

# A FRAMEWORK FOR VISUALIZATION-DRIVEN ECO-CONSCIOUS DESIGN EXPLORATION

Devarajan Ramanujan<sup>1\*</sup>, William Z. Bernstein<sup>1</sup>, William Benjamin<sup>1</sup>, Karthik Ramani<sup>1,2 †</sup>

<sup>1</sup>School of Mechanical Engineering

<sup>2</sup>School of Electrical and Computer Engineering  
Purdue University  
West Lafayette, Indiana 47907

**Niklas Elmqvist**

College of Information Studies  
University of Maryland, College Park  
College Park, Maryland 20742

**Devadatta Kulkarni, Jeffrey Tew**

Cincinnati Innovation Lab  
Tata Consultancy Services  
Milford, Ohio 45150

*A large portion of design activity involves reuse of previous knowledge in order to solve new problems. Therefore, facilitating eco-conscious exploration of archived designs is needed for advancing sustainable product design. It is thus necessary to create integrated exploration tools that share common data representations for design and sustainability-related product metadata. This can allow designers to observe co-variations in design data and develop engineering intuition with regards to environmental sustainability performance. In this work, we present a framework for relating sustainability and product metadata using taxonomy-based representations of lifecycle data. This facilitates simultaneous visualization of environmental indicators along with part similarities. To demonstrate this framework, we implement shapeSIFT, an interactive multi-dimensional visualization tool for eco-conscious design exploration. ShapeSIFT uses a visual analytics-based approach to represent part metadata and environmental indicators. This facilitates query-based dynamic exploration of part repositories.*

## Nomenclature

$EI$ : net cradle-to-gate environmental impact calculated using the cumulative energy demand (CED) method

from Ecoinvent version 1.01 within SimaPro®  
 $e$ : environmental impact associated with the unit process for material extraction  
 $b_v$ : blank/initial volume of material used for manufacturing the specified part  
 $p_i$ : environmental impact associated with the  $i^{th}$  unit manufacturing process. Note that this quantity is also dependent on the type of material that is manufactured  
 $O_i^{dim}$ : operating dimension associated with the  $i^{th}$  manufacturing process  
 $n$ : total number of unit manufacturing processes for a part  
 $\Delta EI^{ext}$ : error in estimating the cradle-to-gate impact for the material extraction stage  
 $\Delta EI_i^{rem}$ : error in estimating the cradle-to-gate impact for the  $i^{th}$  mass removal process  
 $\Delta EI_i^{con}$ : error in estimating the cradle-to-gate impact for the  $i^{th}$  mass conserving process  
 $\Delta EI$ : cumulative error in estimating the net cradle-to-gate impact for a part  
 $V_{mesh}$ : part volume as calculated from the mesh file  
 $D(a_1, a_2)$ : distance measure between nodes  $a_1$  &  $a_2$   
 $d_{lca}(a_1^*, a_2^*)$ : depth (measured from the root) of the lowest common ancestor for nodes  $a_1^*$  and  $a_2^*$   
 $d_{pt}(a_1^*, a_2^*)$ : length of traversed path (number of hops) to reach node  $a_2^*$  from node  $a_1^*$ . For computing  $d_{pt}$ , we allow edge traversals in all directions

\*Address all correspondence [dev@purdue.edu](mailto:dev@purdue.edu)

†Address all correspondence for other issues to [ramani@purdue.edu](mailto:ramani@purdue.edu)

- $m$ : a material type specified using the Ashby taxonomy [1]
- $r$ : a manufacturing process plan consisting of an ordered set of individual processes specified using the Allen and Todd taxonomy [2]
- $f$ : a function description represented by an unordered set of individual functions specified using the categorization by Hirtz et al. [3]
- $d_m(m_1, m_2)$ : scalar distance between materials  $m_1$  &  $m_2$
- $d_r(r_1, r_2)$ : scalar distance between manufacturing process plans  $r_1$  &  $r_2$
- $d_f(f_1, f_2)$ : scalar distance between function descriptions  $f_1$  &  $f_2$
- $S$ : a pixel-based representation for an image, used for matching a sketch input to the 2D projection of the part
- $x$ : histogram-based representation of an image  $S$
- $d_s(x_1, x_2)$ : scalar distance between two image histograms  $x_1$  &  $x_2$

## 1 INTRODUCTION

Life Cycle Assessment (LCA) has become one of the most objective methods for quantifying the environmental sustainability of a product or a process. Even so, conducting a detailed LCA for every part archived in a repository is impractical due to the amount of time and resources required. A more manageable approach is to simplify the assessment process by: (1) reducing the scope of the analysis, or (2) approximating materials and manufacturing by available alternatives from a standardized inventory [4]. Such approximations can generate environmental indicators that can be useful for screening out designs with a significant negative effect on the environment [5]. An important consideration for an environmental indicator to be applicable to product design is automating the estimation process. This allows scaling the assessment process to repositories that contain a large number of designs. Previous research [6] has outlined challenges in this context, such as (1) the availability of accurate mass and volume data, (2) levels of specificity in the description of material and manufacturing processes, and (3) estimations of part features relevant to process. These challenges prevent automation as well as a more accurate computation of life cycle impact. In this work, we overcome these challenges through a taxonomy-based description for part attributes.

The primary contribution of this paper is a framework that enables sustainability-aware exploration and reuse of previous design knowledge in a 3D part repository. For this, we develop taxonomy-based representations for design metadata that allow automated estimation of (1) environmental indicators and (2) part similarities. We also present a visualization approach for creating a user interface that allows querying and exploring the part repository. Our approach also enables visualization of environmental indicators and design metadata by encoding them as “visual variables” [7, p. 42]. Visualizing such data via multi-dimensional representations allows designers to develop engineering intuitions about *impact-metadata* relationships that affect the design process.

## 2 BACKGROUND

A significant challenge for estimating life cycle impacts of designs archived in repositories is clearly defining the nature and the quality of archived data. Although efforts such as the National Institute of Standards and Technology repository [8] and the University of Missouri-Rolla design repository [9] have been successful at defining data standards for design information, they are not designed to contain an exhaustive lifecycle inventory (LCI) that can be used for environmental impact assessment. We review previous research pertaining to (1) methods for estimating eco-indicators using knowledge from existing parts, (2) similarity assessment in design repositories, and (3) the role of information visualization in environmental sustainability.

### 2.1 Estimating eco-indicators from existing part data

Previous research in sustainable design has looked at bridging gaps in lifecycle related information using techniques that leverage implicit knowledge embedded in existing parts. Approaches proposed by researchers include (1) using surrogate measures of environmental impact, (2) developing indices that relate environmental impact to part attributes, and (3) extrapolating impact on the basis of existing similar products. Sousa et al. [10] develop a method for generating approximate life cycle assessment (LCA) metrics through neural networks trained using pre-existing product attributes. On similar lines, a knowledge-based approximate life cycle assessment system (KALCAS) is discussed by Park et al. [11]. Dewulf et al. [12] detail Eco-Pas, a methodology that uses “eco-cost estimating relationships” for anticipative weak point analysis of a product’s environmental impact. Huang et al. [13] establish a life cycle performance index for eco-conscious redesign. This index is used for selecting an environmentally benign material that still meets strength and stiffness constraints. An approach for estimating life cycle impacts by correlating them with product functions has been detailed by Devanathan et al. [14]. Here, the authors develop the “function-impact matrix” that associates impact embodied by a structure to its corresponding function. This allows designers to look at less impactful embodiments for realizing a specific function. Another method for impact estimation based on functional modeling of similar existing products is detailed by Haapala et al. [6]. The estimation method discussed is scalable to large design repositories. Tagged product attributes, such as material, manufacturing processes, and mass, are used for estimating impact. However, this method does not use information contained in 3D part models nor does it allow variable levels of data specification with regards to categorical product attributes, such as material or manufacturing processes.

### 2.2 Similarity assessment in design repositories

Estimating similarity in design repositories necessitates an understanding of design intent by means of coding semantic information with regards to artifact features. This is particularly true for categorical metadata, such as material, manufacturing, or function descriptions, which cannot be quantified on any particular measurement scale. Iyer et al. [15] discuss a context-based inference system to capture

design intent from legacy CAD, such as 2D drawings and 3D models. The authors extract raw data from such systems, classify them, and convert them to design intent using an interactive inference system. A prevalent method for establishing context is the classification of artifact features either in the form of ontologies or taxonomies. This categorization can be used for establishing a non-ambiguous mapping between features. Li et al. [16] discuss a method for ontology-based retrieval of design information. The authors use natural language processing and domain-specific ontologies to construct a semantics-based representation from unstructured design documents. The constructed representation is used for indexing and retrieval of design information. Quay et al. [17] propose a hierarchical data organization where users can select regions of spatial, temporal, or topical interests. While these methods characterize parts based on design attributes and/or 3D models, our goal is to explore repositories based on a combination of shape, environmental sustainability, and other relevant part attributes. This multi-dimensional approach to design exploration is essential for designers to generate insights about relationships between sustainability metrics and traditional design variables.

### 2.3 Visualization in environmental sustainability

Visualization can be defined as the use of computer-supported, interactive, visual representations of abstract data to amplify cognition [18]. The primary objective in data visualization is to allow users to gain insight into an information space by mapping data onto graphical primitives [19]. Visualization can act as a powerful enabler of environmental sustainability by its ability to *make data transparent*. This includes means for (1) generating awareness about specific data, (2) making design exploration more intuitive, and (3) facilitating better decision-making by emphasizing trends and correlations for sustainability-related data. Creative real-time visualizations that quantify energy consumption and carbon loads have been used to promote resource conservation [20]. Developing meaningful visualization of sustainability indicators presents a challenge due to its high dimensionality. An interface for visualizing the QUEST environmental sustainability model is presented by Munzner et al. [21]. The authors provide insights into the successes and challenges in designing visualization schemes required for engaging communities in environmental policy making. An additional requirement for a visualization scheme applicable to 3D repositories is the ability to query and convey shape information. Pousman et al. [22] discuss integration of sustainability-related visualizations for paper printing. The primary goal of their work is to motivate conversations among community members. Providing feedback on individual/group behavior for reducing environmental impact is detailed by Froehlich et al. [23]. Marwah et al. [24] discuss reducing the energy consumption and carbon load of data centers. The authors provide cases that use visualizations of sensor data (e.g. temperature, power load) to understand trends and anomalies in daily operation. While the mentioned efforts are clearly directed towards knowledge discovery and decision-making in the context of sustainabil-

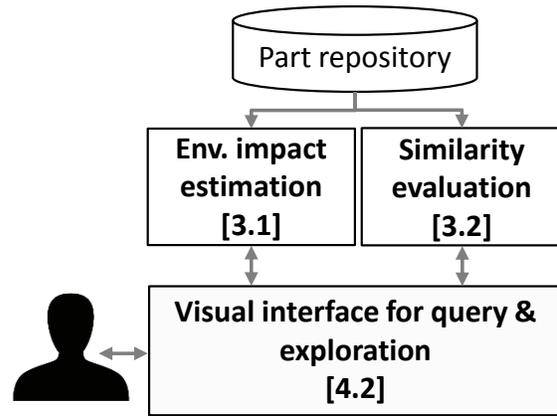


Fig. 1. Overview of our framework for supporting sustainability-aware design exploration in 3D part repositories. Components within the pipeline along with their section numbers are shown in bold font.

ity, prior research has not tied such efforts with design exploration. Similarly, visualization-based methods for design exploration in 3D repositories [25, 26] do not emphasize on sustainable design. In this paper, we develop a framework for bridging these domains to facilitate sustainability-aware design exploration.

### 3 METHODOLOGY

A high level overview of our framework is illustrated in Fig.1. The primary interaction mode for users of our framework is query-based exploration of part similarities. A range of visualizations can be designed to guide these processes. An interface with one such visualization scheme is discussed in this paper. We start this discussion by detailing the methodology behind the three core modules in our framework, namely the (1) environmental impact assessment module, (2) similarity evaluation module, and (3) visualization module and prototype interface. Applying our framework requires access to 3D part data with metadata regarding material, process planning, and functionality. Feature level information is often absent in existing repositories. Therefore, we work with 3D part repositories that do not contain a direct mapping of manufacturing processes to specific part features. For example, if the process plan has two material removal operations such as *milling* and *turning*, we cannot estimate how much material was removed by *milling* as compared to *turning* in order to produce the final shape. In order to make our framework relatively independent of the representation of data present in a part repository (i.e. file formats and granularity), we use low level representations that can be derived from common high level representations of design data. This approach allows users to adapt the developed framework towards their preferred data schemes. Figure 2 illustrates the data model for a part class contained in our framework. The primary inputs to our framework are (1) a 3D model of the part, represented as a mesh, (2) a material definition, (3) an ordered list of manufacturing processes, (4) a function description of the part, and (5) part identifiers for indexing and

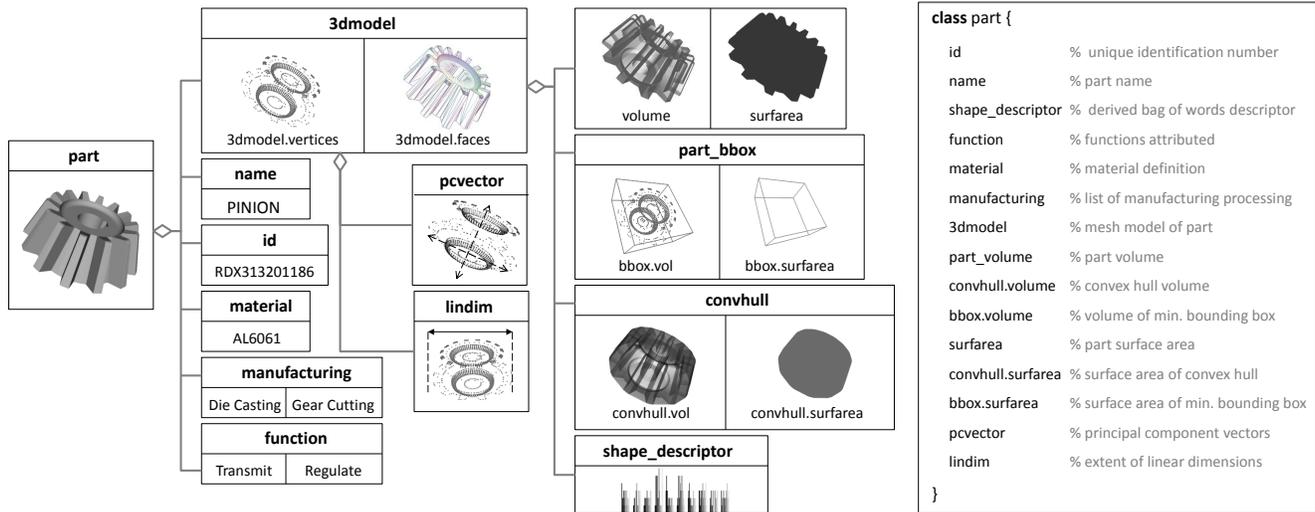


Fig. 2. Data representation model for defining a “part class” in our framework. Here, the arrows represent an aggregation relationship. Metadata contained in the class are either specified as input data during instantiation or subsequently derived from input data. Minimum input data that needs to be specified include (1) the part geometry in the form of a 3D model, (2) the part material, (3) an ordered list of manufacturing processes, (4) part functions, and (5) identifiers for indexing and query.

retrieval. All other part metadata, such as the environmental indicator, shape descriptors, and metadata similarities, are derived from these inputs.

### 3.1 Environmental impact estimation

In this work, we focus on developing an automated indicator for approximating cradle-to-gate impact for mechanical parts. Consequently, our framework is applicable towards parts whose lifecycle impacts are dominated by resource extraction and manufacturing processes. Although this reduction in scope results in higher uncertainties in environmental impact assessment, it is necessary, since information regarding downstream lifecycle stages (i.e. use phase and end-of-life) is rarely available at the design phase. Moreover, our framework is aimed at design-phase exploration with the goal of screening out designs with significant environmental impact.

Given a 3D model of a mechanical part with corresponding metadata, we start by extracting volumetric as well as shape-related data as shown in Fig. 2. This information along with process data is used for estimating the approximate environmental impact of the product. Since a mesh-based representation of the solid model is used, feature level information is unavailable for estimating the cradle-to-gate indicator.

#### 3.1.1 Taxonomy-based representation of lifecycle data

Manufacturing processes are specified in our framework as per the Allen and Todd taxonomy [2]. This taxonomy categorizes processes into 14 major families. This classification taxonomy takes into account workpiece geometry, resulting tolerances, workable materials, and cost. This taxonomy was preferred as the classifications described correspond closely with volumetric information of parts. One of the reasons for incorporating a taxonomy-based specification for manufacturing is the flexibility that it allows in the level of specificity

of a process. For example, a repository might contain a part that is described as being *cast* without further information on the exact nature of the casting process (e.g. die casting, investment casting, and sand casting). However, estimating environmental impact data requires a more specific unit process. In such cases, an approximate measure for environmental impact can be established by averaging the unit impacts of the set of manufacturing process in the induced sub-tree. Similarly, it is possible that unit process information regarding a specific process is unavailable in the used LCI database. Here, we can approximate the resulting impact by substituting it for the most similar manufacturing process that has data available in the LCI. For this, we develop a similarity measure among manufacturing processes based on the structure of the taxonomy.

A corresponding taxonomy for material specification described by Ashby [1] is also incorporated in our framework. Within this scheme, materials are grouped into five classes: ceramics and glasses, fibers and particulates, hybrids, metals and alloys, as well as polymers. Each material class is further classified into multiple material groups. A complete classification scheme is available within the CES Edupack software [27]. In addition to material and manufacturing taxonomies, we implement a function taxonomy adapted from the categorization of functions by Hirtz et al. [3]. Here, the authors develop a reconciled functional basis, where functions are grouped into 8 primary classes. They are further divided into multiple sub-classes. The authors also provide a list of correspondences that allow users to correlate their functional basis with related efforts. By implementing this function taxonomy, we allow designers to compare and filter parts based on similarities in part function.

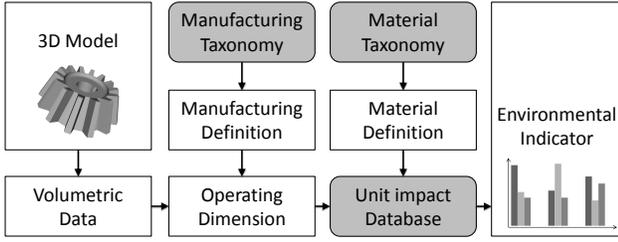


Fig. 3. Pipeline for estimating the environmental impact indicator from a three-dimensional, mesh-based representation of the part and corresponding input metadata. Gray squares represent reference taxonomies and databases used for standardizing data description.

### 3.1.2 Estimating the cradle-to-gate indicator

The pipeline for estimating the environmental indicator is described in Fig. 3. First, we extract volumetric properties from a 3D model of the mechanical part stored in the database. Properties, such as volume, surface area, convex hull volume, and minimum bounding-box volume, are calculated from the stereolithography (.STL) file and indexed. Next, we estimate the operating dimension ( $O^{dim}$ ) for each manufacturing process associated with the part.  $O^{dim}$  is defined as the physical variable pertaining to part geometry (i.e. volume and surface area) that is processed by a manufacturing operation. Table 1 illustrates the definition of  $O^{dim}$  for different kinds of manufacturing processes as per the Allen and Todd taxonomy. Thus, the  $O^{dim}$  for a manufacturing process can be used as a scaling factor on its corresponding unit process. Scaling the impact of a unit process by  $O^{dim}$  results in the net impact of that unit process on the part geometry. In an ideal setting, the operating dimension for each process is specified as input data or encoded as shape changes of the 3D model. Although a well-defined product lifecycle management (PLM) system might also archive such data, most repositories today do not provide any means for obtaining this information. Therefore, in this paper, we estimate the  $O^{dim}$  for a specific manufacturing process based on the following approximations.

- If the volume of the starting stock/blank is not specified, it is taken to be equal to the smaller value of (1) the convex hull volume of 3D part and (2) the volume of the minimum bounding box of the 3D part.
- If there is more than one material removal operation in the list of manufacturing processes, the total removed volume is divided equally among these processes.
- The Allen and Todd taxonomy is used to categorize manufacturing processes into one of the following four types:
  - Mass conserving volumetric (e.g. forging, annealing)
  - Mass reducing volumetric (e.g. turning, drilling)
  - Surficial (e.g. anodizing, electrocoating, dust coating)
  - Joining (e.g. welding, adhesive bonding)

Thus, any process that appears before the first *mass reducing* process always operates on the convex hull volume or surface area. Similarly, any process that occurs after a

Type of Manufacturing Process	Operating Dimension ( $O^{dim}$ )
Mass conserving volumetric	Volume of the part before/after the manufacturing process
Mass reducing volumetric	Volume of the material removed in the manufacturing process
Surficial	Surface area that is coated/transformed by the manufacturing process
Joining	Functional dimension (volume, surface area, length, etc..) depending on to the type of joining process

Table 1. Definition of  $O^{dim}$  based on the manufacturing process.

*mass reducing* process operates on the reduced volume. Although units such as volume and surface area are easily computable from a 3D model, extracting feature level information for calculating the operating dimension for *joining* processes present significant challenges. Therefore, information about the operating dimension (i.e. length of weld, surface area of bonded surfaces) is required to be specified by the user as input to the framework. Once the  $O^{dim}$  for each manufacturing process is estimated, the cradle-to-gate environmental indicator is computed as a linear sum of the impact of material extraction and manufacturing processes (see Eq. 1).

$$EI = e * b_v + \sum_{i=1}^n p_i * O_i^{dim} \quad (1)$$

Approximating the  $O^{dim}$  introduces additional uncertainties in estimating the cradle-to-gate environmental impact computed using Eq. 1. These uncertainties result from a lack of detailed information pertaining to the material extraction and manufacturing stages. As discussed, the availability of detailed lifecycle data in design repositories obviates the need for this approximation. However, we present and discuss uncertainties for a scenario in which feature information (that maps manufacturing process data to part geometry) is absent from the design repository. Apart from inherent uncertainties in the life cycle assessment process, additional approximation errors resulting from our method can be formalized as follows.

- $\Delta V$  : the error resulting from approximating the initial blank. volume  $b_v$  by the convex hull/minimum bounding box volume.
- $\Delta w_i$  : error in removed volume fraction for  $i^{th}$  material removal operation. This results from our approximation that the total removed volume is divided equally among all material removal processes.

$$\Delta EI^{ext} = \Delta V * e \quad (2)$$

$$\Delta EI_i^{rem} = p_i * \left\{ \frac{\Delta V}{n} + \Delta w_i * (b_v - V_{mesh} + \Delta V) \right\} \quad (3)$$

$$\Delta EI_i^{con} = p_i * \left\{ \frac{\Delta V (n-1)}{n} - (b_v - V_{mesh} + \Delta V) \sum_{j=1}^{i-1} \Delta w_j \right\} \quad (4)$$

$$\Delta EI = \Delta EI^{ext} + \sum_{i=1}^n \left\{ H(\Phi_i) * \Delta EI_i^{rem} + H(-\Phi_i) * \Delta EI_i^{con} \right\} \quad (5)$$

$H(\Phi)$  is the Heaviside step function.

$\Phi_i = 1$  if the  $i^{th}$  process is volumetric & mass removing

$\Phi_i = -1$  if the  $i^{th}$  process is volumetric & mass conserving

Equations 2, 3, 4, and 5 represent a closed form solution for the cumulative error in estimating  $EI$  due to approximations in estimating  $O^{dim}$ . These equations are derived by substituting the error terms in Eq. 1. Please note that uncertainties with respect to surficial and joining processes are not considered in these equations due to the dependencies of these errors on the shape of a specific part. In this paper, we use Eq. 1 to compute a cradle-to-gate indicator for the purpose of demonstrating our visualization pipeline. For this, the Cumulative Energy Demand (CED) is used as an indicator of environmental impact. Cumulative Energy Demand for a product is defined as the total quantity of primary energy needed to produce, use, transport and dispose of that particular product. Previous literature has outlined the usefulness of CED to serve as a screening indicator for environmental performance [28]. A lookup table is hard-coded into our system that contains CED values of unit processes for material extraction as well as a given material-manufacturing process combination. The data for these entries have been referenced from the methods library available through SimaPro 7.1 [29]. Our current setup is also capable of estimating cradle-to-gate impacts based on the Eco-Indicator 99 method referenced in SimaPro 7.1. Developing a more holistic indicator is possible if data concerning the transportation, use-phase and end-of-life is made available within the repository. Future implementations of the shapesIFT tool will look at providing a choice for multiple impact estimation methods. In this paper, we limit our focus towards enabling comparisons between a computed environmental indicator and product attributes.

### 3.2 Similarity evaluation

A natural way of quantifying similarity between elements of a set is by establishing a measure of similarity/distance between them. The similarity between two objects is a function of the commonality and the differences they share [30]. We capture these properties using a distance function  $d : \varepsilon \times \varepsilon \rightarrow \mathfrak{R}$  that operates on elements of a taxonomy  $\varepsilon$  and returns a real valued ( $\in \mathfrak{R}$ ) distance measure.

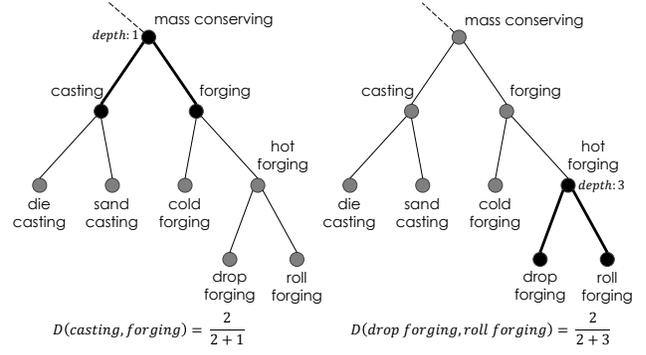


Fig. 4. An example subtree from the Allen and Todd taxonomy [2] for manufacturing processes. This figure illustrates computation of pairwise dissimilarities among manufacturing processes using Eq.6. We can see that this distance measure accounts for hierarchies as it allocates a decreasing value of dissimilarity to a pair of siblings lower down the taxonomy tree. Here, we illustrate that  $D(\text{drop forging}, \text{roll forging}) < D(\text{casting}, \text{forging})$  because the former pair of siblings are at a lower depth.

Although we do not strictly enforce the distance function  $d$  to meet the required conditions to be defined as a metric, we develop a function that possesses the following properties:

1. Non negativity :  $d(e_1, e_2) \geq 0; \{e_1, e_2\} \in \varepsilon$
2. Symmetry :  $d(e_1, e_2) = d(e_2, e_1)$
3. Identity :  $d(e_1, e_2) = 0 \Leftrightarrow e_1 = e_2$

We begin the discussion on similarity computation by defining the involved terms. All mechanical parts are considered to be elements of a set  $\rho$ , with associated materials  $m \in M$ , manufacturing processes  $r \in R$ , functions  $f \in F$ , and a specific shape  $s$ . Here,  $M$ ,  $F$ , and  $R$  are the respective taxonomies adopted to represent these attributes. A manufacturing process  $r \in R$  is treated as an operator  $r : \rho \times \rho \rightarrow P$  such that it operates on a certain part and returns another part with either same or different material and shape properties. Thus, the entire sequence of manufacturing processing can be viewed as a composition of operators that transform an initial blank  $P_0\{m_0, s_0, f_0\}$  to the final part  $P_n\{m_n, s_n, f_n\}$ . The material, manufacturing, function, and shape definition represent significant decisions towards framing design intent. Therefore, we interpret the similarity among parts as a composition of similarities in these four attributes. For this, we define a set of distance functions  $\{d_m, d_f, d_r, \text{ and } d_s\}$  associated with these attributes respectively. Since material, manufacturing and function definitions are represented using corresponding taxonomies, we develop a generalized similarity measure that can be adapted to taxonomies. The distance function for shape is defined using similarities in “shape features” outlined by Squire et al. [31].

#### 3.2.1 Material, manufacturing, and function similarity

Classification trees and taxonomies increase in specificity as we proceed lower down the hierarchy. Therefore, a pair of siblings at a lower level are more similar than siblings higher than them. For example, in a manufacturing taxonomy, any two types of milling processes are more similar to

each other than any two mass reducing processes. Exploiting this property for similarity computation requires making use of the hierarchal nature of the taxonomy. The distance measure discussed in this paper builds on concepts described in Ganesan et al. (2003) [32] and applies them towards the used material, manufacturing, and function taxonomies. Given any two elements in a taxonomy, we calculate a distance measure as follows:

- *Tree Depth Equalization*: When computing the similarity between any two elements of the same tree, only elements at the same depth from the root are evaluated. This is necessary to account for the variation in levels of input specificity. For example, as shown in Fig. 4, the difference between two processes such as *casting* (not very specific) versus *drop forging* (more specific) is essentially the difference between *casting* and *forging* (on a similar level of specificity as casting). Thus, the depth equalization step normalizes the specificity of the items being compared ensuring the validity of the distance measure. The algorithm for the same is illustrated below.

```

if depth(a1) > depth(a2)
then a1* = ancestor(a1) at depth(a2) && a2* = a2
else if depth(a2) > depth(a1)
then a2* = ancestor(a2) at depth(a1) && a1* = a1
else a1* = a1 && a2* = a2

```

- *Distance Estimation*: The next step is to calculate the numerical value of similarity between the entities substituted in the first step. Our distance function is based on the generalized vector-space model discussed in Ganesan et al. [32, p. 71]. We focus on illustrating the applicability of this distance function to material, manufacturing and function taxonomies by demonstrating its hierarchy preserving behavior on the Allen and Todd taxonomy [2]. The corresponding distance function is defined in Eq.(6).

$$D(a_1, a_2) = \frac{d_{pl}(a_1^*, a_2^*)}{d_{pl}(a_1^*, a_2^*) + d_{lca}(a_1^*, a_2^*)} \quad (6)$$

As both  $d_{pl}$  and  $d_{lca}$  lie in the interval  $[0, \infty)$ , the distance measure  $D$  is confined to the interval  $[0, 1]$ . However, when  $d_{pl} = d_{lca} = 0$ , the similarity measure is indefinite. These cases occur only when comparisons are made among elements of taxonomy and its root. As these comparisons do not hold any meaning, we exclude them from the set of allowable comparisons. It can be easily verified that this distance function satisfies the non-negativity, symmetry and identity conditions mentioned earlier. The distance between two elements in a taxonomy  $D(a_1, a_2)$  is equal to 1 only if  $d_{lca} = 0$ . In other words, two elements in the taxonomy are considered to be entirely dissimilar if their lowest common ancestor is the root node of the taxonomy.

Figure 4 illustrates the application of the distance measure to an example subtree. Here, the distance between *casting* and *forging* is  $2/3$  which is greater than  $2/5$ ; the distance

between *drop forging* and *roll forging*. This shows that the distance function accounts for the hierarchical structure of the taxonomy while calculating similarities. Given that we have established a method to compute pairwise-similarities between any two nodes in a taxonomy, we proceed to define our method for composing a scalar distance measure for the specified material, manufacturing, and function definitions. The definition of a part attribute can consist of a single element or, in other cases, a set of elements from the taxonomy. Additionally, the ordering of the associated elements may hold significance in cases such as the definition of a manufacturing process plan. Hence, we develop a measure of each of the attributes that encodes dissimilarity as the maximum deviation of one set of attributes from the other.

In our repository, each part is associated with a single material type. Therefore, for any two materials  $m_1, m_2 \in M$ , the distance function,  $d_m$ , is directly given by the function operating on the material taxonomy as shown in Eq. 7. A manufacturing description  $r = \langle e_{r1}, e_{r2}, \dots, e_{rn} \rangle$ ,  $r \in R$  is considered as an ordered n-tuple of manufacturing processes. Given two manufacturing descriptions,  $r_1$  and  $r_2$ , we define a set  $r_1 \circ r_2$  whose elements are 2-tuples formed by the element-wise product of  $r_1$  and  $r_2$ . The reason behind performing an element-wise operation is that, given two manufacturing descriptions it only makes sense to compare primary production processes with other primary processes, secondary processes with other secondary processes and so on. For example, consider two parts with the following process plans:  $\{\text{casting}, \text{annealing}\}$  and  $\{\text{forging}, \text{nitriding}\}$ . Comparing a primary process of one part (*casting*) with a surface treatment process of another (*nitriding*) will wrongly indicate that the process plans for the two parts are highly dissimilar. Instead, comparing primary processes separate from secondary processes provides a more meaningful measure. In cases where the cardinalities of  $r_1$  and  $r_2$  are different, we restrict the similarity computation to the first  $n$  elements, where  $n$  is the lower of the two cardinalities. The distance function  $d_r$  is defined as the maximum possible value of dissimilarity among the sets of descriptions as given in Eq. 8. A function description  $f = \{e_{f1}, e_{f2}, \dots, e_{fn}\}$  is considered as a set of functions wherein the ordering of the elements are immaterial. Like  $d_m$ , the dissimilarity between two sets of function descriptions is governed by the maximum possible value of dissimilarity among the descriptions. Given two sets of function descriptions,  $f_1$  and  $f_2$ , the distance function  $d_f$  is detailed in Eq. 9. Here,  $f_1 \times f_2$  represents the Cartesian product of the sets  $f_1$  and  $f_2$ . Unlike the manufacturing description, we choose to compare all possible function pairs because there is no concept of *function ordering* in our definition scheme.

$$d_m(m_1, m_2) = D(m_1, m_2) \quad (7)$$

$$d_r(r_1, r_2) = \max(D(r_1 \circ r_2)) \quad (8)$$

$$d_f(f_1, f_2) = \max(D(f_1 \times f_2)) \quad (9)$$

### 3.2.2 Estimation of shape similarity

For estimating shape similarity, we convert 3D models into 2D projections of sketch-like renderings using “sugges-

tive contours” [33]. This allows comparing user sketches and images to 3D models in the repository. In this paper, we use the bag-of-features method (BoF) [31] to develop a metric for shape similarity due to its robustness to noise introduced by affine deformations. Previous literature [34, 35] has shown that the BoF method has commendable performance with regards to 2D shape classification and retrieval. The core idea of the BoF method is to represent images as a histogram of occurrences of “visual words”. The procedure for computing shape similarity is described below.

- *Feature Detection:* In this step, we compute locations of interesting features given by computing the “feature points” on the image using the Harris Detector [36]. Finding such discriminative locations helps in identifying differences between shapes.
- *Feature Description:* In this step, we compute patch descriptors for each detected feature using the Scale Invariant Feature Transform (SIFT) [37]. SIFT embeds these features in a high dimensional space by assigning a 128 dimensional descriptor to the features.
- *Quantizing Features using Visual Vocabulary:* The feature descriptors computed using SIFT have high dimensionality and the complexity of computation increases with the number of features that are detected. To reduce some of the involved complexity, we compute a “visual vocabulary” by clustering features in the database.
- *Image Descriptor Generation:* In this step, we transform the image data into a histogram representing a count of occurrences of cluster center matches. Given any two histograms  $x$  and  $y$  that represent two images  $S_x$  and  $S_y$  respectively, a p-norm distance can be computed by Eq. 10.

$$d_s(x, y) = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (10)$$

In this implementation we use a simple  $L^1$  norm by setting  $p = 1$ . Additionally, in the interest of supporting fast retrieval, we use the fast approximate nearest neighbor method [38] to index queries.

Thus, the overall distance between two parts is given by  $\{d_m, d_r, d_f, d_s\}$  which is a set comprising of pairwise distances among corresponding part attributes. Although it is possible to compose a scalar pair-wise distance measure from this set, there is a possibility that reducing the dimensionality of the data might result in excessive loss of similarity information. Interpreting whether two parts are more similar due to similarities in material, function, or any such attribute is largely decided by the context of the application and therefore by the user. Hence, we focus on creating meaningful multi-dimensional information visualization schemes that aid users in exploring the part repository. The main idea of our visualization scheme involves overlaying computed environmental indicators on similarity information of part attributes for enabling sustainability-aware design exploration of part repositories.

## 4 IMPLEMENTATION

Although there are numerous schemes for visualizing sustainability related data, only a handful of them merge these visualizations with the design exploration process. For creating an interface between the two, we develop a list of design goals that are sensitive to needs of the designer.

### 4.1 Design goals

- *Ability to explore product repositories from a design similarity and sustainability perspective.* The process of exploration should allow the user to build engineering intuitions of the relationship between shape, material/manufacturing data, and environmental sustainability.
- *Intuitive Interaction.* One of our goals is to simplify the design exploration process by providing an intuitive means for navigating and searching for alternate design solutions from a given part database.
- *Exploration Support for Design Process.* We posit that human spatial and visual reasoning skills can be leveraged for effective exploration in the design process. An important element within developing intuitive exploration schemes is the use of cognitively prominent visual variables such as variations in shape, size, and color. This allows pre-attentive processing of decision variables, allowing designers to easily narrow down their focus. For developing a solution that meets the above requirements, we develop a visualization scheme for the shapeSIFT tool that contains the elements discussed in Section 4.2.

### 4.2 Interface description

In this section, we discuss the modified interface that was designed based on expert feedback and our learnings from implementing an initial prototype. Implementation details for the initial prototype can be found in our previous work [39]. 3D parts used for constructing our database were obtained from the Engineering Shape Benchmark (ESB) [40]. The ESB contains a total of 479 models in Stereolithography file format. Synthetic data regarding material, manufacturing, and functionality was added to the part data. This database is stored as an Extensible Markup Language (XML) file and is accessed by our interface. We reimplemented the interface on Java using Processing®, an open source programming language that is geared towards visual design. Our interface uses a mutually coordinated, multi-window framework that allows users to customize the size and position of the windows. A screenshot of the shapeSIFT interface is shown in Fig. 5. The interface elements in the implementation of shapeSIFT are detailed below.

- *Sketch-based input:* Sketching is shown to provide a visible graphic memory that facilitates creativity by providing an easily accessible repository of generated ideas and by stimulating building on earlier ideas [41]. Adopting sketching as one of the primary method for query gives us the advantage of utilizing one of the dominant modes of artifact creation among designers. The sketch window (Fig.5-d) implements a canvas for creating a two-dimensional sketch

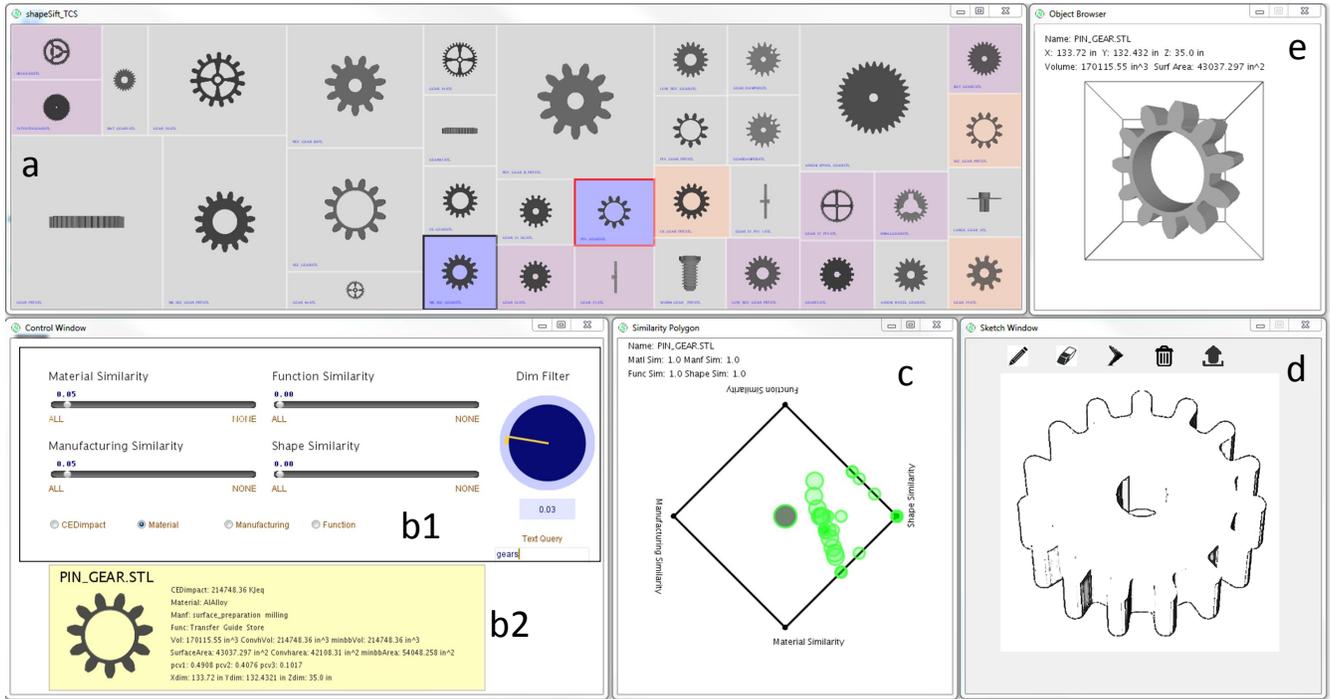


Fig. 5. A screen capture of the modified shapeSIFT interface. It consists of a squarified (tiled) layout window that displays query results (5-a). A control panel (5-b1) is used for setting similarity thresholds for material, manufacturing, function, and shape similarities. A text query box is also provided for users to query part metadata. A label (5-b2) is used to display metadata information and a picture of a particular part that is selected from the squarified layout. The similarity polygon (5-c) allows the user to obtain an understanding of the similarity attributes. The sketch window (5-d) contains a canvas and related controls for creating/modifying the sketch, uploading an image, and querying the repository. The object viewer window (5-e) displays a 3D model of a selected part. A demonstration video can be viewed at <http://goo.gl/talfJm>.

query. Also, users can upload an image onto the sketch canvas. The uploaded image is converted into a sketch-like representation using a neighborhood-based high-pass filter that performs edge detection. On submitting a query, the squarified layout and other windows are automatically updated to reflect the retrieved results.

- *Squarified layout visualization:* Squarified layouts are useful for visually providing a summary of the search results. They can also provide visual cues that allow users to aggregate and discriminate search results. The squarified layout window (Fig.5-a) shows the retrieved results in a layout in which each cell is scaled inversely to the computed environmental indicator (see Eq. 1). Representing the indicator by a prominent visual variable (i.e. size) allows us to nudge designers away from selecting significantly impactful parts. The coloring scheme in this layout encodes a chosen dimension of similarity  $\{d_m, d_r, d_f\}$ . Thus, parts with similar metadata are shown using similar colors. If a particular part does not lie within the threshold set using the control panel, the corresponding cell background is grayed out. Clicking a cell selects the part and displays it on the label, 3D object viewer, and highlights it on the similarity polygon. An additional feature available to the user is changing the reference part for similarity computation from the set of retrieved results by right clicking a cell from this layout.
- *Similarity Polygon:* The similarity polygon visually repre-

sents a barycentric embedding of similarity metadata along chosen attribute dimensions. The similarity polygon window window (Fig.5-c) implements the similarity polygon for the set of retrieved results. Since the similarity polygon uses a barycentric embedding of similarity values for plotting, data points with the same relative weights (e.g.  $\{1,1,1,1\}$  and  $\{0.5,0.5,0.5,0.5\}$ ) are plotted at the same coordinate. This makes it difficult for the user to judge the overall magnitude of the similarity values of a part with respect to the reference part. Therefore, we have implemented a visualization scheme that scales the radius of the circle with the total measure of similarity with respect to the reference part. We have also removed the similarity dimension based on “part class” present in the prototype in favor of a text query box in the control window. Selecting a part from the squarified layout, highlights the corresponding part on the similarity polygon. Conversely, selecting one/multiple parts from the similarity polygon highlights the corresponding cells on the squarified layout using a red border.

- *Sliders for filtering similar results:* Users can also filter results using sliders either by setting individual or multiple thresholds for the set of computed similarities  $\{d_m, d_r, d_f, d_s\}$ . Parts that are dissimilar to the query part in terms of these attributes are grayed out in the squarified layout. The control window implements sliders as well as textual query mechanisms within shapeSIFT (Fig.5-b1). The

control panel contains (1) sliders for setting the similarity thresholds for material, manufacturing, function, and shape similarity, (2) radio buttons that can be used to set the coloring scheme on the squarified layout based on the environmental indicator, material, manufacturing, or function metadata, (3) a dimension filter that screen out parts that are larger or smaller than the reference part in terms of its maximum dimensions, and (4) a text query box that can be used for querying part metadata. A dynamic label (Fig.5-b2) that shows metadata information is also displayed here.

- **3D Model View:** A 3D model of the .STL file associated with the selected part is shown in the object viewer window (Fig.5-e). Users can rotate the displayed model and view the geometric parameters of the selected model on a text label located on the top left of that window.

## 5 CONCLUSIONS AND FUTURE WORK

This paper has presented a new framework for sustainability-aware selection from design repositories through metadata visualization. Part similarities are quantified on multiple dimensions, such as material, manufacturing, and function, using a set of standard taxonomies. Our framework describes methods for automating the computation of environmental impact indicators and similarities in part attributes. This data is visualized using a squarified layout which provides an overview of similar parts and their attributes. We also develop an example interface that allows visualization of part metadata, similarity metrics, and the computed environmental indicator.

Our future work will look into methods for improving the accuracy of environmental indicators and for estimating corresponding uncertainties. For this, we will work towards data representations that can map changes in feature level information in a 3D model to a specific manufacturing step. We will also work towards quantifying and visually representing uncertainties present in impact assessment. Techniques in human-computer interaction, such as perceptually uniform color spaces [42] and human perception of rectangle maps [43], can be useful in solving this problem. We will also work on expert-based evaluation of the shapeSIFT interface to validate its utility with designers in the industry. For this, we plan on evaluating the usefulness of presenting sustainability-based data in a design exploration setting and the usability of the shapeSIFT visualizations and layout. A significant challenge that we plan on addressing in this context is the lack of a real-world design repository with data suitable for eco-conscious design exploration.

## ACKNOWLEDGMENT

This research is supported by the NSF under grants CBET 1512217, DGE 1144843, and the TATA Design Innovation Grant. The contents of this manuscript do not necessarily reflect the views or opinions of the funding agencies.

## References

- [1] Ashby, M., and Cebon, D., 1993. "Materials selection in mechanical design". *Le Journal de Physique IV*, **3**(C7 1–10).
- [2] Todd, R., Allen, D., and Alting, L., 1994. *Manufacturing processes reference guide*. New York:Industrial Press Inc.
- [3] Hirtz, J., Stone, R., McAdams, D., Szykman, S., and Wood, K., 2001. "Evolving a functional basis for engineering design". In Proceedings of the ASME IDETC 2001, Pittsburgh, PA.
- [4] Hunt, R., Boguski, T., Weitz, K., and Sharma, A., 1998. "Case studies examining lca streamlining techniques". *The International Journal of Life Cycle Assessment*, **3**, pp. 36–42.
- [5] Ramani, K., Ramanujan, D., Bernstein, W. Z., Zhao, F., Sutherland, J., Handwerker, C., Choi, J.-K., Kim, H., and Thurston, D., 2010. "Integrated sustainable life cycle design: A review". *Journal of Mechanical Design*, **132**(9), p. 091004.
- [6] Haapala, K., Poppa, K., Stone, R., and Tumer, I., 2011. "Automating environmental impact assessment during the conceptual phase of product design". In AAAI 2011 Spring Symposium: Artificial Intelligence and Sustainable Design, pp. 21–23.
- [7] Bertin, J., 1983. *Semiology of graphics*. Central Asia book series. University of Wisconsin Press.
- [8] Szykman, S., 2002. "Architecture and implementation of a design repository system". In Proceedings of DETC2002, DETC2002/CIE-34463, Montreal, Canada.
- [9] Bohm, M., Stone, R., Simpson, T., and Steva, E., 2008. "Introduction of a data schema to support a design repository". *Computer-Aided Design*, **40**(7), pp. 801–811.
- [10] Sousa, I., Wallace, D., and Eisenhard, J. L., 2000. "Approximate life-cycle assessment of product concepts using learning systems". *Journal of Industrial Ecology*, **4**(4), pp. 61–81.
- [11] Park, J.-H., and Seo, K.-K., 2006. "A knowledge-based approximate life cycle assessment system for evaluating environmental impacts of product design alternatives in a collaborative design environment". *Advanced Engineering Informatics*, **20**(2), pp. 147–154.
- [12] Dewulf, W., and Dufflou, J., 2006. "A web based application for the eco-pas tool". In CIRP - 13th International Conference on Life Cycle Engineering location:Leuven, Belgium, Vol. 1, pp. 143–147.
- [13] Huang, H., Liu, Z., Zhang, L., and Sutherland, J., 2009. "Materials selection for environmentally conscious design via a proposed life cycle environmental performance index". *The International Journal of Advanced Manufacturing Technology*, **44**, pp. 1073–1082.
- [14] Devanathan, S., Ramanujan, D., Bernstein, W., Zhao, F., and Ramani, K., 2010. "Integration of sustainability into early design through the function impact matrix". *Journal of Mechanical Design*, **132**(8), p. 81004.
- [15] Iyer, G. R., Mills, J. J., Barber, S., Devarajan, V., and Maitra, S., 2006. "Using a context-based inference approach to capture design intent from legacy CAD". *Computer-Aided Design & Applications*, **3**(1-4), pp. 269–278.
- [16] Li, Z., and Ramani, K., 2007. "Ontology-based design information extraction and retrieval". *AI EDAM: Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, **21**(02), pp. 137–154.
- [17] Quay, R., and Hutunuwatr, K., 2009. "Visualization of sustainability indicators: A conceptual framework". In *Visualizing Sustainable Planning*. Springer, pp. 203–213.
- [18] Card, S. K., Mackinlay, J. D., and Shneiderman, B., 1999. *Readings in information visualization: using vision to think*. Morgan Kaufmann.
- [19] Senay, H., and Ignatius, E., 1994. "A knowledge-based

- system for visualization design”. *Computer Graphics and Applications, IEEE*, **14**(6), pp. 36–47.
- [20] Holmes, T. G., 2007. “Eco-visualization: combining art and technology to reduce energy consumption”. In Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition, C&C '07, ACM, pp. 153–162.
- [21] Munzner, T., Barsky, A., and Williams, M., 2009. “Reflections on questvis: A visualization system for an environmental sustainability model”. *Scientific Visualization: Interactions, Features, Metaphors*, **2**, pp. 240–259.
- [22] Pousman, Z., Rouzati, H., and Stasko, J., 2008. “Imprint, a community visualization of printer data: designing for open-ended engagement on sustainability”. In Proceedings of the 2008 ACM conference on Computer supported cooperative work, ACM, pp. 13–16.
- [23] Froehlich, J., Findlater, L., and Landay, J., 2010. “The design of eco-feedback technology”. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 1999–2008.
- [24] Marwah, M., Sharma, R., Shih, R., Patel, C., Bhatia, V., Mekanapurath, M., Velumani, R., and Velayudhan, S., 2009. “Data analysis, visualization and knowledge discovery in sustainable data centers”. In Proceedings of the 2nd Bangalore Annual Compute Conference, ACM, p. 2.
- [25] Pu, J., Kalyanaraman, Y., Jayanti, S., Ramani, K., and Pizlo, Z., 2007. “Navigation and discovery in 3d cad repositories”. *Computer Graphics and Applications, IEEE*, **27**(4), July-Aug., pp. 38–47.
- [26] Kim, V. G., Li, W., Mitra, N. J., DiVerdi, S., and Funkhouser, T., 2012. “Exploring collections of 3d models using fuzzy correspondences”. *ACM Trans. Graph.*, **31**(4), July, pp. 54:1–54:11.
- [27] EduPack, C., 2012. *version 11.9.9*. Granta Design, Cambridge, United Kingdom.
- [28] Huijbregts, M. A., Rombouts, L. J., Hellweg, S., Frischknecht, R., Hendriks, A. J., van de Meent, D., Ragas, A. M., Reijnders, L., and Struijs, J., 2006. “Is cumulative fossil energy demand a useful indicator for the environmental performance of products?”. *Environmental Science & Technology*, **40**(3), pp. 641–648.
- [29] Goedkoop, M., Oele, M., Schryver, A., and Vieira, M., 2008. *SimaPro Database Manual-Methods library*. Pré Consultants, Netherlands.
- [30] Lin, D., 1998. “An information-theoretic definition of similarity”. In Proceedings of the 15th international conference on Machine Learning, Vol. 1, San Francisco, pp. 296–304.
- [31] Squire, D., Müller, W., Müller, H., and Pun, T., 2000. “Content-based query of image databases: inspirations from text retrieval”. *Pattern Recognition Letters*, **21**(13), pp. 1193–1198.
- [32] Ganesan, P., Garcia-Molina, H., and Widom, J., 2003. “Exploiting hierarchical domain structure to compute similarity”. *ACM Transactions on Information Systems (TOIS)*, **21**(1), pp. 64–93.
- [33] DeCarlo, D., Finkelstein, A., Rusinkiewicz, S., and Santella, A., 2003. “Suggestive contours for conveying shape”. In ACM Transactions on Graphics (TOG), Vol. 22, ACM, pp. 848–855.
- [34] Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C., 2004. “Visual categorization with bags of keypoints”. In Workshop on statistical learning in computer vision, ECCV, Vol. 1, p. 22.
- [35] Sivic, J., and Zisserman, A., 2006. “Video google: Efficient visual search of videos”. *Toward Category-Level Object Recognition*, pp. 127–144.
- [36] Harris, C., and Stephens, M., 1988. “A combined corner and edge detector”. In Alvey vision conference, Vol. 15, Manchester, UK, p. 50.
- [37] Lowe, D., 2004. “Distinctive image features from scale-invariant keypoints”. *International journal of computer vision*, **60**(2), pp. 91–110.
- [38] Muja, M., and Lowe, D., 2009. “Fast approximate nearest neighbors with automatic algorithm configuration”. In International Conference on Computer Vision Theory and Applications (VISSAPP'09), pp. 331–340.
- [39] Ramanujan, D., Benjamin, W., Bernstein, W. Z., Elmqvist, N., and Ramani, K., 2013. “ShapeSIFT: Suggesting Sustainable Options in Design Reuse From Part Repositories”. In ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, pp. V004T05A041–V004T05A041.
- [40] Jayanti, S., Kalyanaraman, Y., Iyer, N., and Ramani, K., 2006. “Developing an engineering shape benchmark for cad models”. *Computer-Aided Design*, **38**(9), pp. 939–953.
- [41] McKim, R. H., 1972. *Experiences in visual thinking*. Brooks/Cole Pub. Co, Monterey, CA.
- [42] Paschos, G., 2001. “Perceptually uniform color spaces for color texture analysis: an empirical evaluation”. *Image Processing, IEEE Transactions on*, **10**(6), pp. 932–937.
- [43] Kong, N., Heer, J., and Agrawala, M., 2010. “Perceptual guidelines for creating rectangular treemaps”. *Visualization and Computer Graphics, IEEE Transactions on*, **16**(6), pp. 990–998.