Toward an Extensive Data Integration to Address Reverse Engineering Issues

Jonathan Dekhtiar, Alexandre Durupt, Matthieu Bricogne, Dimitris Kiritsis, Harvey Rowson, Benoit Eynard

▶ **To cite this version:**

Jonathan Dekhtiar, Alexandre Durupt, Matthieu Bricogne, Dimitris Kiritsis, Harvey Rowson, et al.. Toward an Extensive Data Integration to Address Reverse Engineering Issues. 13th IFIP International Conference on Product Lifecycle Management (PLM), Jul 2016, Columbia, SC, United States. pp.478-487, 10.1007/978-3-319-54660-5_43 . hal-01699728

HAL Id: hal-01699728
https://inria.hal.science/hal-01699728

Submitted on 6 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Toward an extensive data integration to address reverse engineering issues.

Jonathan Dekhtiar*[1], Alexandre Durupt[1], Matthieu Bricogne[1], Dimitris Kiritsis[2], Harvey Rowson[3] and Benoit Eynard[1]

[1] Université de Technologie de Compiègne, Department of Mechanical Systems Engineering, UMR UTC/CNRS 7337 Roberval CS 60319, 60203 Compiègne Cedex, France

[2] Swiss Federal Institute of Technology at Lausanne (EPFL), STI-IPR-LICP, CH-1015 Lausanne, Switzerland

[3] DeltaCAD, 795 Rue des Longues Raies, 60610 La Croix-Saint-Ouen, France

[*] *Corresponding author*. Tel.: +33-770-411-384 // +33-344-234-971
E-mail Address: contact@jonathandekhtiar.eu & jonathan.dekhtiar@utc.fr

Abstract.

Mechanical Reverse Engineering has been getting increasingly more attention from the industry. It aims rebuilding a broad Digital Mock Up (DMU) in order to redesign and/or remanufacture a product. Some of the reverse engineering challenges are to perform an efficient knowledge extraction out of the original product, and then to process the data it and consolidate them for further analysis. These data could be extracted from a vast number of different data as such as Manufacturing Data, Technical Reports, Design Data (e.g. CAD Files, technical drawings, etc.), Quality procedures, etc. Moreover, the amount of data stored by the companies' information system keep on rapidly growing. We propose to use data science in order to cope with the previous issues. This paper aims to detail the different possibilities offered by the data science field of expertize, more precisely in terms of machine learning, text mining and computer vision, but also to give a brief overview on the future works we will research. This paper position itself as a roadmap for our further proceedings.

Keywords: Reverse Engineering, Knowledge Based Engineering, Data Science, Machine Learning, Deep Learning, Computer Vision

1 Introduction

Over the last decades, mechanical reverse engineering has been getting more and more interest from the industry, many industrial situations require to perform some reverse engineering process, such as remanufacturing a part which has been originally produced without any Computer-Aided Design (CAD) system. For instance, we could mention the need of remanufacturing a part of a Nuclear Power Plant which is likely to have been manufactured between 1960 and 1980 in France and thus it is unlikely that CAD software has been used to design most parts. Each situation requires a specific analysis. What are the different elements at our disposal? How reliable are they? Do we need to preprocess them? What do we aim to reverse engineer? How precisely do we need to perform these operations? Reverse Engineering is a wide field that could and should operate all along the product lifecycle. Let's have a closer look on how Reverse Engineering fit into a PLM approach.

2 PLM, a product and DMU centric approach.

Product Lifecycle Management (PLM) systems are, nowadays, widely used and a common response to cope with numerous mechanical engineering and manufacturing issues. They appear to be the main answer to many challenges induced by the ever-demanding market, the global competition, the inevitable growing customer needs and shortening products' and components' lifecycle.

Moreover, the concept of "*extended enterprise*" is becoming the norm, and thus companies are facing a rapid and massive increase in data exchanges, but also in collaborations between a vast amount of expertizes. This trend could be an obstacle if not well managed. PLM systems are thought and designed to address such problematic during the whole product lifecycle.

As this study partly focuses on data that could be found in such systems, we considered in this study this definition of PLM system: it is a "*collaborative backbone allowing people throughout extended enterprises to work together more efficiently*". It is "*a holistic business concept developed to manage a product and lifecycle including not only items, documents and BOM's, but also analysis results, test specifications, environmental component information, quality standard, engineering requirements, change orders, manufacturing procedures, product performance information, component suppliers, and so forth*" [1]. As a matter of fact, a PLM system is a *product-centric* approach used during the whole product lifecycle. Its different steps have been specified and described in the scientific literature [2], and could be schematized as follows :

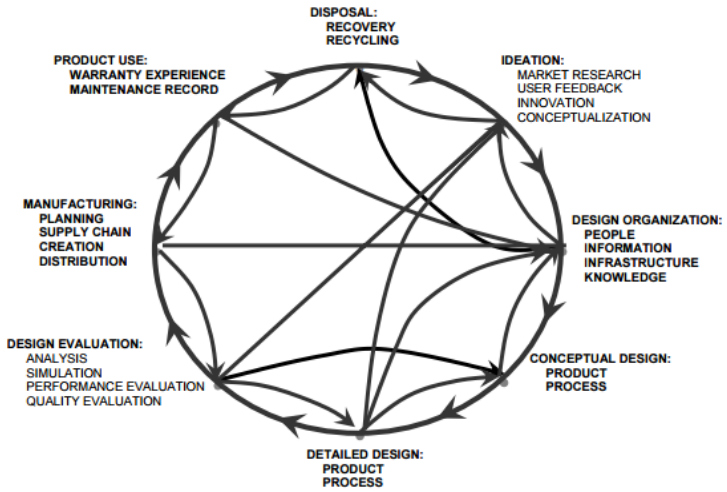

Figure 1 - PLM systems are used during the whole Product Lifecycle [2]

In our frame of research, the Digital Mock-Up (DMU) concept seems to have a prominent role. We will use this notion as a *numerical container* during the whole product lifecycle for every kind of product-related documents, files, technical reports, photos, videos, Bill of Materials (BOM), Computer-Aided Design (CAD) files, etc. We believe that expert knowledge could be extracted or deduced from all kinds of data, and therefore it seems essential to agree on an extensive definition of the DMU.

3. Related Works

3.1 Reverse Engineering, an Essential Part of the Product Lifecycle.

Our approach aims knowledge integration from products and for the products. But also to enhance and speed up the design tasks carried out in an industrial manufacturing context. Reverse Engineering (RE) is precisely the field of expertize which focuses on such goals. But also on understanding the design intents and manufacturing processes [3]. In contrast with the RE process, we find Forward Engineering (FE) which is the traditional engineering workflow process (designing, manufacturing, maintaining, etc.) [4]. FE goes from high-level ideas and concepts to a manufactured part or product, contrary to RE, which goes from an analysis of the existing situation to higher-level.

So how could RE be part of the product lifecycle described by Subrahmanian E. et al. [2] ? Here are a few examples, highlighted in the book written by Raja V. et al. [4].

- Conceptual Design and Detailed Design:
    - The original manufacturer doesn't exist any more or doesn't manufacture the part or product any more. However, the need is still here (e. g. long life parts in nuclear power plants).

- There is no CAD model. It may never have existed, or have been lost. We need to re-engineer a complete CAD 3D model.
- A company wants to speed-up its FE activities by using previous knowledge and experiences already inside the company's Information System (IS) (e. g. rapid-prototyping activities [5]).
- Evolving a product for the purpose of improving some characteristics or to modify the product's specifications (e. g. reinforcing a part which is too easily damageable).

- Manufacturing:
  - Rapidly adapt the manufacturing resources (human, materials, machines, etc.)
  - Automate and systemize the quality check all along the manufacturing chain (e. g. comparing a *real* product to a *numeric* ideal version or to a *supposed perfect* previously produced part).
  - Speeding up CNC Code generation by using previous knowledge on each specific machine. [6]
- Disposal:
  - Recycling and dismantling process has been lost or has never produced. Nonetheless such documents could become mandatory by law depending on the country.

Out of this list, we can mention that it is not necessary to reverse engineer the whole DMU in each use-case. We can and we should focus on multiple levels of granularity. In some cases, we will need to reverse engineer the 3D CAD File. In other cases, it could be just the BOM or the documentation. And finally we could also have an industrial need which requires a combination of all the previous ones. To summarize, the use case determines which part of the DMU and how precisely we will need to reverse engineer.

3.2 Data Heterogeneity in Companies' Information Systems

In order to perform a broad and efficient a reverse engineering workflow, we need to gather, clean, sort, process and make sense of a large amount of data. As these data are highly related to some specific engineering expertise, they could be labelled *engineering data*.

These data could be found inside a vast variety of systems, such as PLM systems, Product Data Management (PDM) systems, Enterprise Resource Planning (ERP) systems, Knowledge Based Engineering (KBE) systems, Databases, etc.
They are, by nature, utterly heterogeneous. We mean that the data type and format vastly vary between the data. Here are some examples to illustrate our point.

Raw Data *(no defined structure)*	**1D Data** *(tabular data)*	**Imagery Data**	
		2D Data	**3D Data**
Quality Procedures Technical Reports Documentation	Databases BOMs (*if well formatted)* Assembly Trees Basic characteristics *(colour, weight, etc.)* Manufacturing Logs	Photos Plans X-Rays Screenshots	CAD File Exchange formats (*stl/step/iges)* Laser Scans Tomography Point Clouds CMM

Table 1. We propose four categories to classify the different data types and formats with a non-exhaustive list of data-sources.

In addition to the heterogeneous aspect, companies are facing a rapidly growing amount of generated and stored data. Most manufacturing machines are now connected to the local network and operated by a computer. They produce manufacturing logs which can be stored in raw files, in databases or in systems such as ERPs (which basically run over databases). Emerging technologies such as NFC/RFID tend to produce even more data allowing advanced and complex tracking of every part during the whole product lifecycle. It is in this way that car manufacturers can gather a very large amount of data, during maintenance phase, about the product. Moreover, generalization of extended companies, extreme personalization and the recent emergence of the *Internet of Things* tend to reinforce this trend. All of these facts have led to an explosion in data amounts stored in Information Systems.

As seen previously, engineering data are, by nature, heterogeneous. Some are structured, others are raw files. And even if we only wanted to focus on the structured data, which is not our goal, the data's structure vastly varies. This great disparity implies that a specific method must be developed for each data type.

3.3 Available Solutions on the Market

Over the last decades, improvements on data acquisition techniques have permitted to elaborate complex methods to perform geometrical recognition and canonical form detection [7]. Mathematical and geometrical analysis has been previously performed in order to *clean* and fill holes in a 3D mesh obtained by laser-scan for instance [8]. Such advances could, of course, be the ground basis of RE oriented software. If we have a look to the available solutions on the market to answer the RE needs, we could find the well-known market leading solution GeoMagic Design X (formerly RapidForm XOR) developed by 3D systems. This software allows the user to generate CAD Models from 3D Scan Data (mesh or point clouds). Metris 3D developed similar software: Focus Reverse Engineering which basically offers the same features and possibilities. Unfortunately, they all produce a non-editable solid model or weakly feature-based 3D Model, and therefore perform little to no expertize knowledge integration. Even if, such approach could be compliant for the emerging 3D printing industry, it provides insufficient information to be a solid backbone for products redesign and remanufacturing.

3.4 Knowledge-Based Approaches in Reverse Engineering

Based on that observation, many research projects were launched aiming knowledge integration. In this sense, Knowledge-Based Engineering (KBE) appears to be the answer for capturing and reusing previous knowledge. These approaches gain meaning especially through a reverse engineering process. The main aspects of KBE could be summarized as follows: *knowledge*, *engineering* and *automation* [9]. One of the most-known methodologies is MOKA (Methodology and software tools Oriented to Knowledge-based engineering Applications) [10], an eight steps KBE lifecycle. We could also highlight the methodology KOMPRESSA (Knowledge-Oriented Methodology for the Planning and Rapid Engineering of Small-Scale Applications) [11] which focus on Small to Medium Enterprises. More specifically dedicated to the reverse engineering context, we could find the Knowledge Based Reverse Engineering (KBRE) [12] methodology developed for the project PHENIX. Nonetheless, none of these approaches gives an effective framework to perform knowledge extraction for a vast amount of heterogeneous data. Some initiatives try such as the METIS project [13–15] aims to solve such problematic. However, it appears that only a few data types have been covered and we might expect to find expertize knowledge in many more of them. Moreover, the proposed approach is mainly data-type specific and strongly relying on a domain expert to operate the different reverse engineering steps.

We propose to cope with this issue by using technologies and methods of the ***Data Science*** field of expertize. Let's have a closer look to the offered possibilities.

4 Prospective approach

4.1 What do we mean by data science?

Data Science is the field of expertize which focus on *making sense of the data* in a very broad view. Analysing and understanding the data are two pillars of the methods. It aims to elaborate a model able to explain the data and make accurate predictions on new data. Prediction accuracy is measured by a set of statistical tools that we won't develop here. On the other hand, predictions mainly refer to classification tasks.

It is fundamental to point out that Data Science and more specifically Machine Learning (ML) techniques are not a solution for every type of problem. However, ML techniques can greatly help to solve problems that would be inconvenient to treat in a conventional way.

- The rules are hidden, or not very well known: Many human tasks require complex decision processes, the rules may vastly vary between each of us and we might not even be conscious of this process and why we made such a choice. For instance, a spam filter is a fully personalized program that learns, from your habits, which emails are spam. We could also mention some effortless tasks for a human being such as detecting a

face. So much effortless, that sometimes our brain may mistakenly identify faces in nature (e. g. clouds, trees, etc.) or human-made objects. It's called Pareidolia. However, it is a tremendously difficult task for a computer to learn how to detect faces.

- Scaling is not humanly possible: It might be possible for a human to classify a few hundred emails a day and decide whether they are spam or ham (non-spam). This task may become cumbersome when you need to process millions of emails each day, as Google, Yahoo, Microsoft and other companies do. Nonetheless, ML techniques are pretty efficient to address such problems, moreover, many of them might become more accurate and efficient while working at large scales.

Both situations are very similar to the context of our study. And this is why we assume we could use data science and machine learning to improve state of the art results for the issues mentioned beforehand.

4.2 What data science and machine learning techniques could we use to address some of the aforesaid issues?

There are two main types of machine learning techniques. The first is *Supervised Learning* (SL), it requires an access to the domain expertize knowledge during the model construction. The second is known as *Unsupervised Learning* (UL), contrary to the previous case, it doesn't use any kind of knowledge during the model construction. It is important to notice that there exists other types such as *weakly supervised learning* [16], *semi supervised learning* [17], *reinforcement learning* [18, 19], etc. However, we won't focus on these types of algorithm for a feasibility study, we still could come back to them afterwards once the validity of our approach has been proven.

To be brief, we expect, from a ML algorithm (supervised or not), to produce a statistical model able to sort out the data in different categories. One may distinguish two major steps in using a ML algorithm, the first one is called *training*, the model is fit to explain at best the *training data*. However, we need to be careful on many pitfalls such as over/under fitting in order to preserve a proper accuracy and generalizability of our model in order to explain new data. Once the model has been estimated sufficiently correct, it's ready to be used in production and deployed. Applications and softwares could be produced over it. (e.g. an accurate spam detection model allows to create a comfortable to use anti-spam solution). There also exist models that are able to handle regression problems, in this particular case, the output is no longer a category, but an equation that describes the relationship between two or more variables. Nevertheless, this type of output is out of the scope of this paper.

With SL, the number of categories is fixed by the expert. And a name is associated with each category. The expert needs to label every training data. In other words, the expert needs to give the expected output for each training data (e. g. We have available of 5000 emails to train a spam filter model, the expert needs to label each of the email whether it's spam or ham, but we could have done a 3 categories model: Spam / Ham / Important emails).

With UL, the number of categories depends on how many categories we want to have in our model, even if we don't really know exactly what they could be. We call these categories: *clusters.* The algorithm will then look for the best split to form *X* clusters, where *X* is the desired quantity of clusters. UL is especially useful when the expertize knowledge is hard to obtain or too expensive. UL helps to highlight the underlying structure of the data in finding similarities between data and forming groups with statistically similar data.

4.3 Text Mining –Raw Data Focused Approach

Let's imagine that we have 1,000,000 reports or bills produced during ten years of manufacturing activities in an aeronautical context. These documents come from every part of the company. They are all associated with one plane and only one. However, no tags are associated with these documents and I would like to retrieve every document related to thermic aspects. On the next day, we need, for a quality audit, to extract every quality procedure and report. It might be quite a tedious task to operate by hand. However, *Text Mining* (TM) is the subfield of ML which aims to address such problems. Given a few documents of each type (e. g. quality check, thermic, materials, design, financial, manufacturing, disposal, etc.), we could imagine a system trying to solve such an issue. If it is proven to work, it would not only work in the reverse engineering context but also in the forward engineering context. It would allow to display a filtered and context-based view of the DMU. Such issues have been highlighted in the European Project LinkedDesign [20, 21].

4.4 Computer Vision – 2D and 3D Data Focused Approach

We have seen we could use TM to classify, sort and filter textual documents. However, how could we classify, sort or filter imagery documents (e. g. photos, videos, plans, 3D scans, etc.) depending on the context or the desired expected output. A realistic industrial situation could be: we have 3D scanned a car engine, and we would like to reverse engineer the BOM in order to understand its functioning.

We propose to use another domain of data science known to address such problems: *Computer Vision* (CV). Basically CV aims to acquire, process, analyse describe, and understand at best the content of an imagery data [22–25]. Some of the typical tasks of CV are, for instance, object recognition (also called object classification), pose estimation, image segmentation, object tracking in image sequences. Messy backgrounds tend to make these tasks even more difficult.

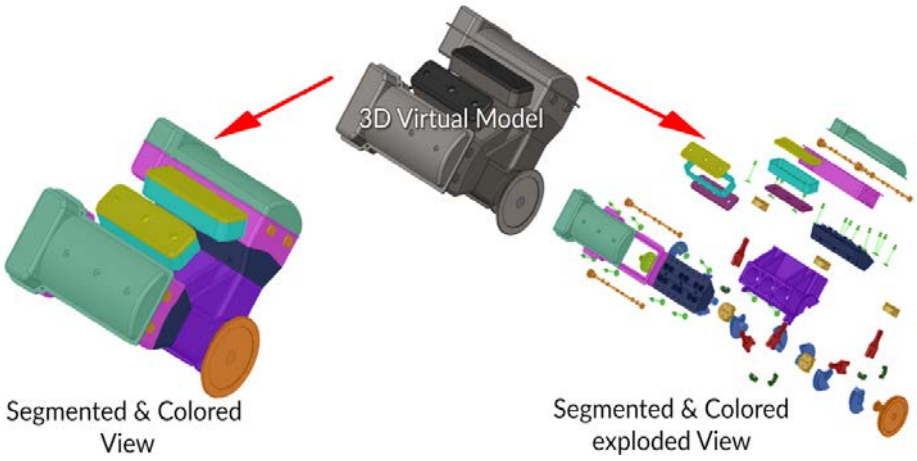

Figure 2 – Picture of the original 3D engine on the middle. On both sides the CV algorithm gives a theoretical output: a **segmented** and **classified** view of the assembly**.**

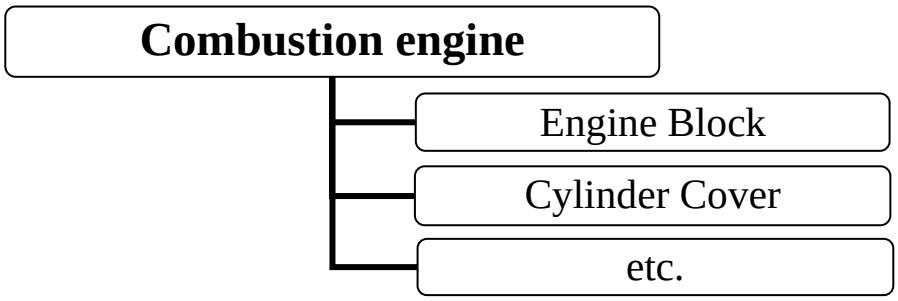

Figure 3 – The overall process could aim at producing a mechanical expertize oriented view based on the output given by the CV algorithm.

We also would like to use such technologies in order to measure and characterize the mechanical parts in imagery data (2D or 3D). Being able to describe each part in terms of physical properties (e.g. diameter, length, etc.) appears to us to be also a crucial and essential goal to achieve. We identified a few publications in the scientific literature that could help to answer such problematic [26, 27].

5 Conclusion

This contribution is a prospective approach aiming to summarize goals being pursued but also the challenges we would like to overcome. Massive and multi-source data is the tall order of tomorrow.

We believe that, beside the creation of new challenging issues for the manufacturing industry and the researchers, it could create new ways to solve and address classical engineering issues. And since knowledge and heterogeneous data integration is one of the main courses of action in reverse engineering to perform, such an approach takes on even more significance in a reverse engineering workflow.

It is for these reasons that given the limitations of typical Knowledge-Based Engineering approaches to perform a broad knowledge integration, but also bearing in mind the fast-growing amount of data stored inside the companies' information systems and the needs of reverse engineering, we are convinced of the revelance of a data-oriented methodology.

Our future works will focus to prove its effectiveness and will also try to build different metrics in order to compare the performance of the different working approaches. We hope and aspire to prove not only the revelance but also the effectiveness of a data-oriented approach to address issues related to reverse engineering and knowledge-based engineering.

References

1. Saaksvuori, A., Immonen, A.: *Product Lifecycle Management*. Springer Berlin Heidelberg, Berlin, Heidelberg (2008).
2. Subrahmanian, E., Rachuri, S., Fenves, S.J., Foufou, S., Sriram, R.D.: *Product lifecycle management support: a challenge in supporting product design and manufacturing in a networked economy*. International Journal of Product Lifecycle Management. 4–25 (2005).
3. Várady, T., Martin, R.R., Cox, J.: *Reverse engineering of geometric models - An introduction*. Computer-Aided Design. vol. 29, 255–268 (1997).
4. Raja, V., Fernandes, K.J.: *Reverse Engineering: An Industrial Perspective*. Springer Science & Business Media (2007).
5. Yan, X., Gu, P.: *A review of rapid prototyping technologies and systems*. Computer-Aided Design. vol. 28, 307–318 (1996).
6. Danjou, C., Le Duigou, J., Eynard, B.: *Closed-loop Manufacturing, a STEP-NC Process for Data Feedback: A Case Study*. Procedia CIRP. vol. 41, 852–857 (2016).
7. Várady, T., Facello, M.A., Terék, Z.: *Automatic extraction of surface structures in digital shape reconstruction*. Computer-Aided Design. vol. 39, 379–388 (2007).
8. Pernot, J.-P., Moraru, G., Véron, P.: *Filling holes in meshes using a mechanical model to simulate the curvature variation minimization*. Computers & Graphics. vol. 30, 892–902 (2006).
9. Ammar-Khodja, S., Perry, N., Bernard, A.: *Processing Knowledge to Support Knowledge-based Engineering Systems Specification*. Concurrent Engineering. vol.

16, 89–101 (2008).
10. Stokes, M.: *Managing Engineering Knowledge: MOKA: Methodology for Knowledge Based Engineering Applications*. (2001).
11. Lovett, P.., Ingram, A., Bancroft, C..: *Knowledge-based engineering for SMEs — a methodology*. Journal of Materials Processing Technology. vol. 107, 384–389 (2000).
12. Durupt, A., Remy, S., Ducellier, G., Bricogne, M.: *KBRE: a proposition of a reverse engineering process by a KBE system*. International Journal on Interactive Design and Manufacturing (IJIDeM). vol. 4, 227–237 (2010).
13. Bruneau, M., Durupt, A., Roucoules, L., Pernot, J.-P., Rowson, H.: *A methodology of reverse engineering for large assemblies products from heterogeneous data*. In: Tools and Methods of Competitive Engineering (2014).
14. Bruneau, M., Durupt, A., Roucoules, L., Pernot, J.-P., Eynard, B.: *Towards new processes to reverse engineering digital mock-ups from a set of heterogeneous data*. In: INGEGRAPH - ADM - AIP - PRIMECA (2013).
15. Ouamer-Ali, M.-I., Laroche, F., Bernard, A., Remy, S.: *Toward a Methodological Knowledge based Approach for Partial Automation of Reverse Engineering*. In: CIRP Design Conference. pp. 270–275 (2014).
16. Crandall, D.J., Huttenlocher, D.P.: *Weakly supervised learning of part-based spatial models for visual object recognition*. In: Leonardis, A., Bischof, H., and Pinz, A. (eds.) ECCV'06 Proceedings of the 9th European conference on Computer Vision - Volume Part I. pp. 16–29. Springer Berlin Heidelberg, Berlin, Heidelberg (2006).
17. Chapelle, O., Schlkopf, B., Zien, A.: *Semi-Supervised Learning*. (2010).
18. Kaelbling, L.P., Littman, M.L., Moore, A.W.: *Reinforcement Learning: A Survey*. Journal of Artificial Intelligence Research. vol. 4, 237–285 (1996).
19. Sutton, R.S., Barto, A.G.: *Reinforcement Learning: An Introduction*. (1998).
20. Nadoveza, D., Kiritsis, D.: *Ontology-based approach for context modeling in enterprise applications*. Computers in Industry. vol. 65, 1218–1231 (2014).
21. Nadoveza, D.: *Towards Ontology-based Context Modeling for Manufacturing Applications*, http://infoscience.epfl.ch/record/203793, (2014).
22. Klette, R.: *Concise Computer Vision*. Springer London, London (2014).
23. Stockman, G., Shapiro, L.G.: *Computer Vision (1st ed.)*. Prentice Hall PTR (2001).
24. Morris, D.: *Computer Vision and Image Processing*. Palgrave Macmillan (2003).
25. Jähne, B., Haußecker, H.: *Computer vision and applications: a guide for students and practitioners*. Academic Press, Inc. (2000).
26. Lee, H., Kwon, H., Bency, A.J., Nothwang, W.D.: *Fast Object Localization Using a CNN Feature Map Based Multi-Scale Search*. (2016).
27. R, S.S., Babu, R.V.: *Salient Object Detection via Objectness Measure*. In: Image Processing (ICIP). pp. 4481 – 4485 (2015).