Université Grenoble Alpes

Initiatives de Recherche à Grenoble Alpes (IRGA)

Année 2023

Acronyme du projet	Deep Twin
Titre du projet	Deep Learning-based Digital Twin of Manufacturing Systems
Porteur du projet	Pierre DAVID, Nuno GARCIA (université Lisbonne)
Dispositif	Projet international Unite!
Demande de moyens	☒ Fonctionnement/Investissement ☒ ADR - Ecole doctorale : I-MEP² - Ingénierie - Matériaux, Mécanique, Environnement, Energétique, Procédés, Production ☐ Post-doctorant ☐ BIATS/IT

Description du projet

■ Projets internationaux (5 pages maximum)

- State of the art and project originality
- Scientific or technologic challenges
- Scientific approach and work plan
- Results and deliverables
- International positioning of the project; Current status of the international collaboration and its strengthening

State of the art and Project originality

A digital twin in a manufacturing system is a virtual replica of a physical system that can be used to analyze, monitor, and optimize processes in real-time. These digital models are important in modern smart industries, helping in terms of production planning and control decisions if they are correctly representing their physical counterparts. More specifically in operations management digital twin can be used [16] for example for real-time monitoring, production control or production planning. The digital is a means to organize and merge past, current and future information on the production system to make the smartest decisions in the management of operations.

Many challenges occur along the lifecycle of a production system digital twin. Namely, the creation of the digital twin necessitates skills, data and time. Then, during the operational life of the twin, it has to be fed with timely consistent and accurate data to be valuable in possible decisions. Lugaresi and Matta [15] highlights that the ability to take online decisions is strongly based on the assumption that models are properly aligned with the real system. As a consequence, practical implementations of digital twins remain scarce due to the challenges of real time alignment. For the digital twin creation, the challenge of data production remains the same. Halaska and Sperka [17] highlighted how discovery algorithms for digital twin creation perform better with more extensive event logs when discovering complex models.

Computer vision can aid in the process of Automated Digital Twins Generation and feeding by using cameras and sensors to capture data about the physical asset's movements, dimensions, and properties.
The complex environment of a factory poses hard computer vision problems that we aim to address in this work, such as 2D object recognition, 3D volume estimation, integrating information from multiple views, enriching the vision model with data from logs, and providing a flexible way to adapt the model to new objects and actions happening in the manufacturing environment.
Object recognition and segmentation are extensively studied tasks in computer vision [1]. Deep learning architectures ranging from fully convolutional to recent transformers have been pushing the capabilities of vision recognition systems.
Fusing RGB images with other visual modalities such as 3D data [4], Infrared [3], or thermal [5], can make models more robust to mistakes [2]. Multiple views can also be considered as multiple modalities [8].
The additional modalities might not be always available at inference time, due to sensor failure or other malfunction in the computing pipeline. One way to address this problem is to train the model using all modalities possible, and let it use whatever number of modalities available to make the prediction. This paradigm is related to learning using privileged information [6, 7].
The ultimate trend of deep learning for computer vision tasks is to leverage foundation models that are pre-trained on very large vision-language datasets [9,10]. The ability of these models to integrate different types of information, including tabular data and log information, points to future architectures that exploit the strengths of both pure vision and language models.
Besides object recognition, the digital twin generation and feeding need to consider humans in the scene. As pointed by Liu et al. 2021, significant research effort needs to be made on the topic of digital twin for people in the manufacturing phase. Understanding human actions is usually cast as a video recognition task in the computer vision literature, using only RGB or other modalities such as depth or even audio [11]. The specific case of human-object interactions is especially important for our application. Several datasets have been of great importance in the development of these methods [12][13]. As in the object recognition and segmentation task, foundation models are starting to play a key role in this area [14].

This project will build on these recent advances in classic computer vision tasks and advance digital twin generation for smart factories. It will contribute to more precisely consider human behavior in the production system digital twin. It is also to contribute to developed digital twin feeding technology by focusing on a unique technology capable of replacing a sensor network and capable of providing high level information namely on operator's activity.
The originality of the work this work is to consider computer vision as the major source of data for two phases of a manufacturing system digital twin: its creation and its real time feeding.

[1] Csurka, G., Volpi, R., & Chidlovskii, B. (2022). Semantic Image Segmentation: Two Decades of Research. *Foundations and Trends® in Computer Graphics and Vision*, *14*(1-2), 1-162.
[2] Zhang, Y., D. Sidibé, O. Morel, and F. Mériaudeau (2021), 'Deep Multimodal Fusion for Semantic Image Segmentation: A Survey'. Image and Vision Computing 105(104042).
[3] Liang, Y., R. Wakaki, S. Nobuhara, and K. Nishino (2022), 'Multimodal Material Segmentation'. In: CVPR.
[4] Caltagirone, L., M. Bellone, L. Svensson, and M. Wahde (2019), 'LIDAR- Camera Fusion for Road Detection Using Fully Convolutional Neural Networks'. Robotics and Autonomous Systems (RAS) 111, 125—-131.
[5] Sun, Y., W. Zuo, and M. Liu (2019d), 'RTFNet: RGB-Thermal Fusion Net- work for Semantic Segmentation of Urban Scenes'. IEEE Robotics and Automation Letters 4(3), 2576–2583.
[6] Lopez-Paz, D., Bottou, L., Schölkopf, B., & Vapnik, V. (2015). Unifying distillation and privileged information. arXiv preprint arXiv:1511.03643.
[7] Garcia, N. C., Morerio, P., & Murino, V. (2019). Learning with privileged information via adversarial discriminative modality distillation. IEEE transactions on pattern analysis and machine intelligence, 42(10), 2581-2593.
[8] Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. (2015). Multi-view convolutional neural networks for 3d shape recognition. In Proceedings of the IEEE international conference on computer vision (pp. 945-953).
[9] Yuan, L., Chen, D., Chen, Y. L., Codella, N., Dai, X., Gao, J. & Zhang, P. (2021). Florence: A new foundation model for computer vision. arXiv preprint arXiv:2111.11432.
[10] Gao, Mingfei, et al. "ULIP: Learning a Unified Representation of Language, Images, and Point Clouds for 3D Understanding." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.
[11] Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.
[12] Heilbron, F. C., Escorcia, V., Ghanem, B., & Niebles, J. C. (2015, June). Activitynet: A large-scale video benchmark for human activity understanding. In 2015 IEEE conference on computer vision and pattern recognition (CVPR) (pp. 961-970). IEEE.
[13] Bhatnagar, Bharat Lal, et al. "Behave: Dataset and method for tracking human object interactions." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[14] Ko, D., Choi, J., Choi, H. K., On, K. W., Roh, B., & Kim, H. J. (2023). MELTR: Meta Loss Transformer for Learning to Fine-tune Video Foundation Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20105-20115).
[15] Giovanni Lugaresi, Andrea Matta. Automated Digital Twins Generation for Manufacturing Systems: a Case Study. IFAC-PapersOnLine, 2021, 54 (1), pp.749-754.
[16] Mengnan Liu, Shuiliang Fang, Huiyue Dong, Cunzhi Xu,
Review of digital twin about concepts, technologies, and industrial applications, Journal of Manufacturing Systems,Volume 58, Part B, 2021,
[17] Halaška, M., & Šperka, R. (2019). Performance of an automated process model discovery–the logistics process of a manufacturing company. Engineering Management in Production and Services, 11(2), 106-118.

Scientific or technological challenges

Industrial Engineering scientific challenges: to include operator behavior in the digital twin, to automate the generation of plant supervision digital twin including operators, to merge computer vision data and workshop data for digital twin supervision.

Computer vision scientific challenges: object recognition with occlusions, 3D scanning and volume prediction, motion analysis and action prediction, real-time requirements, to perform anomaly detection for the quality control module.

Scientific approach and work plan

The scientific approach is based on addressing two technical parts: defining the computer vision algorithms for the manufacturing system analysis and defining a methodology for the manufacturing system digital twin creation and exploitation. To guide and validate the proposals, it is decided to define a reference industrial scenario on which experimenting the algorithms. This scenario will permit to define the reference industrial situation to analyze with the camera and to illustrate the use of the digital twin to operate the system. Figure 1 illustrates the main phases envisioned for the project. It is to be realized between Grenoble and Lisboa. Grenoble will held the main experimental site with its platform for real size production system simulation. Lisboa is the excellence center for computer vision activities.

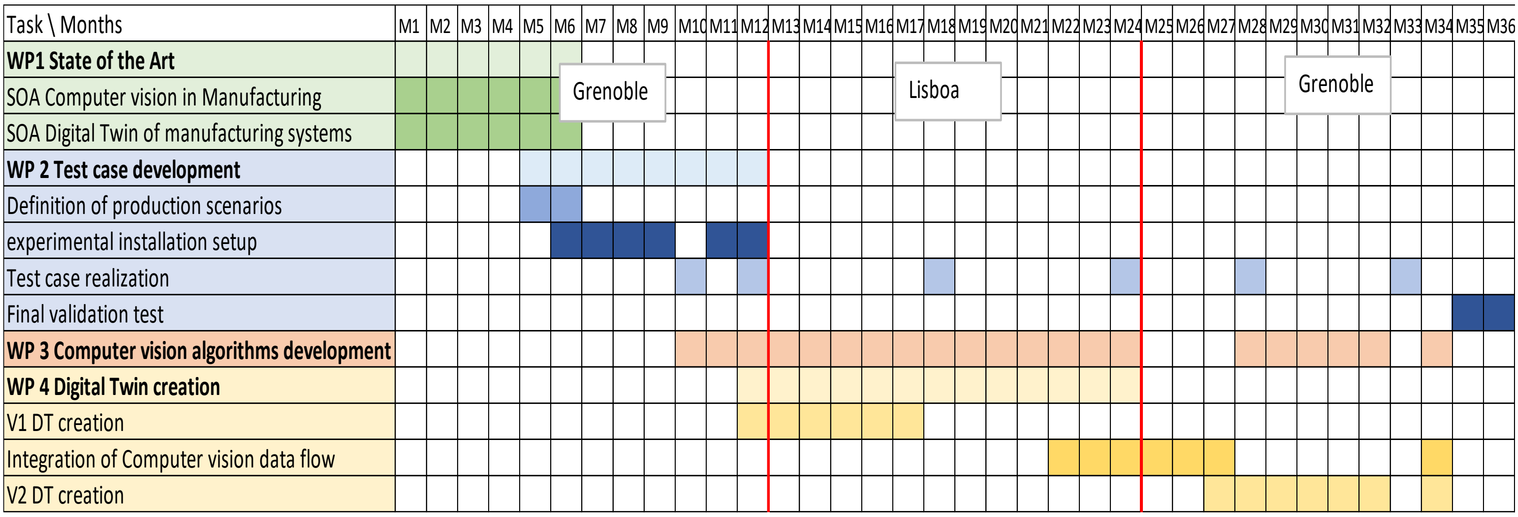

Figure 1. Project outline

WP 1. The project will start by giving a complete state of the art on the following point:

- Use of digital in production systems
- Human factor integration in digital twin
- Digital twin creation based on process mining technologies
- Usage of camera based technologies in production systems' digital twin
- Usage of camera for productive activities identification

The realized state of the art will permit to refine the research questions and identify the promising technologies to deploy in the digital creation phase.

WP 2. In parallel a test and validation case has to be developed to support and evaluate the project's technical development. It is about defining the production scenario we want to cover within the project. The scenario will be based on the existing experiment environment and test case available at G-SCOP lab. It will be a production scenario intensively embedding human participation in the production activities. For example, a hydro-turbine assembly production system is in the finalization phase within the lab. The project development will be developed and tested in real environments, both in Portugal and in France. In particular, the facility at Grenoble is fully equipped to test manufacturing systems mixing human and industry 4.0 technologies. The platform available at Grenoble Institute of technology emulates a flexible production system. It is composed of several equipment that can be integrated with each other: reconfigurable assembly stations in which human operators are handling the products: Autonomous Mobile Robots for intralogistics ; Cobots ; Real Time Location System ; Automatic Stocker.
The validation and test benches will also be defined by its realization protocol and by gathering a database of registered experiments. Scenarios demonstrating the value-added by the digital twin guided decision will have to be defined and tested.

WP 3. A work package will be devoted to the creation of the computer vision algorithm capable of retrieving the expected data from the shop floor for the digital twin creation and feeding phases. The plan is as follows: to define the highest priority data to obtain (ex: recognition of human activity, stock levels, parts paths, *etc.*)

; define the data retrieving techniques ; implementation of the retrieving algorithms (coding, training data pipeline, deploying, real-time monitoring); conduct individual validation tests.

WP 4. A work package will be devoted to the creation of the digital twin based on discrete event simulation techniques. It will also focus on the integration of the data retrieving services (camera-based factory analysis) in a digital twin feeding process. The appropriate data pipelines, dashboards and visualizations will be developed and served using cloud services. The digital twin can also be complemented using a unifying view using 3D maps and CAD models of the factory floor. Hence, a step of this project involves collecting visual data, namely 3D maps using mono or stereo vision systems, and CAD models obtained from AI models that use the acquired 3D point clouds. The work package will address the discrete event simulation model development that will be used to analyze the possible future state of the production system. An integration phase between this model and the real-time data retrieved from the camera-based analysis will be developed to provide the simulation model with real-time accurate data. The raw data from the camera analysis will be also used for the first model creation in a fashion mimic process mining technology, showing the ability to generate important parts of the digital twin.

Results and deliverables

The output of this work is an application for automated digital twin generation in real-life manufacturing scenarios, given as input a set of camera feeds and information logs. The system has to be flexible to adapt to new situations, which means a configurable continued training process of the deep learning models, reducing the time and cost of constructing a digital twin. Furthermore, it has to operate in real-time.

The setup validation scenario will also be considered as testbenches that can be made available to the scientific community.

Taken independently we can list the following technical and scientific results:

- computer vision-based algorithms for factory activities monitoring including humans' activities analysis
- a digital twin architecture to merge production logs and camera-based data
- a methodology to synchronize discrete event-based production digital twin with camera based real-time data
- a methodology to exploit camera-based factory analysis to generate production system digital twin.

International positioning of the project; Current status of the international collaboration and its strengthening

This project is the result of an application to the Unite! PhD Fellowship Call between Faculty of Sciences at University of Lisbon and Université Grenoble Alpes.
The Unite! - University Network for Innovation, Technology and Engineering - is one of the 17 European Universities pilots that were granted funding by the European Commission. This alliance is composed of nine universities and promotes collaboration towards a trans-European Campus.
This project is a step towards the realization of the Unite! goals of a more trans-regional network for open innovation and entrepreneurship.
Currently a joint master's program in Industry 4.0 is under development within Unite!, in which University of Lisbon and Grenoble INP/UGA are partners (together with TU Darmstadt, Polito and UPC). For the success of this joint master's program, the professors from all participating partners should also get to know each other and collaborate in research. This will generate further collaboration opportunities and proposals of joint research topics within this master's program. We believe that our research project will provide a valuable example for future joint collaborations among the professors who are providing input to this Master's program.

This project is an important opportunity to bring together the competencies of Lisbon team on complex computer-vision analysis and the expertise of Grenoble team in industrial operation management. It creates a synergy between two scientific specialties to advance the state of practices in Digital Twin creation and operation. The project will be the opportunity to bring new project for example within the frame of the EU funded EIT Manufacturing calls. Indeed, the produced demonstrator of the project will be a strong basis to respond to EIT manufacturing innovation calls. Both Grenoble INP and University of Lisboa belong to this EIT.