenarios and present the most suitable similarity criteria to be satisfied for each application [3]. For instance, if a designer wants to reuse a digital model, he/she can start with a rough query (i.e. a simple CAD model with few details) to retrieve a more detailed similar one, which can then be adapted and modified according to new requirements. In this scenario, the overall shape could be one of the salient characteristics to evaluate the similarity. Another scenario requires the retrieval of similar models to get access to existing product information (e.g. simulation results, assembling strategies or failure reports). In this case, only considering the shape to evaluate the similarity of the products could be limiting, while evaluating the similarity according to the functional point of view could be more interesting. In addition, other useful characteristics exist and can be used to evaluate the similarity (e.g. products dimensions, required manufacturing processes or materials, design intent). Thus, a good similarity assessment technique should be able to consider multiple similarity criteria.

A second issue concerns the *availability of the information* necessary to compare models according to multiple similarity criteria. Actually, depending on the criteria, the required information is not always directly available as it is neither encoded in the CAD models nor in the supporting systems, such as Product Data Management (PDM) or Product Lifecycle Management (PLM) systems. In this case, the similarity assessment technique should embed specific reasoning mechanisms to extract information implicitly encoded in the CAD models.

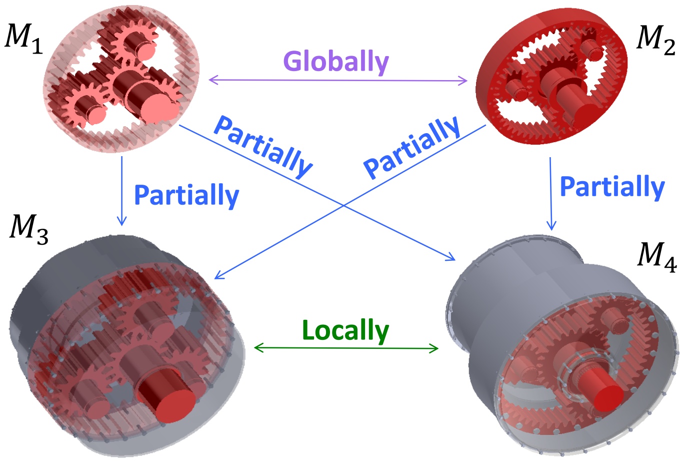


Fig. 1. Example of different similarity levels for the retrieval of CAD assembly models (local, partial and global).

A third issue concerns the plurality of the *similarity levels* according to which the similarity can be assessed. Indeed, retrieval methods have to deal with *global* and *partial* matching, where the partial matching can be further split into *part-in-whole* (i.e. the input model is inside a whole retrieved object) and *whole-to-whole by partial similarity* (i.e. a subpart of the input object is similar to a subpart of a retrieved object) [4, 5]. An example of the different similarity levels is depicted in Fig. 1. The CAD assembly model M_1 is *globally similar* to the model M_2, since they have analogous parts. The model M_1 (resp. M_2) is considered *partially similar* (i.e. part-in-whole) to M_3 (resp. M_4), since it is completely included in the second one. Finally, models M_3 and M_4 are *locally similar* (i.e. whole-to-whole by partial similarity), since they share similar parts.

A fourth issue originates from the two previous ones. Due to the multiple similarity criteria and levels characterizing the similarity between two CAD assembly models, it is not straightforward *to properly present to the user the ranked results* of a given query according to specific criteria.

This paper aims at introducing a new approach to assess the similarity of CAD assembly models according to multiple criteria. The contribution is threefold: (i) a new approach to retrieve CAD models according to multiple similarity criteria based on both extrinsic and intrinsic characteristics extracted from CAD assembly models; (ii) assessment of the similarity of assembly models according to different similarity levels (local, partial and global); (iii) an ad-hoc visualization interface to enhance the user experience when analyzing the results of the similarity assessment. The rest of the paper is organized as follows. Section 2 reviews some related works. Section 3 reports some details on the adopted framework for the comparison of assembly models, while section 4 and section 5 detail the similarity criteria and the proposed measures for their evaluation. Finally, section 6 presents the considered database and discusses some results obtained using our prototype software. Section 7 ends the paper providing conclusions and future steps.

2. Related works

The DMU of a product gathers together all the digital information generated and used all along the PDP. The data are usually managed by PDM and PLM systems which also integrate functionalities to ease the work between the actors of the PDP. Not only it contains CAD models, but the tracked data usually involve the technical specifications of the product, provisions for its manufacturing and assembling, the types of materials that will be required to produce it and many more information. However, design information is mostly contextual, i.e. many information is contained in the CAD model itself, and such systems provide a limited support in geometric searches [6]. To overcome these limitations, content-based 3D model retrieval has become an active research area and many works emerged in the last decade. Those techniques usually make use of pre-computed shape descriptors or signatures to evaluate the similarity among the models and thus to allow the retrieval of similar models. In this section, solely the methods that directly address the retrieval of CAD assembly models are reviewed. For a comprehensive survey of the state-of-the-art on general 3D model retrieval, one can refer to the survey papers [4, 7, 8, 9].

There exist retrieval methods that describe assemblies just through the description of the shape of their parts. One of the most popular shape descriptor is the shape distribution proposed by Osada et al. [10], and which has been used in the methods developed by Renu and Mocko [11], Wang et al. [12] and Zhang et al. [13]. These approaches strongly rely on the shape information and do not use any assembly relationships (e.g. geometric constraints, kinematic links or parts arrangement). However, in case of complex assemblies made of several parts, a method based only on the shape is not sufficient to retrieve efficiently the target assembly models. Indeed, 3D models with similar shapes can be assembled in different ways, involving different kinematic characteristics and then different relationships between their parts.

The relationships between the parts of an assembly models are usually represented by graph-based descriptors. For instance, Miura and Kanai [14] represent models by attributed graphs, which encode structural information and other data, e.g. contact and interference stages and geometric constraints. Tao and Huang [15] use component attributed relational graphs, where arcs represent the adjacency relationships between two components encoding the types of surface (e.g. planar or cylindrical surfaces) and the connection relations (e.g. screw connection, pin joint, key joint, rivet joint). Deshmukh et al. [3] propose a flexible retrieval system exploiting the data present in CAD models represented by mating graphs. Despite the fact that these three methods describe assembly models at a local level, their matching approaches allow to recognize just global similarities among assembly models. Moreover, if some information is missing it must be made explicit by the user.

A more complete system able to detect also partial similarities has been proposed by Chen et al. [16]. Their assembly descriptor considers different information levels including the topological structure, the relationships between the components of the assembly, as well as the geometric information. Anyhow, it assesses partial similarity by exploiting a hierarchical structure. This suggests that if two assembly models M_1 and M_2 represent the same object but have a different parts organization, then the system is not able to recognize M_1 and M_2 as similar, not even with a lower similarity measure.

Recently, Han et al. [17] proposed a retrieval system using a semantic representation of assembly models that allows to capture both the design and the assembly model intent of a model. In this method, the similarity is evaluated by a many-to-many matching procedure comparing each part in the query model with each part in the target model. Despite the fact that this method presents an interesting novelty in the description of an assembly model, the adopted matching procedure allows to evaluate only one level of similarity, the global similarity.

As a conclusion, even if content-based 3D model retrieval is an active topic, most of the existing works focus on single parts, and the methods which tackle assembly models can hardly manage a large variety of similarity criteria together with the possibility to assess the similarity at different levels, i.e. partial, local and global similarities. This paper aims at introducing measures to evaluate assembly similarities according to multiple criteria.

3. Assembly retrieval framework

For a comprehensive understanding of the new evaluation measures and procedures presented in sections 4 and 5, the overall assembly retrieval framework is first introduced in this section.

Basically, the similarity evaluation is a two-steps procedure: a proper descriptor for the assembly is first evaluated, then this signature is used for the models comparison and similarity assessment. The assembly retrieval framework adopted in this paper is based on the so-called Enriched Assembly Model (EAM) descriptor [18]. The EAM descriptor is computed both for the query model, i.e. the model according to which the similarity is to be assessed, and for all the CAD assembly models of the inquired dataset. Then, EAMs are compared two-by-two to evaluate the similarity of the query model with respect to all the models of the dataset. The results are stored and provided to the user who can rank the target models, i.e. the CAD assembly models resulting from the matching, according to the level of similarity he/she is interested in. The framework includes both real-time and batch processes. The batch process computes the complete EAM descriptors for all the CAD assembly models in the dataset, while the real-time processes compute partial EAM descriptors for the query including only the data that are pertinent to the user-specified similarity criteria, and then perform the comparison.

The framework as well as the EAM descriptor have been devised to perform retrieval of mechanical parts assemblies, under the assumption that the products are rigid objects that cannot be deformed. Moreover, the considered assembly models are supposed to contain a number of parts contacts which does not change over the time, whereas the portions of the parts involved in the contacts can change. The creation of the EAM relies on both explicit information directly available, and implicit information to be extracted through dedicated procedures. The CAD models are represented by means of a standard CAD format (STEP AP 203/214) so as to be independent of any CAD system. These data provide four different types of information (structure, statistic, interface and shape) which are stored in an attributed multi-graph structure, i.e. a graph structure that allows both multiple arcs between a pair of nodes and attributes for nodes as well as for arcs. The following subsections provide details on the information present in the EAM to clearly understand how these data contribute in the similarity assessment developed in sections 4 and 5.

3.1. Structure information

The structure information characterizes the hierarchical structure of an assembly, i.e. how the parts are gathered together in the CAD model. This information is available in CAD models and it is represented in the EAM by nodes and directed arcs. The root node corresponds to the entire assembly model and the leaves represent the parts constituting the assembly. Directed arcs between nodes represent the relation "made-of" between the components of the assembly. Two attributes belonging to the sets *CompType* and *PatternType* characterize nodes representing parts. The attribute in *CompType* defines the type of the component by the following values:

$$\begin{aligned} CompType = \big\{&\text{bearing, c-clip, gear, shaft, spacer, key, nut,linkage arm, parts of bearing, screw and bolt,} \\ &\text{cylinder-like, cube-like, sphere-like, torus-like, miscellaneous}\big\} \end{aligned} \tag{1}$$

The attribute in *PatternType* characterizes the arrangement in the 3D space of a set of repeated parts (indicated by *RP*). It can take the following values:

$$PatternType = \big\{\text{linear translation, circular translation, circular rotation, reflection}\big\} \tag{2}$$

Lastly, an attribute that represents all the patterns in the model can be associated to the root node, whose values belong to:

$$PatternList = \Big\{\big(\text{patType}_i, \text{RP}_i\big),\ \forall i \in \{1, ..., N_p\}\Big\} \tag{3}$$

where N_p is the number of patterns in the model, patType_i corresponds to the type of the i^{th} pattern and RP_i to its constituting repeated parts.

The information regarding the type of components and the type of patterns is generally not included in the CAD models and then it is necessary to extract them. Details on how these attributes are computed can be found in [19, 20, 21, 22].

3.2. Statistic information

The statistic information has been designed to ease the filtering of large datasets and reduce the number of models to be compared. Statistics for single parts include: 1) percentage of the overall model area covered by surfaces of a specific type (i.e. planar, cylindrical, spherical, toroidal and free-form); 2) number of maximal faces (i.e. adjacent faces sharing the same underlying surface and which can be considered as a single face) of a specific type of surface. The use of such information allows discarding directly some shapes, thus reducing the number of candidates for the matching process. Thus, the generic statistic attribute for parts has values in the set *PartStat*, which gathers together five pairs of values in the range $([0, 1] \times \mathbb{N})$, one pair for each type of surface (i.e. planar, cylindrical, spherical, toroidal and free-form). Statistics for assembly are represented by an attribute whose values are in *AssemblyStat*, which includes the number of patterns of each of the four pattern types. Thus, it comes that:

$$PartStat = ([0, 1] \times \mathbb{N})^5 \tag{4}$$

$$AssemblyStat = \mathbb{N}^4 \tag{5}$$

3.3. Interface information

The interface information describes the relationships between parts of an assembly model regardless its structure. These relationships are expressed by contact information and joint information. In most configurations, contacts between two parts can be of type Surface, Curve or Point (Fig. 2.a). However, sometimes, there may be unrealistic configurations where two parts share a volume, i.e. they intersect each other (Fig. 2.b, 2.c and 2.d). In the proposed framework, volumetric interferences can be solved while deducing the correct original contact when one of the involved part is a sphere (Fig. 2.c and 2.d). For those configurations which cannot be solved, the attribute UnSolved is used to characterize the type of contact (Fig. 2.b). Within the EAM, a contact information is represented by an arc with attributes belonging to the two following sets:

$$\begin{aligned} ContactType &= \{\text{Surface, Curve, Point, UnSolved}\} && (6) \\ DOF &= Tra \times Rot && (7) \end{aligned}$$

The degree of freedom (DOF) of a given contact is computed while identifying the set of allowed translations (Tra) and rotations (Rot) between the parts in contact [21]. Here, both sets are expressed by normalized vectors characterizing either a translation direction or a rotation axis according to the global reference frame.

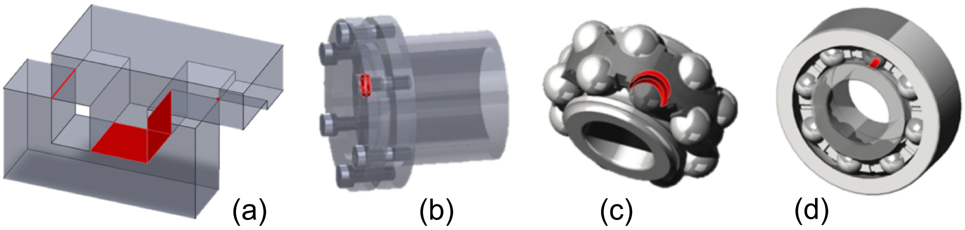


Fig. 2. Possible contacts between parts of a CAD assembly model: (a) contacts of type Surface, Curve and Point, (b) Unsolved contact type, (c) volumetric interference solved as Surface contact type, (d) volumetric interference solved as Curve contact type.

Joint information describes the motion resulting from all the contacts between two parts. Its characterization involves attributes that define the type of joint (i.e. the type of contacts involved in the joint), and the degree of freedom resulting from all the constraints imposed by the contacts: JointType = {Surface, Curve, Point, UnSolved, Mixed}

The so-called Mixed type is used to characterize a joint derived from contacts of different types. The type of contact, the degree of freedom and the type of joint are not explicitly encoded in the CAD models and should be computed as described in [20].

3.4. Shape information

The shape information gathers together two attributes used to characterize the shape and size of the parts. The first attribute characterizes the shape and it makes use of 3D spherical harmonics [23] defined by an histogram of 544 bins. While the second attribute characterizes the size and it gathers together the volume and the surface area of a part. Thus, the shape information is defined by two attributes, whose values belong to the *Shape* and the *Size* sets:

$$\begin{aligned} Shape &= \mathbb{R}^{544} && (8) \\ Size &= \mathbb{R}^{+} \times \mathbb{R}^{+} && (9) \end{aligned}$$

3.5. Graph representation

Nodes, arcs and attributes described in the previous subsections contribute to the creation of an attributed multi-graph that represents an EAM descriptor. Let $G(\mathcal{N}, \mathcal{A}, \Phi_{\mathcal{N}}, \Phi_{\mathcal{A}})$ be an attributed multi-graph representation of an EAM descriptor, where $\mathcal{N}$ is the set of nodes, $\mathcal{A}$ is the set of arcs, and $\Phi_{\mathcal{N}}$ and $\Phi_{\mathcal{A}}$ are respectively the node and arc attribute functions. Different types of nodes and arcs are defined according to the different types of information extracted from the assembly model or deducted by reasoning processes. In particular, $\mathcal{N} = \mathcal{N}_P \cup \mathcal{N}_A$ and $\mathcal{A} = \mathcal{A}_S \cup \mathcal{A}_C \cup \mathcal{A}_J$, where $\mathcal{N}_P$ and $\mathcal{N}_A$ are the sets of nodes associated respectively with parts and sub-assemblies, while $\mathcal{A}_S$, $\mathcal{A}_C$ and $\mathcal{A}_J$ are the sets of arcs representing respectively the assembly hierarchical structure, the contacts between parts and the joints between parts. The node attribute function is defined as:

$$\Phi_{\mathcal{N}} : \left| \begin{array}{ccl} \mathcal{N}_P \cup \mathcal{N}_A & \rightarrow & T_{\mathcal{N}_P} \cup T_{\mathcal{N}_A} \\ n & \mapsto & \Phi_{\mathcal{N}}(n) = \begin{cases} \Phi_{\mathcal{N}_P}(n) \text{ if } n \in \mathcal{N}_P \\ \Phi_{\mathcal{N}_A}(n) \text{ if } n \in \mathcal{N}_A \end{cases} \end{array} \right. \tag{10}$$

where the parts and sub-assemblies functions are expressed as follows:

$$\Phi_{\mathcal{N}_P} : \left| \begin{array}{ccl} \mathcal{N}_P & \rightarrow & T_{\mathcal{N}_P} = \mathit{Shape} \times \mathit{Size} \times \mathit{CompType} \\ & & \qquad\qquad \times \mathit{PatternType} \times \mathit{PartStat} \\ n & \mapsto & \left(\Phi^{Sh}_{\mathcal{N}_P}(n), \Phi^{Si}_{\mathcal{N}_P}(n), \Phi^{CT}_{\mathcal{N}_P}(n), \Phi^{PT}_{\mathcal{N}_P}(n), \Phi^{PS}_{\mathcal{N}_P}(n)\right) \end{array} \right.$$

$$\Phi_{\mathcal{N}_A} : \left| \begin{array}{ccl} \mathcal{N}_A & \rightarrow & T_{\mathcal{N}_A} = \mathit{PatternList} \times \mathit{CompType} \\ & & \qquad\qquad \times \mathit{AssemblyStat} \\ n & \mapsto & \left(\Phi^{PL}_{\mathcal{N}_A}(n), \Phi^{CT}_{\mathcal{N}_A}(n), \Phi^{AS}_{\mathcal{N}_A}(n)\right) \end{array} \right.$$

The structure arcs in $\mathcal{A}_S$ have no attribute. As a consequence, and similarly to what has been defined for nodes, the arc attribute function sets up the attributes of the elements of $\mathcal{A}_C$ and $\mathcal{A}_J$ as follows:

$$\Phi_{\mathcal{A}} : \left| \begin{array}{ccl} \mathcal{A}_C \cup \mathcal{A}_J & \rightarrow & T_{\mathcal{A}_C} \cup T_{\mathcal{A}_J} \\ a & \mapsto & \Phi_{\mathcal{A}}(n) = \begin{cases} \Phi_{\mathcal{A}_C}(n) \text{ if } a \in \mathcal{A}_C \\ \Phi_{\mathcal{A}_J}(n) \text{ if } a \in \mathcal{A}_J \end{cases} \end{array} \right. \tag{11}$$

where the contact and joint arc functions are expressed as follows:

$$\Phi_{\mathcal{A}_C} : \left| \begin{array}{ccl} \mathcal{A}_C & \rightarrow & T_{\mathcal{A}_C} = \mathit{DOF} \times \mathit{ContactType} \\ a & \mapsto & \left(\left(\Phi^{Tra}_{\mathcal{A}_C}(a), \Phi^{Rot}_{\mathcal{A}_C}(a)\right), \Phi^{CT}_{\mathcal{A}_C}(a)\right) \end{array} \right.$$

$$\Phi_{\mathcal{A}_J} : \left| \begin{array}{ccl} \mathcal{A}_J & \rightarrow & T_{\mathcal{A}_J} = \mathit{DOF} \times \mathit{JointType} \\ a & \mapsto & \left(\left(\Phi^{Tra}_{\mathcal{A}_J}(a), \Phi^{Rot}_{\mathcal{A}_J}(a)\right), \Phi^{JT}_{\mathcal{A}_J}(a)\right) \end{array} \right.$$

Fig. 3 illustrates an example of the EAM graph structure created from a CAD model and enriched with semantic information extracted by the geometric reasoning algorithms. For readability purposes, only a subset of the attributes is represented. The single line-circled nodes are associated with parts, while the double-line nodes are associated to parts belonging to regular patterns. The labels of the nodes represent the type of component. The straight arcs connect two parts which are in contact, and the associated label indicates the DOF. If the label is not present, then the contact is of type UnSolved. The wavy arcs indicate a line contact and according to the description of the interface layer, the DOF between parts in contact by a vertex or an edge are not considered. Thus in these cases, labels specifying the corresponding degree of freedom are not specified.

3.6. The matching procedure

Using attributed graphs as assembly descriptors, the problem of finding the local similarity between two CAD assembly models is equivalent to find the Maximum Common Sub-graph (MCS) between two graphs, namely G_q and G_k, representing respectively the *query* and *target* models. Among the different techniques to solve the MCS problem [24], our problem is transformed into a Maximum Clique (MC) problem [25]. As a consequence, a clique (i.e. a complete subgraph where each pair of nodes is connected) has to be detected in a suitable *association graph* derived from G_q and G_k.

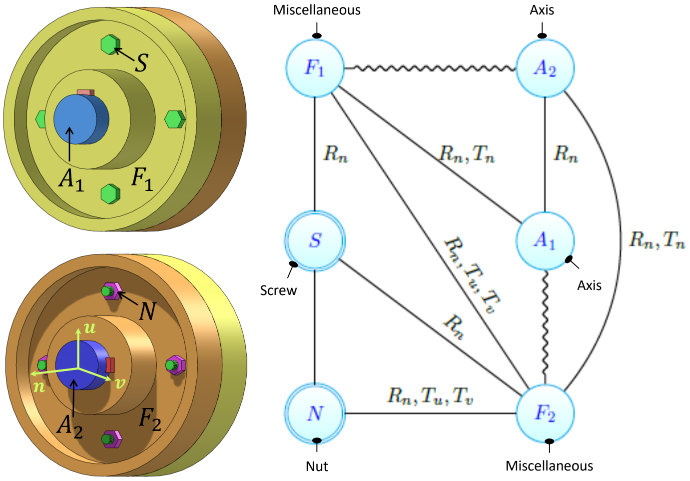


Fig. 3. Example of a partial EAM descriptor, associated to the CAD model of flange represented from the front and rear views.

In an association graph, nodes represent pairs of *similar nodes* between G_q and G_k, whereas arcs identify *similar relationships*. Here, c_N and c_A respectively refer to the node and arc similarity criteria and they are described in section 4. The association graph is not unique and its definition depends on the similarity criteria. Thus, once c_N and c_A are specified, there exist a unique association graph G_{q,k,c_N,c_A}. The maximum cliques in G_{q,k,c_N,c_A} represent the common sub-graphs between G_q and G_k according to the criteria c_N and c_A. The generic r^{th} clique in the graph G_{q,k,c_N,c_A} is expressed as $(C_{q,k,c_N,c_A})_r \subseteq G_{q,k,c_N,c_A}$ and the set of all the cliques for the association graph G_{q,k,c_N,c_A} is denoted as $\mathcal{D}_{q,k} = \left\{ (C_{q,k,c_N,c_A})_r \mid 1 \leq r \leq N_{qk} \right\}$, where N_{qk} is the number of maximum cliques in the association graph G_{q,k,c_N,c_A}.

In the proposed framework, cliques are detected using an exact graph matching method based on the well-known Bron and Kerbosh approach [26].

4. Similarity criteria

Considering nodes, the similarity criteria c_N that can be set up concern the shape, size, component type and pattern type information. Each of these criteria is specified by the values of its attribute function. The shape similarity is based on the shape descriptor of each node, i.e. the 3D spherical harmonics. Two 3D spherical harmonics can be compared in different ways. However, Kazhdan et al. [23] studied some properties of the spherical harmonics, and in particular, they demonstrated that the L_2-norm is the most appropriate for the evaluation of the shape similarity. The evaluation of component and pattern type similarity requires to check if the values of the *CompType* and *PatternType* attributes are the same for the nodes. Following this analysis, similarity criteria can be defined as follows:

Definition 1. Two nodes n_q^i and n_k^j are considered similar according to the shape criterion if and only if:

$$\left\| \Phi_{\mathcal{N}_P}^{Sh}(n_q^i) - \Phi_{\mathcal{N}_P}^{Sh}(n_k^j) \right\|_2 < \epsilon_{shape}$$

where ϵ_{shape} represents the threshold set in the query.

Definition 2. Two nodes n_q^i and n_k^j are considered similar according to the size criterion if and only if:

$$\mathrm{abs}\Big(\Phi_{\mathcal{N}_P}^{Si}(n_q^i) - \Phi_{\mathcal{N}_P}^{Si}(n_k^j) \Big) < \epsilon_{size}$$

where ϵ_{size} represents the threshold set in the query.

Definition 3. Two nodes n_q^i and n_k^j are considered similar according to the component type criterion if and only if:

$$\Phi_{\mathcal{N}_P}^{CT}(n_q^i) = \Phi_{\mathcal{N}_P}^{CT}(n_k^j)$$

Definition 4. Two nodes n_q^i and n_k^j are considered similar according to the pattern type criterion if and only if:

$$\Phi_{\mathcal{N}_P}^{PT}(n_q^i) = \Phi_{\mathcal{N}_P}^{PT}(n_k^j)$$

Once pairs of similar nodes have been generated in the association graph, it has to be checked if there exist an arc between the generic pair of association nodes (n_q^i, n_k^l) and (n_q^j, n_k^h) that satisfies similar relationship criteria c_A, or if both the node pairs (n_q^i, n_k^l) and (n_q^j, n_k^h) are not linked by an arc. The generic arc in G_q (resp. G_k) between the node pair (n_q^i, n_q^j) (resp. (n_k^l, n_k^h)) is indicated as a_q^{ij} (resp. a_k^{lh}). Here, a_q^{ij} and a_k^{lh} are respectively part of the sets $\mathcal{A}_q$ and $\mathcal{A}_k$. Since two nodes can be linked by multiple contact arcs, the set of contact arcs between the node pair (n_q^i, n_q^j) (resp. (n_k^l, n_k^h)) is indicated as $\mathcal{A}_{C_q}^{ij}$ (resp. $\mathcal{A}_{C_k}^{lh}$). The number of elements in a set $*$ is indicated by $|*|$. To evaluate the relationships between parts, three possible similarity conditions are defined according to the following definitions:

Definition 5. Two arcs a_q^{ij} and a_k^{lh} are considered compatible according to the contact type criterion if and only if:

$$(a_q^{ij} \in \mathcal{A}_{C_q}) \text{ and } (a_k^{lh} \in \mathcal{A}_{C_k})$$

Definition 6. The set of arcs $\mathcal{A}_{C_q}^{ij}$ and $\mathcal{A}_{C_k}^{lh}$ are considered compatible according to the allowed DOF for contacts criterion if and only if $\forall a_s \in \mathcal{A}_{C_q}^{ij}$ $\exists\, a_t \in \mathcal{A}_{C_k}^{lh}$ such that:

$$\begin{aligned}&\Big[\big|\Phi_{\mathcal{A}_C}^{Tra}(a_s)\big| = \big|\Phi_{\mathcal{A}_C}^{Tra}(a_t)\big| \text{ and } \big|\Phi_{\mathcal{A}_C}^{Rot}(a_s)\big| = \big|\Phi_{\mathcal{A}_C}^{Rot}(a_t)\big|\Big]\\ &\quad \text{or } \big[\Phi_{\mathcal{A}_C}^{CT}(a_s) = \text{UnSolved}\big] \text{ or } \big[\Phi_{\mathcal{A}_C}^{CT}(a_t) = \text{UnSolved}\big]\end{aligned}$$

Definition 7. Two arcs a_q^{ij} and a_k^{lh} are considered compatible according to the allowed DOF for joints criterion if and only if:

$$\begin{aligned}&\Big[\big|\Phi_{\mathcal{A}_J}^{Tra}(a_q^{ij})\big| = \big|\Phi_{\mathcal{A}_J}^{Tra}(a_k^{lh})\big| \text{ and } \big|\Phi_{\mathcal{A}_J}^{Rot}(a_q^{ij})\big| = \big|\Phi_{\mathcal{A}_J}^{Rot}(a_k^{lh})\big|\Big]\\ &\quad \text{or } \big[\Phi_{\mathcal{A}_J}^{JT}(a_q^{ij}) = \text{UnSolved}\big] \text{ or } \big[\Phi_{\mathcal{A}_J}^{JT}(a_k^{lh}) = \text{UnSolved}\big]\end{aligned}$$

The right part of Fig. 4 shows an assembly model representing a flange with three screws and a portion of its attributed multi-graph. For sake of readability, the root node corresponding to the entire assembly model is omitted and only $\mathcal{A}_J$ arcs are depicted. The right part of Fig. 4 shows a similar model together with a portion of its attributed multi-graph. Here, the assembly model only contains two screws. The nodes of the two attributed multi-graphs represent the parts of the CAD models, same type of line indicates same value of the spherical harmonic shape descriptor (i.e. parts with similar shapes) and parts with the same color belong to patterns of a specific type (i.e. green for circular translation and red for linear translation). The arcs represent the joint contacts where the labels indicate the DOF allowed between two linked parts.

In this example, the hypothesis is that the user seeks assemblies in which parts with similar shape are connected by the same joint relationships. This means that two nodes create an association node if their corresponding parts have similar shape, while association arcs are added if the joint arcs (between the related pairs of nodes in the attributed

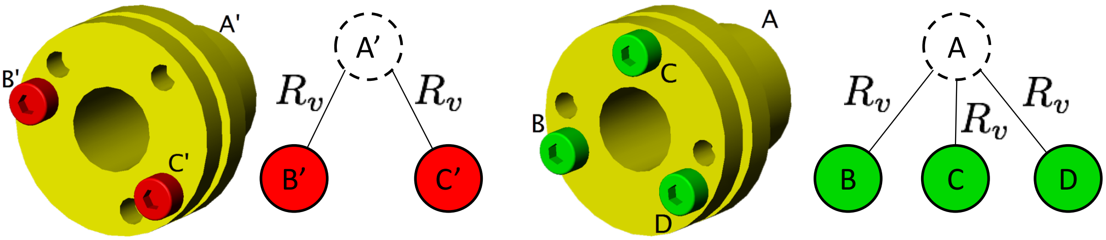


Fig. 4. Example of two assembly models with a different number of screws, and their EAM attributed multi-graphs G_1 and G_2.

multi-graphs) have the same number of rotations and same number of translations. The association graph, resulting from these criteria, is illustrated in Fig. 5. Here, there are six possible maximum cliques: $(C_{1,2,c_N,c_A})_1 = \{AA', BB', CC'\}$, $(C_{1,2,c_N,c_A})_2 = \{AA', BC', DB'\}$, $(C_{1,2,c_N,c_A})_3 = \{AA', CB', DC'\}$, $(C_{1,2,c_N,c_A})_4 = \{AA', CC', DB'\}$, $(C_{1,2,c_N,c_A})_5 = \{AA', BB', DC'\}$, $(C_{1,2,c_N,c_A})_6 = \{AA', BC', CB'\}$.

Each clique represents a possible sub-graph matching between the two attributed multi-graphs G_1 and G_2 (Fig. 4). The detected cliques are used to evaluate the similarity, as explain in the following section.

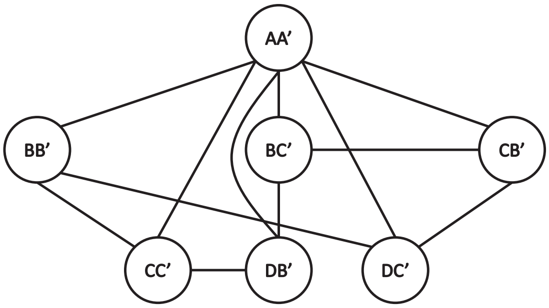


Fig. 5. Association graph for the objects in Fig. 4 built with shape and joint criteria of similarity

5. Similarity assessment

Depending on the user-specified similarity criteria c_N and c_A, the similarity between two CAD assembly models can be assessed differently. To evaluate the similarity, a bundle of 4 measures $\mathcal{S} = [\mu^{shape}, \mu^{joint}, \mu^{position}, \mu^{structure}]$ is proposed, where each of them characterizes a single aspect of the similarity between two assembly models. As previously said, the similar elements between the query and the compared assemblies correspond to the common sub-graphs between the two corresponding attributed multi-graphs. Thus, the similarity is computed on the detected cliques. To simplify the writing, in the following, the generic r^{th} clique $(C_{q,k,c_N,c_A})_r$ will be denoted as C, the set of its nodes as $\mathcal{N}_C$, and the number of nodes in $\mathcal{N}_C$ as $|\mathcal{N}_C|$.

5.1. Shape similarity measure: μ^{shape}

The $\mu^{shape}(C)$ shape similarity measure is based on the shape descriptor of each node involved in the clique C, i.e. the 3D spherical harmonics and the size values. Since two objects can have exactly the same shape but different dimensions, the assessment of the shape similarity is based uniquely on the 3D spherical harmonics, while the size values are considered as a standalone measure, which can be used to refine the retrieval results. As previously discussed, the

L_2-norm is an appropriate norm to compare 3D spherical harmonics [23]. Thus, the shape similarity of a clique C is defined as the average of the shape similarity of each node in the clique:

$$\mu^{shape}(C) = \frac{1}{|\mathcal{N}_C|} \sum_{(n_q^i, n_k^j) \in C} 1 - \left\| \frac{\Phi_{\mathcal{N}_P}^{Shape}(n_q^i)}{\|\Phi_{\mathcal{N}_P}^{Shape}(n_q^i)\|_2} - \frac{\Phi_{\mathcal{N}_P}^{Shape}(n_k^j)}{\|\Phi_{\mathcal{N}_P}^{Shape}(n_k^j)\|_2} \right\|_2 \tag{12}$$

5.2. Joint similarity measure: μ^{joint}

The $\mu^{joint}(C)$ joint similarity measure is defined to assess how much two assemblies are similar in terms of the relative DOF among their parts. A joint can arise from contacts of different types (Surface, Curve, Point or UnSolved). In case of joints deriving from contacts of type Surface, the allowed DOF of the two linked parts is available, otherwise only the information of the type of joint is accessible. Since joint arcs have two different types of attributes, this measure is defined as a combination of two other measures $\mu_{surf}^{joint}(C)$ and $\mu_{curve,pt}^{joint}(C)$ whose computation is not straightforward.

Actually, if two assembly models, identifying the same object, are embedded in different reference frames or simply rotated or translated, then their joint similarity measure should be the same. However, in the EAM, the information related to the DOF depends on the reference frame of the assembly model. As a consequence, a simple comparison between the DOF of the corresponding elements is not appropriate. This is illustrated in Fig. 6. The deviation between the axes characterizing the DOF between the parts P_1 and P_4 of Fig. 6.a, and the axes characterizing the DOF of the same parts in Fig. 6.b, corresponds to the angle defined by the axis u and n that is equal to 90 degrees even if the objects are the same.

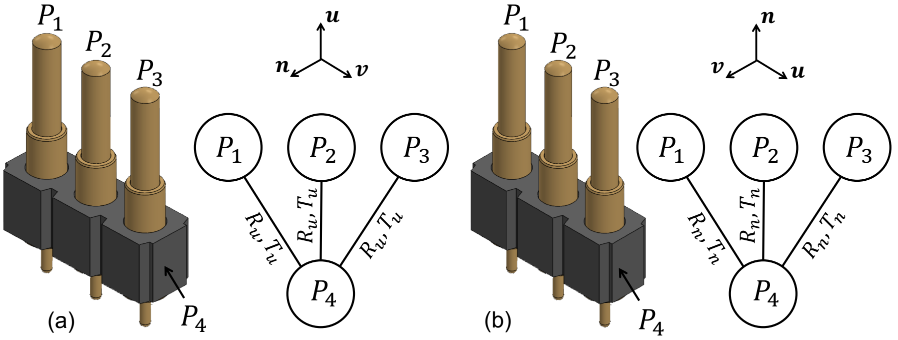


Fig. 6. Two instances of the same object embedded in two different reference frames: (a) query model, (b) target model.

On the contrary, the variations of each pair of axes defined by the DOF between the parts (P_1, P_4), (P_2 P_4) and (P_3 P_4) are the same in both configurations. Thus, the variation of the rotation and translation axes defined by the DOF of a part is considered according to all the parts in contact. Therefore, μ_{surf}^{joint} is computed on the nodes of the cliques instead of using the arcs. The configurations in which there exist several angles defined by the axes of the DOF are managed using matrices which capture all the possible variations of the rotation/translation angles related to a part. The elements of the matrices are computed using the dot product of a pair of axes, as specified in the following definitions:

Definition 8. $Var_{Tra}(n)$ is the matrix characterizing the variations of translations related to the node n. Its generic element is defined as $(Var_{Tra}(n))_{i,j} = t_i.t_j$, with $t_i, t_j \in Tra(n)$, and $Tra(n)$ the set of all the joint translations between the node n and its generic adjacent node n^*:

$$Tra(n) = \{\textstyle\bigcup_a \Phi_{\mathcal{A}_J}^{Tra}(a),\ \forall a : a = (n, n^*) \in \mathcal{A}_J\}$$

Definition 9. $Var_{Rot}(n)$ is the matrix characterizing the variations of rotations related to the node n. Its generic element is defined as $(Var_{Rot}(n))_{i,j} = r_i.r_j$, with $r_i, r_j \in \mathit{Rot}(n)$, and $\mathit{Rot}(n)$ the set of all the joint rotations between the node n and its generic adjacent node n^*:

$$\mathit{Rot}(n) = \{\bigcup_a \Phi^{Rot}_{\mathcal{A}_J}(a),\ \forall a : a = (n, n^*) \in \mathcal{A}_J\}$$

Since the axes are normalized, the dot product corresponds to the cosine of the angle between the considered axes. The final variations of a node in the clique are obtained by computing the averages of those matrices. Here, the average $\sigma(M)$ of a matrix M of size $N \times N$ is meant as the arithmetic mean of the elements in the matrix divided by the number of elements:

$$\sigma(M) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} m_{i,j} \qquad \forall m_{i,j} \in M \tag{13}$$

Following this method, when a single arc is incident to a node, its variation is equal to 1. Moreover, the joint measure has to take into account also contacts and joints of type UnSolved, for which the DOF information is not available. In such a configuration, the only possible difference between two models is the number of joints concurring to the definition of the variation matrix, i.e. joints of surface type. Thus, this information is used to distinguish these cases by dividing the average of each variation matrix by the number of translations/rotations involved in their definition, i.e. $|Tra(n)|$ and $|\mathit{Rot}(n)|$.

Finally, as the DOF are not computed for joints arisen from curve or point contacts, the defined joint measure $\mu^{joint}_{curve,pt}$ is based on the type of contacts and has maximum value 1 if the joints are of the same type (both Curve or Point) and a lower value if the joints are of different types. The lower value is set to 0.8 and it was chosen empirically to slightly decrease the measure.

As a result of this analysis, the overall joint similarity measure is defined as follows:

$$\mu^{joint}(C) = \Big(\mu^{joint}_{surf}(C) + \mu^{joint}_{curve,pt}(C)\Big)/2 \tag{14}$$

where the measures related to Surface and Curve/Point are computed as follows:

$$\mu^{joint}_{surf}(C) = \frac{1}{|\mathcal{N}_C|} \sum_{(n^i_q, n^j_k) \in C} \left[1 - \frac{d_{Tra}\big((n^i_q, n^j_k)\big) + d_{Rot}\big((n^i_q, n^j_k)\big)}{2}\right] \tag{15}$$

$$\mu^{joint}_{curve,pt}(C) = \frac{1}{|\mathcal{N}_C|} \sum_{((n^i_q, n^j_k),(n^l_q, n^h_k)) \in C^2} \Big[1 - d_{edge}\Big((a^{ij}_q, a^{lh}_k)\Big)\Big] \tag{16}$$

with:

- $d_{Tra}\big((n^i_q, n^j_k)\big) = \text{abs}\left(\frac{\sigma(Var_{Tra}(n^i_q))}{|Tra(n^i_q)|} - \frac{\sigma(Var_{Tra}(n^j_k))}{|Tra(n^j_k)|}\right)$
- $d_{Rot}\big((n^i_q, n^j_k)\big) = \text{abs}\left(\frac{\sigma(Var_{Rot}(n^i_q))}{|\mathit{Rot}(n^i_q)|} - \frac{\sigma(Var_{Rot}(n^j_k))}{|\mathit{Rot}(n^j_k)|}\right)$
- $d_{edge}\big((a^{ij}_q, a^{lh}_k)\big) = \begin{cases} 0 & \text{if } \Phi^{JT}_{\mathcal{A}_J}(a^{ij}_q) = \Phi^{JT}_{\mathcal{A}_J}(a^{lh}_k), \\ 0.2 & \text{otherwise.} \end{cases}$

So far, in equation (14), the combination of the two individual measures on Surface and Curve/Point has been weighted equally, but a different weighting could also be imagined to give more or less importance to the type of joints.

5.3. *Position similarity measure:* $\mu^{position}$

Another salient characteristic affecting the level of similarity between two assembly models is the relative arrangement of the assembly components. For instance, the assembly models in Fig. 7 have parts with similar shape and relationships (i.e. the colored parts are not in contact among them) but a different arrangement. The objective of the position similarity measure is to be able to distinguish such configurations while characterizing the position of the similar parts. To this aim, we consider the directional versors between the center of gravity of each couple of parts in the clique not in contact each other. The use of versors makes the measure size independent. Since the versors are dependent on the reference frame, to overcome this problem, the relative position similarity is computed following the same approach adopted for the computation of the joint similarity in case of surface type joint:

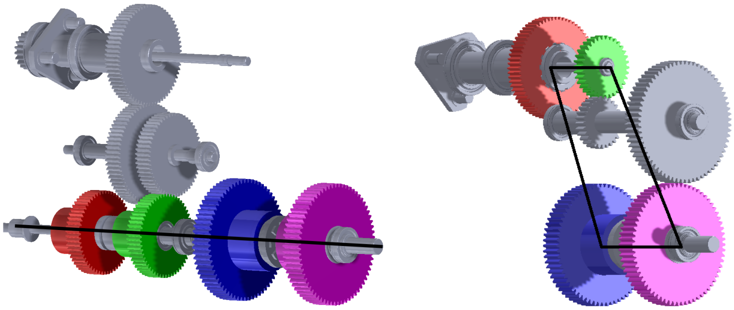


Fig. 7. Example of assembly models with similar parts arranged according to different configurations.

$$\mu^{position}(C) = \frac{1}{|\mathcal{N}_C|} \sum_{(n_q^i, n_k^j) \in C} \Big[1 - d_{Dir}\big((n_q^i, n_k^j)\big) \Big] \tag{17}$$

with:

- $d_{Dir}\big((n_q^i, n_k^j)\big) = \text{abs}\big(\sigma\big(Var_{Dir}(n_q^i)\big) - \sigma\big(Var_{Dir}(n_k^j)\big)\big)$
- $Var_{Dir}(n_q^i)$ is the variation matrix of the directional versors between the gravity centers of the parts corresponding to the node n_q^i and the nodes n_q^l such that $(n_q^i, n_q^l) \notin \mathcal{A}_{C_q}$,
- $Var_{Dir}(n_k^j)$ is the variation matrix of the directional versors between the gravity centers of the parts corresponding to the node n_k^j and the nodes n_k^h such that $(n_k^j, n_k^h) \notin \mathcal{A}_{C_k}$.

5.4. *Structure similarity measure:* $\mu^{structure}$

Similar products and CAD models can be organized in sub-assemblies in different ways according to the designer objectives, therefore it is important to be able to capture the differences at the level of the structure. Thus, the structure similarity measure characterizes the way parts are assembled in the assembly tree of a CAD model. The EAM descriptor encodes the hierarchical structure of an assembly model by a set of arcs $\mathcal{A}_S$.

Therefore, the proposed measure is based on the comparison of the structural relations of the pairs of nodes (n_q^i, n_q^l) and (n_k^j, n_k^h), where (n_q^i, n_k^j) and (n_q^l, n_k^h) are nodes of the clique. Its evaluation requires verifying if the nodes n_q^i and n_q^l belong (or not) to the same sub-assembly in G_q, and similarly if the nodes n_k^j and n_k^h belong (or not) to the same sub-assembly in G_k. Using a distance function equal to 0 if the pair of nodes $(n_q^i,\, n_q^l)$ has the same relation as the pair $(n_k^j,\, n_k^h)$, or 1 otherwise, the structure similarity measure of a clique is defined as follow:

$$\mu^{structure}(C) = \frac{1}{|\mathcal{N}_C|^2} \sum_{((n_q^i,n_k^j),(n_q^l,n_k^h))\in C^2} \Big[1 - d_{Str}\big((n_q^i, n_k^j), (n_q^l, n_k^h)\big)\Big] \tag{18}$$

with:

$$d_{Str}\big((n_q^i, n_k^j),(n_q^l, n_k^h)\big) = \begin{cases} 0 & \text{if } \big[\exists n_q^* \in \mathcal{N}_q \text{ and } \exists n_k^* \in \mathcal{N}_k\big] \text{ s.t.} \\ & \Big[\big((n_q^i, n_q^*), (n_q^l, n_q^*)\big) \in \mathcal{A}_{S_q} \\ & \quad \text{and } \big((n_k^j, n_k^*), (n_k^h, n_k^*)\big) \in \mathcal{A}_{S_k}\Big] \\ 1 & \text{otherwise.} \end{cases}$$

5.5. Combination of similarity measures

The definition of measures able to rank the retrieved models requires combining the above measures. Here, three overall measures are defined to characterize the local, partial and global similarities between CAD assembly models. Through the combinations, it should also be possible to weight differently each similarity measure using a factor either specified in the query or chosen during the browsing of the results. In this paper, the weights are specified by the user. However, in future works, the idea is to study the possibility to assign default weight values according to particular usage scenarios.

Definition 10. The assembly *local similarity* measure between two models is the weighted average of the four individual similarity measures and it is defined by the function:

$$\gamma_\ell : \left| \begin{array}{lcl} \mathcal{D}_{q,k} \times \mathcal{W} & \to & [0,1] \\ (C, w) & \mapsto & \dfrac{\begin{bmatrix} w^{sh}\mu^{shape}(C) + w^{jo}\mu^{joint}(C) \\ \quad + w^{po}\mu^{position}(C) + w^{st}\mu^{structure}(C) \end{bmatrix}}{w^{sh} + w^{jo} + w^{po} + w^{st}} \end{array} \right.$$

with $w = \{w^{sh}, w^{jo}, w^{po}, w^{st}\} \in \mathcal{W} = [0,1]^4$.

To provide information on how many parts are similar among all of the query and the target models, two so-called coverage factors have been identified. They are used to weight the local similarity measure and to define the partial and global similarity ones. The term *coverage factor* refers to the percentage of elements of the query and target models that are considered similar with respect to all the elements in the two models.

Definition 11. The partial and global coverage factors (PCF and GCF) are computed as follows:

$$PCF(C) = \frac{|\mathcal{N}_C|}{|\mathcal{N}_q|} \quad \text{and} \quad GCF(C) = \frac{2|\mathcal{N}_C|}{|\mathcal{N}_q| + |\mathcal{N}_k|}$$

where $\mathcal{N}_C$, $\mathcal{N}_q$ and $\mathcal{N}_k$ represent respectively the set of nodes in the clique C, the nodes in the query model and the nodes in the target model.

The global and partial similarity measures are defined weighting the local similarity according to these coverage factors. Right now, those measures do not yet consider the "relevance" of the parts. For instance, the contribution of a small part (i.e. rivet, c-clip) could be weighed differently from a bigger part (i.e. gear, shaft, bearing). In some cases, it could also be considered as negligible.

Definition 12. The assembly *partial similarity* measure between two models is evaluated by the function:

$$\gamma_p : \left| \begin{array}{lcl} \mathcal{D}_{q,k} \times \mathcal{W} & \to & [0,1] \\ (C, w) & \mapsto & \gamma_p(C, w) = PCF(C) \cdot \gamma_\ell(C, w) \end{array} \right.$$

Definition 13. The assembly *global similarity* measure between two models is evaluated by the function:

$$\gamma_g : \left| \begin{array}{ccl} \mathcal{D}_{q,k} \times \mathcal{W} & \rightarrow & [0,1] \\ (C,w) & \mapsto & \gamma_g(C,w) = GCF(C) \cdot \gamma_\ell(C,w) \end{array} \right.$$

One can finally notice that global similarity implies partial and local similarities. However, while partial similarity implies local similarity, the vice versa does not hold. In the end, the following rules stand:

- Two models are 100% *globally similar* if and only if they are 100% *locally similar* and $GCF = 1$.
- Two models are 100% *partially similar* if and only if they are 100% *locally similar* and $PCF = 1$.

6. Experimentations and results

To test the effectiveness of the proposed framework, an assembly model dataset is required. Unfortunately, the most known mechanical shape benchmarks in literature, as the Princeton Shape Benchmark (PSB) [27], the National Design Repository (NDR) [28] and the Engineering Shape Benchmark (ESB) [29], are not proper for our purpose. Indeed, they classify just parts and do not consider assembly models, and, so far, no public database exists to evaluate and compare assembly retrieval systems [30, 31]. Moreover, the authors of assembly retrieval methods have developed their own datasets, and these benchmarks are not public.

As a consequence, the proposed assembly retrieval framework has been benchmarked using our own assembly models dataset that contains 140 models arranged in 12 classes (Table 1). There are 15057 parts in total, out of which 5343 parts are unique.

Category	Number	Category	Number
Bearing	36	Landing gear	7
Coupling flange	5	Linear actuator	10
Double rotor turbine	13	Mill max	8
Hinge	4	Piston	5
Hydraulic reduction	6	Propeller mixer	18
Hydraulic rotor	6	Rotor wind turbine	22

Table 1. Classification of the 140 CAD assemblies forming our testing set.

To display and browse the results in an intuitive and user-friendly manner, multi-view dynamic web pages have been developed based on HTML5, jQuery, Ajax, PHP and X3D for model visualization. In this way, the user does not see a list of names, but a 3D overview of the target models with their matched components highlighted. This is illustrated in Fig. 8. In the 3D view, the matched components are colored in blue while the unmatched ones are in red. The bars of the histograms indicate the values of the local (yellow bar), global (green) and partial (purple) overall similarity measures.

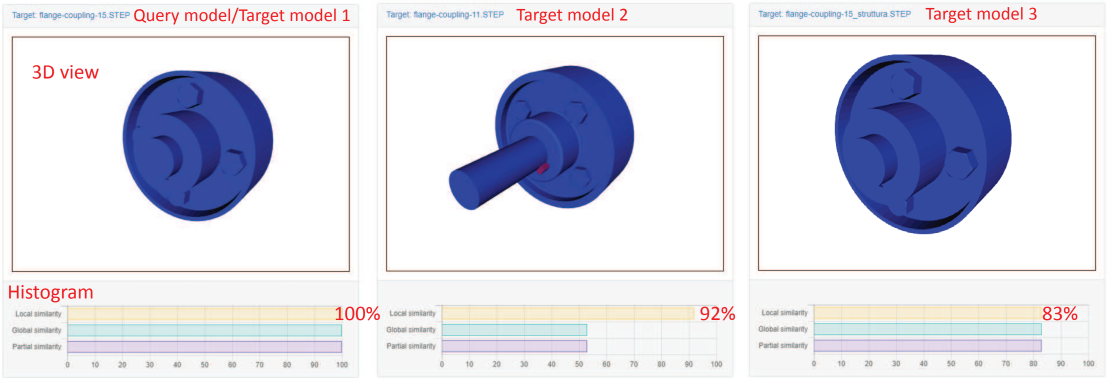


Fig. 8. Example of two target models (middle, right) retrieved from a given query model (left).

Each model can be further analyzed in another view, where the query and the target models are displayed. This is illustrated in Fig. 9. In the 3D view, each pair of matched parts is highlighted by a different color, and the values of the single measures are reported in a radar chart. This example is further discussed in the next section.

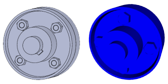
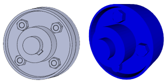
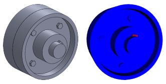
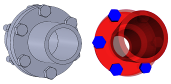
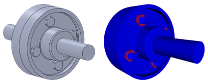
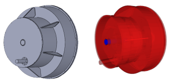
CAD models & matched parts	#1 	#2 	#3 	#4 	#5 	#6 
μ^{shape}	1.00	1.00	0.77	0.89	0.74	0.77
μ^{joint}	1.00	1.00	0.68	1.00	1.00	0.00
$\mu^{position}$	1.00	1.00	0.99	0.87	0.95	0.82
$\mu^{structure}$	1.00	0.35	1.00	1.00	1.00	1.00
γ_ℓ	1.00	0.83	0.86	0.94	0.92	0.65
γ_p	1.00	0.83	0.74	0.54	0.52	0.19
γ_g	1.00	0.83	0.74	0.54	0.52	0.05

Table 2. Similarity evaluation between a coupling flange query model (#1) and several target models (#1 to #6) with a weighting vector $w_{flange} = \{1, 1, 1, 1\}$.

6.1. Coupling flange assembly models retrieval

The first experimentation aims at identifying the mechanical coupling flanges contained in the database. A mechanical coupling flange is a set of components linking two parts of a product. In this example, the query model corresponds to the first model (#1) illustrated in Table 2. It contains four screws and fours nuts arranged in a circular translation pattern, two main flanges, two shafts and two keys. All the parts are organized in a flat structure, i.e. without any sub-assembly. The CAD model does not present any volumetric intersection and each screw is in contact with the corresponding nut through an idealized cylindrical face (i.e. the screw thread is not modeled).

The user-specified similarity criteria c_N require that the nodes have to be similar according to the shape, the component type and the pattern type. In this case, the threshold ϵ_{shape} used for the shape criterion is set up to 0.20, thus two components should have shapes similar at 80% according to the values of their 3D spherical harmonics. The user-specified similarity criteria c_A requires that two pairs of compatible nodes should have the same number of allowed rotations and translations. Based on these criteria, the single similarity measures μ^{shape}, μ^{joint}, $\mu^{position}$ and $\mu^{structure}$ can be computed. Then, the overall similarity measures γ_ℓ, γ_p and γ_g can be evaluated. Here, the weights are all equal and $w_{flange} = \{1, 1, 1, 1\}$. This means that the single similarity measures have the same importance to compute the local, partial and global similarity measures. The numerical results are gathered together in Table 2.

The first model (#1) corresponds to the query model, it is therefore straightforward that all its measures have maximum values since it perfectly matches itself.

The second model (#2) has the same components as the query model, i.e. same number of parts, same shape and same contacts but organized in a different way. Indeed, the structure of the query model is flat, while in this target model, the set of screws and the set of nuts are gathered together forming two sub-assemblies. Thus, its $\mu^{structure}$ is less than 1 and this factor decreases the final value of the local similarity measure. Since all the components of the query and of the target model are matched, the values of the partial and global similarities correspond to the local one, i.e. $PCF = 1$ and $GCF = 1$ according to definitions 12 and 13.

In the third model (#3), screws and nuts present a volumetric intersection and are matched thanks to the use of the UnSolved attribute. Differently from the arcs in the query model, in the target model (#3) the arcs of type UnSolved do not have a proper number of allowed rotations and translations. Thus, according to the definitions 15, this difference affects their similarity at the level of the joint. In this example, the number of matched components is twelve and the number of components in the query and in the target models is fourteen, then the overall partial and global similarity measures are lower than the local one according to the same factor, i.e. $GCF = \frac{2 \cdot 12}{14+14} = 0.86$ and $PCF = \frac{12}{14} = 0.86$.

The fourth (#4) and fifth (#5) models have very similar measures, however at first glance, the coverage of these two target models seams different. Actually, the coverage of the models is measured according to the number of matched elements and the two models have the same number of matched elements: four screws and four nuts for the fourth model against four screws, two main flanges, a shaft and a key in the fifth one. An evaluation using the volume may improve the visual perception of similarity, but in general, it is more meaningful to consider the relevance of the matching parts, i.e. fastener elements should be less important than a shaft. Of course, such a consideration requires a study on the relevance for each component category of the different types of mechanical assembly models.

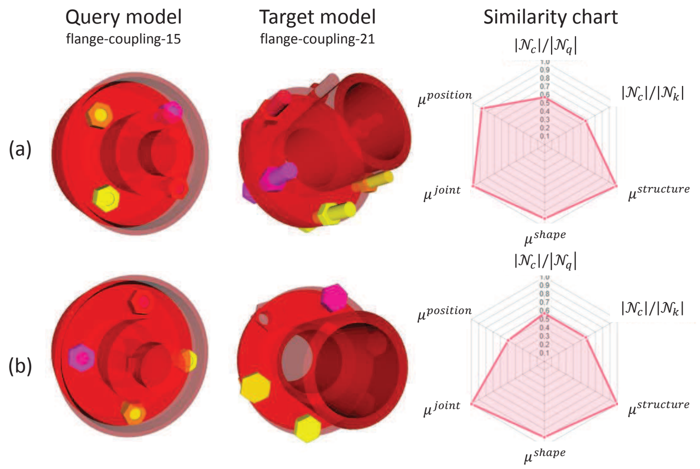


Fig. 9. Two different cliques (a, b) when comparing the target model (#4) to the query model (#1) in Table 2. The arrangement of the parts is different which is reflected by different values of $\mu^{position}$ in the radar charts.

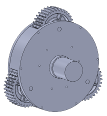
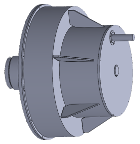
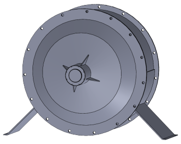
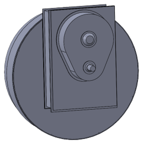
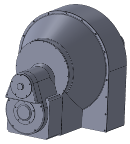
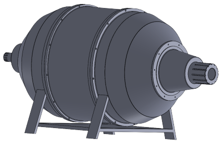
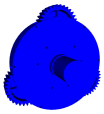
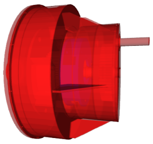
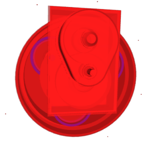
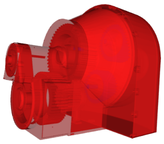
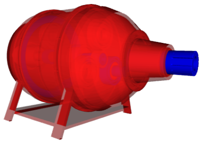
	#1	#2	#3	#4	#5	#6
CAD models						
Matched parts						
μ^{shape}	1.00	1.00	0.78	0.66	0.76	0.74
μ^{joint}	1.00	1.00	0.00	0.00	0.00	0.00
$\mu^{position}$	1.00	1.00	1.00	1.00	0.80	0.61
γ_ℓ	1.00	1.00	0.59	0.55	0.52	0.45
γ_p	1.00	1.00	0.44	0.42	0.39	0.34
γ_g	1.00	0.18	0.06	0.04	0.06	0.10

Table 3. Similarity evaluation between a planetary gearbox query model (#1) and several target models (#1 to #6) with a weighting vector $w_{gear} = \{1,1,1,0\}$.

For the target model (#4), Fig. 9 reports two different cliques, which correspond to two different sets of similar parts. Both the query and the target models have a circular pattern of screws and nuts, but with a different number of repeated elements (i.e. four in the query model and six in the target model). Thus, in the target model, it is not possible to find four equidistant screws and nuts that cover an entire circumference. This affects the position similarity measure as depicted on the radar charts which give a global overview of the similarity.

The fifth model (#5) is very similar to the query model (#1), and the shape of its parts differs only for the shafts and the main flanges which have a border thicker than the query one. However, in this model the screws and the nuts present clearances, which means that these components are not in contact. The fact that not all the components are matched is reflected by the partial and global measures that are lower than the local one. This difference indicates that the matched parts are very similar, but do not cover all the query model and neither the target model. This is confirmed analyzing the values of the single similarity measures. As expected, the values of joint, position and structure are very high, while the value of the shape similarity highlights small differences in the matched components. The most significant variation is in the number of matched components. In this example, both the query and the target models have fourteen parts and eight of them are matched, then $PCF = GCF = 0.57$ and these values affect negatively partial and global measures. In addition, what hinders a full match is the type of contact. In particular, the fours nuts in the target model are not in contact with the screws, and the key and the shaft present different contacts. Indeed, the key and the shaft in the query model are in contact by three planar faces, thus a translation is allowed, while in the target model the two parts are in contact by four planar faces and no motion is possible.

Finally, from the values of the different levels of similarity measures, the user can easily understand that the sixth model (#6) is not suited for design reuse. Indeed, the partial similarity measure is very low indicating that many elements of the query model have not been matched. Actually, this model has been incorporated in the analysis to demonstrate how the measures can help discarding models not similar to a query one.

6.2. Planetary gearbox assembly models retrieval

The second experimentation aims at retrieving planetary gearboxes contained in the database. This functional set has several sun-planet and ring-planet gear pairs as the first model (#1) in Table 3, which has been used as query model. In this example, the similarity criteria for the nodes c_N are relaxed while considering the component type and the pattern type. The similarity criteria for the arcs c_A require that two pairs of similar parts should have the same number of allowed rotations and translations. Here, the weights used to compute γ_ℓ, γ_p and γ_g are set up to $w_{gear} = \{1, 1, 1, 0\}$. Thus, the weight $w^{st} = 0$ which means that the $\mu^{structure}$ similarity measure is not considered to computed the overall similarity values (global, partial and local). Note that not considering the structure similarity criterion does not imply not evaluating it. Indeed, the similarity criteria are used to build the association graph, while the similarity evaluation allows a ranking of the retrieved solutions.

In general, for all the retrieved models in Table 3, one can observe that the global similarity measure is much lower than the others. This suggests that the query model (#1) is included in the target models. Of course, the first model being the query model, all its measures have maximum values since it perfectly matches itself. The second model (#2) has high values of local and partial similarities, this indicates that the single similarity values are high and that the entire query is included in the target model. Indeed, its similarity measures have almost maximum values and only the global measure is low, due to the fact that the query model is entirely included in a bigger target model.

For the third retrieved model (#3), the matched components are the three gears and three axes, whose shapes are similar to the ones in the query model, then μ^{shape} has an high value. The measure μ^{joint} is equal to zero since no contacts are present between the matched components. See that all the components are disconnected, there is no variation of incident rotation/translation to compare. This affects negatively the final value of the local similarity measure when the weight w^{joint} is not null. Anyhow, even if it would be technically possible, assigning $\mu^{joint} = 1$ when no contact is present could be misleading for the user in his/her model analysis.

For the fourth model (#4), the matched components are also the three gears and three axes, but differently from model (#3), the gears are modeled in a simplified form and they are recognized thanks to the attribute *CompType* of the EAM, which identifies three simple rings as gears exploiting the surrounding context of the components. Note that including shape compatibility, and depending on the chosen shape similarity threshold, this configuration would be probably not retrieved since the shapes of the planar gears are quite different from the ones proposed in the query model.

For the last two models (#5) and (#6), the same considerations as for the third model hold, i.e. three gears and three shafts are retrieved whose shapes are similar to the ones of the query model (#1). The measure μ^{joint} is also zero since there is no contact between the retrieved parts.

7. Conclusion and perspectives

This work proposed a system to evaluate the similarity of assembly models according to multiple criteria and levels of similarity. This goal is achieved using explicit information encoded in the STEP descriptors of CAD models, as well as implicit information to be extracted. The collected data are represented in an attributed multi-graph structure and the similarity is recognized detecting the maximum common sub-graph between two graphs representing the assembly models. Since two models can be similar according to different criteria, in this paper, several measures have been introduced to evaluate different aspects of assembly similarity. In particular, the shape, joint, position and structure similarity measures have been defined. The experimentations and results confirm the usefulness of the information extracted from CAD models and stored in the EAM files for their similarity assessment. An ad-hoc visualization interface has been designed and used to enhance the user experience when analyzing the results. In the future, other measures could be defined. For instance, a size measure could assess the similarity of two models according to their dimensions. In the current implementation, the size is only involved in creation of the association graph.

Different levels of similarity (global partial and local) have also been defined. The proposed method is also able to retrieve locally similar assemblies whose matched components are disconnected. This is possible since in the definition of the association graph, two association nodes are connected if they have the same relationship, where ”same relationship” indicates also that both the original pairs of nodes are not in contact.

So far, the weights used to combine the set of measures to compute the similarity are set by the user. To discover which weights best fulfill the user requirements according to different use scenarios, it is necessary to investigate how weight combinations affect the final score and, most of all, it is essential to include the user feedback to validate which results are considered pertinent for the specific query. In the end, using the reciprocal comparison of all the models in the dataset can facilitate the browsing by visualizing similar models according to all the measures defined.

References

[1] Bai, J, Gao, S, Tang, W, Liu, Y, Guo, S. Design reuse oriented partial retrieval of CAD models. Computer-Aided Design 2010;42(12):1069 – 1084.

[2] Brière-Côté, A, Rivest, L, Maranzana, R. Comparing 3D CAD models: Uses, methods, tools and perspectives. Computer-Aided Design and Applications 2012;9(6):771–794.

[3] Deshmukh, AS, Banerjee, AG, Gupta, SK, Sriram, RD. Content-based assembly search: A step towards assembly reuse. Computer-Aided Design 2008;40(2):244–261.

[4] Zhen-Ba, L, Shu-Hui, B, Kun, Z, Shu-Ming, G, Jun-Wei, H, Jun, W. A survey on partial retrieval of 3D shapes. Journal of Computer Science & Technology 2013;28(5):836–851.

[5] Ferreira, A, Marini, S, Attene, M, Fonseca, MJ, Spagnuolo, M, Jorge, JA, et al. Thesaurus-based 3d object retrieval with part-in-whole matching. International Journal of Computer Vision 2010;89(2-3):327–347.

[6] Iyer, N, Jayanti, S, Lou, K, Kalyanaraman, Y, Ramani, K. Shape-based searching for product lifecycle applications. Computer-Aided Design 2005;37(13):1435–1446.

[7] Tangelder, JW, Veltkamp, RC. A survey of content based 3D shape retrieval methods. Multimedia tools and applications 2008;39(3):441–471.

[8] Iyer, N, Jayanti, S, Lou, K, Kalyanaraman, Y, Ramani, K. Three-dimensional shape searching: state-of-the-art review and future trends. Computer-Aided Design 2005;37(5):509 – 530.

[9] Cardone, A, Gupta, SK, Karnik, M. A survey of shape similarity assessment algorithms for product design and manufacturing applications. Journal of Computing and Information Science in Engineering 2003;3(2):109–118.

[10] Osada, R, Funkhouser, T, Chazelle, B, Dobkin, D. Shape distributions. ACM Transactions on Graphics (TOG) 2002;21(4):807–832.

[11] Renu, R, Mocko, G. Retrieval of solid models based on assembly similarity. Computer-Aided Design and Applications 2016;13(5):628–636.

[12] Wang, P, Li, Y, Zhang, J, Yu, J. An assembly retrieval approach based on shape distributions and Earth Mover's distance. The International Journal of Advanced Manufacturing Technology 2016;86(9-12):2635–2651.

[13] Zhang, J, Pang, J, Yu, J, Wang, P. An efficient assembly retrieval method based on hausdorff distance. Robotics and Computer-Integrated Manufacturing 2018;51:103–111.

[14] Miura, T, Kanai, S. 3D shape retrieval considering assembly structure. In: Proceeding of Asian Symposium for Precision Engineering and Nanotechnology 2009 (ASPEN 2009). 2009, p. 11–13.

[15] Tao, S, Huang, Z. Assembly model retrieval based on optimal matching. In: Software Engineering and Knowledge Engineering: Theory and Practice. Springer; 2012, p. 327–336.

[16] Chen, X, Gao, S, Guo, S, Bai, J. A flexible assembly retrieval approach for model reuse. Computer-Aided Design 2012;44(6):554–574.

[17] Han, Z, Mo, R, Yang, H, Hao, L. CAD assembly model retrieval based on multi-source semantics information and weighted bipartite graph. Computers in Industry 2018;96:54–65.

[18] Lupinetti, K, Giannini, F, Monti, M, Pernot, JP. Multi-criteria retrieval of CAD assembly models. Journal of Computational Design and Engineering 2018;5(1):41 – 53.

[19] Lupinetti, K, Chiang, L, Giannini, F, Monti, M, Pernot, JP. Regular patterns of repeated elements in CAD assembly model retrieval. Computer-Aided Design and Applications 2017;14(4):516–525.

[20] Lupinetti, K, Giannini, F, Monti, M, Pernot, JP. Automatic extraction of assembly component relationships for assembly model retrieval. Procedia CIRP 2016;50:472–477.

[21] Lupinetti, K, Giannini, F, Monti, M, Pernot, JP. Identification of functional components in mechanical assemblies. Procedia CIRP 2017;60:542–547.

[22] Lupinetti, K, Giannini, F, Monti, M, Pernot, JP. Identification of functional sets in mechanical assembly models. In: ICIDM2017 conference (Milan, Italy, 2017 July 17-19). 2017,.

[23] Kazhdan, M, Funkhouser, T, Rusinkiewicz, S. Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Symposium on geometry processing; vol. 6. 2003, p. 156–164.

[24] Conte, D, Foggia, P, Sansone, C, Vento, M. Thirty years of graph matching in pattern recognition. International journal of pattern recognition and artificial intelligence 2004;18(03):265–298.

[25] Bomze, IM, Budinich, M, Pardalos, PM, Pelillo, M. The maximum clique problem. In: Handbook of combinatorial optimization. Springer; 1999, p. 1–74.

[26] Bron, C, Kerbosch, J. Algorithm 457: finding all cliques of an undirected graph. Communications of the ACM 1973;16(9):575–577.

[27] Shilane, P, Min, P, Kazhdan, M, Funkhouser, T. The princeton shape benchmark. In: Shape Modeling Applications, 2004. Proceedings. IEEE; 2004, p. 167–178.

[28] Regli, WC, Foster, C, Hayes, E, Ip, CY, McWherter, D, Peabody, M, et al. National design repository project: a status report. In: International Joint Conferences on Artificial Intelligence (IJCAI) and AAAI/SIGMAN Workshop on AI in Manufacturing Systems, Seattle, Washington, USA. 2001,.

[29] Jayanti, S, Kalyanaraman, Y, Iyer, N, Ramani, K. Developing an engineering shape benchmark for CAD models. Computer-Aided Design 2006;38(9):939–953.

[30] Hu, KM, Wang, B, Yong, JH, Paul, JC. Relaxed lightweight assembly retrieval using vector space model. Computer-Aided Design 2013;45(3):739–750.

[31] Qin, F, Gao, S, Yang, X, Li, M, Bai, J. An ontology-based semantic retrieval approach for heterogeneous 3D CAD models. Advanced Engineering Informatics 2016;30(4):751–768.

Recent titles from the IMATI-REPORT Series:

2018

18-01: *Arbitrary-order time-accurate semi-Lagrangian spectral approximations of the Vlasov-Poisson system*, L. Fatone, D. Funaro, G. Manzini.

18-02: *A study on 3D shape interaction in virtual reality*, E. Cordeiro, F. Giannini, M. Monti, A. Ferreira.

18-03: *Standard per la gestione e procedure di validazione di dati meteo da sensori eterogenei e distribuiti,* A. Clematis, B. Bonino, A. Galizia.

18-04: *Strumenti e procedure per la gestione e la visualizzazione di dati meteo prodotti da sensori eterogenei e distribuiti tramite interfacce web basate su mappe geografiche,* L. Roverelli, G. Zereik, B. Bonino, A. Gallizia, A. Clematis.

18-05: *TopChart: from functions to quadrangulations*, T. Sorgente, S. Biasotti, M. Livesu, M. Spagnuolo.

18-06: *Adaptive sampling of enviromental variables (ASEV*), S. Berretta, D. Cabiddu, S. Pittaluga, M. Mortara, M. Spagnuolo, M. Vetuschi Zuccolini.

18-07: *Multi-criteria similarity assessment for CAD assembly models retrieval*, K. Lupinetti, F. Giannini, M. Monti, J,-P. Pernot.