Automation, Robotics & Communications for Industry 4.0 (ARCI' 2022):

Proceedings of the 2nd Winter IFSA Conference on Automation, Robotics & Communications for Industry 4.0

2-3 February 2022

Edited by Sergey Y. Yurish

IFSA Publishing

Sergey Y. Yurish, *Editor*
Automation, Robotics & Communications for Industry 4.0
ARCI' 2022 Conference Proceedings

E-mail (for orders and customer service enquires): ifsa.books@sensorsportal.com

Visit our Home Page on http://www.sensorsportal.com

ISBN: 978-84-09-37741-1
BN-20220201-XX
BIC: TJFM

Contents

Foreword

On behalf of the ARCI' 2022 Organizing Committee, I introduce with pleasure these proceedings with contributions from the *2nd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 (ARCI' 2022)*, 2-3 February 2022.

According to the modern market study, the global Industry 4.0 market will reach US$ 165.5 Billion by 2026 growing at the CAGR of slightly above 20.6% between 2021 and 2026. The Industry 4.0 means the usage of an integrated system, which consists of an automation tool, robotic control and communications. The key factors fuelling the growth of the industry 4.0 market include rapid adoption of Artificial Intelligence (AI) and Internet of Things (IoT) in manufacturing sector, increasing demand for industrial robots, rising government investments in additive manufacturing, and growing adoption of blockchain technology in manufacturing industry.

Industry 4.0 represents the 4th industrial revolution that marks the rising of new digital industry. It is defined as an integrated system that comprises numerous technologies such as advanced robotics control, automation tools, sensors, artificial intelligence, cloud computing, digital fabrication, etc. These technologies help in developing machines that will be self-optimized and self-configured. It helps in enhancing asset performance, technology usage, material usage and other industrial processes that are involves in various industries. Numerous benefits are offered by these technologies such as low operational cost, improved productivity, enhanced customer satisfaction, improved customization, and increased efficiency. The Industry 4.0 holds a lot of potentials and is expected to register a substantial growth in the near future. There are several conferences on automation, robotics and communications, but they are not meet the Industry 4.0 challenges.

The series of annual ARCI Winter IFSA conferences have been launched to fill-in this gap and provide a forum for open discussion of state-of-the-art technologies related to control, automation, robotics and communication - three main components of Industry 4.0. It will be also to discuss how to adopt the current R&D results for Industry 4.0 and to customize products under the conditions of highly flexible (mass-) production.

The conference is organized by the International Frequency Sensor Association (IFSA) - one of the major professional, non-profit association serving for sensor industry and academy more than 20 years, in technical cooperation with media partners – journals: MDPI Processes (Switzerland), MDPI Machines (Switzerland) and Soft Measurements and Computing (Russia). The conference program provides an opportunity for researchers interested in signal processing and artificial intelligence to discuss their latest results and exchange ideas on the new trends.

I hope that these proceedings will give readers an excellent overview of important and diversity topics discussed at the conference.

We thank all authors for submitting their latest works, thus contributing to the excellent technical contents of the Conference. Especially, we would like to thank the individuals and organizations that worked together diligently to make this Conference a success, and to the members of the International Program Committee for the thorough and careful review of the papers. It is important to point out that the great majority of the efforts in organizing the technical program of the Conference came from volunteers.

Prof., Dr. Sergey Y. Yurish,
ARCI' 2022 Conference Chairman

(002)

Measure of Complexity of the Spatial Environment of a Mobile Object

A. N. Karkishchenko [1], V. Kh. Pshikhopov [2]
[1] Southern Federal University, [2] Research and Design Bureau for Robotics and Control Systems,
2 Shevchenko Str., 347928 Taganrog, Russia
Tel.: + 7 8634371694, fax: +7 8634681894
E-mail: karkishalex@gmail.com

Summary: Formal determination of the complexity indicators of the environment of a mobile object operating in three-dimensional space in the presence of obstacles is considered. The mathematical substantiation of the method for calculating the complexity is given. The concepts of local and integral complexities of the environment are introduced. Formulas for calculating the complexity are given.

Keywords: mobile object, scene, triangulation, local complexity, integral complexity.

1. Introduction

The planning of the movement of autonomous mobile objects (MO) in uncertain environments is currently of considerable interest [1, 2]. In particular, special requirements are imposed on such algorithms if the environment has elements of unpredictability and parametric uncertainty [3]. Experimental studies have shown that the choice of strategies and planning algorithms significantly depends on the characteristics of the environment in which MO operate. In many cases, it is advisable to use different planning algorithms at different stages of MO movement. It is shown in [3] that in complex environments this can lead to a significant, up to 50 %, improvement in performance indicators. At the same time, the use of the same planning methods in simple environments leads to an insignificant, about 10 %, change in quality indicators. Therefore, the key concept is the complexity of the environment, which is difficult to formalize; nowadays this concept is often defined and used intuitively. As a result, this does not allow formalizing the procedure for choosing one or another approach to planning a movement corresponding to the complexity of the environment and, therefore, increasing the efficiency of the MO.

The purpose of this work is to construct a measure of the complexity of the environment, which allows optimal selection of motion control algorithms.

The concept of complexity arises in various fields, but there is no uniformly understood definition, since specific tasks lead to the need to introduce specific definitions of this concept. "The meaning of this quantity should be very close to certain measures of difficulty concerning the object or the system in question: the difficulty of constructing an object, the difficulty of describing a system, the difficulty of reaching a goal, the difficulty of performing a task, and so on. The definition of complexity cannot be unique, simply because there are many different ways to quantify these difficulties …" [4].

In [5], an approach to determining the complexity of scenes from the point of view of the visual visibility of the surfaces forming is considered, as well as under conditions of diffuse illumination of these surfaces. In [6], a method for estimating the complexity of polygonal scenes based on reachability graphs is given. The work [7] is devoted to the introduction of the concept of the complexity of the visible part of a three-dimensional scene, depending on the point of observation. In [8], a method is given for assessing the complexity of a scene during animation, i.e., an attempt was made to develop methods for assessing the complexity of changing scenes. At the same time, in [6], a geometric approach to describing complexity is used, which is computationally laborious with a large number of elements on the scene, and in [5, 7, 8], a computationally simpler statistical approach is used to introduce the concept of complexity from the point of view of information theory. At the same time, all these methods are generated by and associated with computer graphics problems and have little relation to assessing the complexity of the environment in which an autonomous mobile object operates.

Previously, the authors considered the problem of determining the complexity measure of the environment of a mobile object operating on a plane in the presence of obstacles [9]. This work generalizes for the spatial case the results obtained earlier.

2. Problem Statement and Assumptions

An autonomous mobile object moves in space with the task of getting to the point that is the target. At the same time, there may be other stationary or moving objects in the area of movement of the MO, which make it difficult to plan the trajectory of movement, giving rise to the risk of a potential collision. Objects that interfere with the movement of a moving object will be called obstacles. To simplify the model, it is assumed that all such obstacles are replaced by the minimal balls representing them. If the real obstacle is

not such, we will assume that it is replaced by a minimal ball containing this obstacle inside.

It is assumed that the MO has range sensors that allow scanning a spatial area in the direction of movement. Geometrically, the scanned area is a pyramid S with the top that coincides with the MO.

The pyramid is always symmetrical about the axis of movement of the MO. Moreover, it has a fixed apex angle equal to $2\theta_h$ in the horizontal plane, and equal to $2\theta_v$ in the vertical plane (Fig. 1). The pyramid with the obstacles indicated in it and, possibly, the target, we will sometimes call the scene and also denote S.

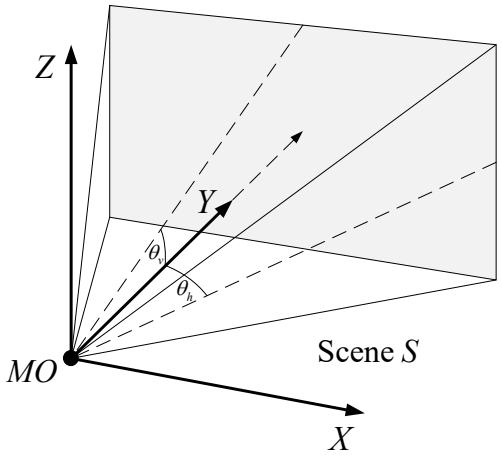


Fig. 1. Scene S and scanning angles.

We will assume that the faces forming the pyramid are always flat, while the base opposite to the top of the pyramid in reality can have an arbitrary shape, including kinks and breaks, since it is the result of scanning the real space in front of MO. Nevertheless, for the sake of simplicity, we will assume that it has the shape of a flat rectangle (Fig. 1), at a distance H from the vertex (the length of the centerline of the pyramid).

We also assume that there is an external coordinate system that allows to accurately position the object and the target. However, further constructions will be considered in the coordinate system associated with the MO, in which the origin of coordinates coincides with the MO center, the ordinate axis – with the MO axis in the direction of its movement, and the abscissa axis is directed to the right, so that together with the applicate axis, the coordinate axes form right triplet (Fig. 1).

We will assume that the space outside the scanned pyramid is completely occupied by obstacles, i.e. it is impossible to build trajectories outside the pyramid. In other words, all possible trajectories of the object's movement lie strictly inside the pyramid.

Depending on the position of the target and MO, as well as the parameters of the scene (position, number, size of obstacles), the difficulty of achieving the target may vary. This "difficulty" can vary from extremely simple, if there are no obstacles at all on the scene, up to maximum, if obstacles block the target and do not allow the MO to move to the target. In order to reasonably choose a strategy for the formation of the trajectory, it is necessary to have a quantitative measure $\delta(S)$ of the complexity of the scene S.

3. General Requirements for the Measure of Complexity

The choice of a measure of complexity is, generally speaking, not so much a mathematical question as a question reflecting the physical content of what is meant by the complexity of a real scene. It is possible to present different requirements to the complexity of this measure, meaning various parameters of the scene, characterized and/or taken into account by this measure. Therefore, the introduction of such a measure can be carried out in many ways. In particular, taking into account that dynamically changing scenes are generally considered, the degree of complexity can depend on time t, i.e., $\delta\left(S(t)\right)=\delta\left(S,t\right)$.

It should also be noted that such a measure will only relatively characterize the complexity of the scene, in the sense that the attainability of a given target will depend not only on the parameters of the scene itself, in particular, the size and number of obstacles, but also on the size of the mobile object itself. The scene, which is relatively simple for a small MO, can be difficult for a large mobile object, since the target may be unattainable for it due to the lack of sufficiently wide passages between obstacles.

At the same time, it is possible to formulate a general requirement, the fulfilment of which is naturally required from any measure of complexity. The measure should be positive and preferably normalized, i.e., $0\le\delta\left(S(t)\right)\le 1$ for any S and t. In this case, equality to zero should correspond to the simplest scene, on which there are no obstacles, and the trajectory can be a segment of a straight line. The most difficult scene, on which it is impossible to lay a trajectory to the target, should have a difficulty equal to one.

4. Triangulation and Spatial Partitioning

We will assume that there are n obstacles $B_1, B_2, ..., B_n$ on the scene S. Each obstacle in space will be characterized by four real numbers $B_i=\left(x_i, y_i, z_i, r_i\right)$, where x_i, y_i, z_i are the coordinates of the center of the obstacle, and r_i is its radius. For convenience, the center of the i-th obstacle will also be denoted $p_i(x_i, y_i, z_i)$. We will also designate by $P=\{p_1, p_2, ..., p_n\}$ – a set of obstacle centers located in space.

Consider the vertices of the pyramid S as new virtual obstacles B_A, B_B, B_C, B_D, B_E and construct the spatial Delaunay triangulation on expanded set of obstacles [10]. With this construction, all internal regions of the resulting partition will be tetrahedra, and its surface will be the convex hull of the triangulation (Fig. 1).

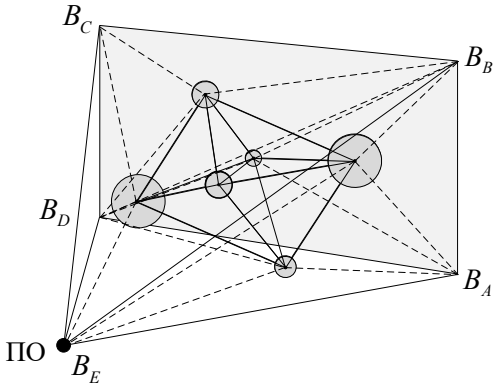


Fig. 2. Triangulation of an expanded set of obstacles.

Any trajectory of a MO to the target can be associated with a sequence of regions from the partition, which this trajectory intersects. Such a trajectory and the corresponding sequence of regions can be called *conjugate*. Conversely, any sequence of regions in which any two adjacent regions are adjacent, the first of which contains the MO, and the last is the target, generates a family of trajectories conjugate to it.

We will assume that all trajectories satisfy the constraints that are natural for the problem under consideration:

1) They cannot cross the border of each area in more than two points;

2) They cannot pass through points that belong to three or more partition areas.

5. Topological Description of Triangulation

The resulting partition of the pyramid S will be described by a topological graph $G = (X, U)$ with a set X of vertices, $|X| = N$, and a set U of edges, $|U| = M$. The vertices of the graph are in one-to-one correspondence with N areas of the pyramid partition. Two vertices are considered adjacent if and only if the corresponding areas have a common border of non-zero area. All points lying inside some open area of the partition are mapped to the same vertex of the graph.

There is a certain relationship between the number N of graph vertices, the number M of edges and the number n of obstacles, which is determined by the particular qualities of the Delaunay partition. The Delaunay partition can be viewed as a net in three-dimensional space of a four-dimensional polyhedron of a special type. Linear relations for the number of faces of different dimensions of this polyhedron are described by the Dehn-Sommerville equations [11]. If we denote by f_1 the number of edges (one-dimensional faces) of the Delaunay polyhedron, then with the help of some additional constructions it can be shown that the following relations hold:

$$\begin{cases} M - 2N + 3 = 0, \\ n - f_1 + N + 7 = 0. \end{cases} \quad (1)$$

As it will be shown below, the parameters of the graph ultimately determine the complexity of the reachability computation.

The point defining the location of the target (if any) will be inside a certain area of spatial partition, therefore it will also be identified with some vertex of the constructed graph. However, as follows from the construction, the point at which the MO is located is a virtual obstacle B_E; therefore, it is identified with all areas for which B_E is the vertex of the corresponding tetrahedron.

If the target and the MO turn out to be identified with the same vertex of the graph, then the complexity is zero. The solution in this case is a trajectory, which is a segment of a straight line connecting the MO and the target. Therefore, we will assume that the mobile object and the target correspond to different vertices of the graph.

Any trajectory of the object's motion corresponds to a conjugate sequence of partition regions, and to it, in turn, a path on the graph G from some vertex corresponding to the MO to the vertex corresponding to the target is in one-to-one correspondence. To obtain constructive results, this model requires metric refinement, since different paths can be very different in terms of the safety of their passage.

6. Path Width. Target Reachability

When moving, a mobile object intersects the imaginary tetrahedron faces in the Delaunay partition. The faces have different sizes and, therefore, different complexity of passing. This complexity can be considered as the throughput, hereinafter called "width", and measured as the area of the region, which is the locus of points, possessing the following property: if the center of the object is at such a point, then the object passes unhindered through this face (Fig. 3).

Let an arbitrary face be formed by three obstacles with centers at points p_1, p_2, p_3 and, respectively, radii r_1, r_2 and r_3. Then, using geometric reasoning, it can be shown that the width w of such an area is determined by the expression

$$w = \frac{1}{2}\left|\left[\overline{p_1p_2}, \overline{p_1p_3}\right]\right| - \frac{1}{2}\sum_{i=1}^{3}\left(r_i + \rho + \sigma\right)^2 \alpha_i,$$

where $[\bullet]$ is the vector product, ρ – radius of the mobile object, α_i are the angles in radian measure at the corresponding vertices of the face, $\sigma > 0$ is a certain "gap" of safety. The more σ, the safer the passage.

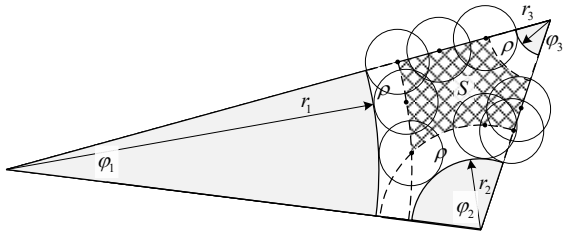


Fig. 3. The region of passage of the object through the face of the two regions of the partition.

The entered widths can be considered as a weight function defined on the edges of the graph. Thus, we consider a weighted topological graph $G = (X, U, w)$ with a marked set X_{MO} of vertices corresponding to the position of the object, and a vertex x_T corresponding to the target.

The width of a path on a graph is a value equal to the smallest width among all the edges that form this path. Consider all the paths on the graph that connect the object to the target. *Reachability* γ of the target refers to the maximum width among all such paths.

7. Matrix Procedure for Computing Reachability

To calculate the reachability $\gamma(x_T)$, one can use a simple matrix procedure. Consider the adjacency matrix $R = \|r_{ij}\|$ of a graph $G = (X, U, w)$, where

$$r_{ij} = \begin{cases} w(u_{ij}), & \text{if } x_i \text{ and } x_j \text{ are adjacent}, \\ 0, & \text{otherwise.} \end{cases}$$

Now define the maximin composition $R^{(2)} = R \circ R$ of the matrix R with itself according to the rule $R^{(2)} = R \circ R = \|r_{ij}^{(2)}\|$, where $r_{ij}^{(2)} = \max\limits_{k \in I_N} \left(\min\left(r_{ik}, r_{kj} \right) \right)$, where $I_N = \{1, 2, ..., N\}$. Then $r_{ij}^{(2)}$ is equal to the maximum width among all widths of all paths of length 2 connecting the vertices x_i and x_j. In particular, this means that there is at least one maximally "safe" path of length 2 from x_i to x_j, the width of which is $r_{ij}^{(2)}$.

Next, inductively define $R^{(3)} = R^{(2)} \circ R = \|r_{ij}^{(3)}\|$, …, $R^{(n+1)} = R^{(n)} \circ R = \|r_{ij}^{(n+1)}\|$, …

For any i and j all paths from x_i to x_j fall into disjoint classes of paths of topological length 1, 2, … However, some of these classes may be empty. Therefore, finding the width of the safest path is reduced to choosing the safest path among the safest paths in each class, i.e., $\max\limits_{k \in I_\infty} r_{ij}^{(k)}$.

Taking into account the reflexivity of the relation set by the graph G, when calculating the maximin "degrees" of a matrix, it is sufficient to restrict ourselves only to matrices $R^{(1)} = R$, $R^{(2)}, ..., R^{(N-1)}$ since $R^{(N-1)} = R^{(N)} = R^{(N+1)} = ...$. Therefore, if, for example, $x_{\text{MO}} = x_i$, and $x_{\text{T}} = x_j$ then $\gamma(x_{\text{T}}) = \max\limits_{x_i \in X_{\text{MO}}} \max\limits_{k \in I_{N-1}} r_{ij}^{(k)}$.

Note that the matrix R has size $N \times N$, and by virtue of relations (1), equality $N = f_1 - n - 7$ holds. Thus, to determine the exact size of the matrix R, it is necessary to know the number f_1 of edges of the resulting partition. However, it can be seen that the size of the matrix is proportional to f_1. Meanwhile, an asymptotic estimate is known [12], namely, any triangulation in a space of dimension d contains $O\left(n^{\lceil d/2 \rceil}\right)$ simplices, where $\lceil \bullet \rceil$ is the nearest integer greater than or equal to that element, so in our case it is $O(n^2)$. Thus, for large n, there is a quadratic dependence of the matrix size on the number of obstacles. In this case, the number of edges of a topological graph is related to the number of vertices by the relation $M = 2N - 3$, therefore, the number of nonzero elements of the matrix R is exactly determined by its size and by virtue of symmetry is equal to $4N - 6$. Therefore, at large N, the matrix R is very sparse, which can be used to optimize computations.

Reachability γ as a characteristic of a scene is inconvenient, since, on the one hand, it is expressed in absolute units of distance, i.e., depends on the chosen scale, and, on the other hand, it does not satisfy the axiomatic requirements for a measure of complexity. Next, we will consider an expression for a complexity measure associated with the reachability and devoid of the indicated disadvantages.

8. Local Complexity Measure

Note that with growth γ it is natural to assume that the complexity $\delta(\gamma)$ should decrease. Moreover, if γ is large enough, and accordingly $\delta(\gamma)$ small, then further increase will only slightly reduce the complexity. On the contrary, for small reachability, the complexity of the scene should be large, and small increases of γ should lead to a rapid decrease in complexity. Thus, we can assume that the rate of

change of complexity at any value of reachability γ is inversely proportional to the value of complexity at the same value, i.e., $\delta'(\gamma) = -\alpha\delta(\gamma)$, where α is some positive constant characterizing this dependence. Hence, we obtain that $\delta(\gamma) = Ce^{-\alpha\gamma}$, where C is the constant of integration, which can be established on the basis of the following considerations.

When $\gamma = 0$ the target is unattainable for the MO, and the complexity of such a scene is maximal. Taking into account the axiomatically imposed requirement that the complexity measure be normalized, i.e., $0 \le \delta(\gamma) \le 1$, we obtain $\delta(0) = C = 1$ and, therefore $\delta(\gamma) = e^{-\alpha\gamma}$.

The parameter α determines how quickly the complexity of the scene decreases with increasing γ. The choice of α can be made in various ways, for example, as follows. Since the limiting value $\delta(\gamma) = 0$ is formally reached only at $\gamma = \infty$, then for real problems it can be assumed that the complexity is "practically" equal to 0, if, say, $\delta(\gamma) \le \varepsilon$, where ε is a small positive number. On the other hand, it can be assumed that the complexity of the implementation of the movement of the MO is practically zero if the path width (i.e., reachability) exceeds the area of the diametrical section $\pi\rho^2$ of the mobile object by more than a factor k. In other words, we can assume that $\varepsilon = e^{-\alpha \cdot k\pi\rho^2}$, from where $\alpha = -\dfrac{\ln\varepsilon}{k\pi\rho^2}$. With this in mind, the expression for the complexity measure takes the form $\delta(\gamma) = \varepsilon^{\frac{\gamma}{k\pi\rho^2}}$. Assuming, for example $\varepsilon = 0.01$ and $k = 20$, we get $\delta(\gamma) = 10^{-\frac{\gamma}{10\pi\rho^2}}$.

The reachability value depends on the coordinates of the target, so the last expression should be understood as a function

$$\delta\left(\gamma(p)\right) = e^{-\alpha\gamma(p)},$$

Therefore, it is reasonable to call this expression a *local measure* of the scene complexity.

9. Integral Complexity Measure

The resulting expression is an estimate of the complexity of the scene with the selected target, i.e. with the given coordinates of the target. The question arises how to characterize the complexity of the scene in general, i.e., irrespective of the given target of the movement. One of the possible approaches is to introduce some characteristic that takes into account the complexity of reaching each point of the scene, which may be a potential target. It is natural to call such a measure an *integral measure* of the complexity of the scene. A similar problem arises, for example, if the coordinates of the target are not known in advance, but are reported to the mobile object in the process of moving it on the scene, or they can change at different times. An integral measure can be constructed as follows.

For simplicity, we consider the case when all obstacles are point, i.e. have a zero radius. Let us assume for definiteness that the tetrahedron in which the mobile object is located has number 1. Let us denote by $\delta_1^*(p)$ the complexity of reaching the point $p(x, y) \in S$ by the mobile object located at the top of the pyramid. Then, as an integral characteristic of the complexity of the scene, we can take the averaging over the local measures of the complexity of all points of the scene:

$$\Delta(S) = \frac{1}{m(S)}\int_S \delta_1^*(p)dp,$$

where $m(S)$ is the volume of the pyramid S.

It is easy to see that the function $\delta_1^*(p)$ is piecewise constant on S. If we denote by the S_k - k -th region of the spatial partition of the pyramid, and by $m(S_k)$ its volume, then

$$\Delta(S) = \frac{1}{m(S)}\sum_{k=1}^{N}\int_{S_k}\delta_1^*(p)dp = \sum_{k=1}^{N}\frac{m(S_k)}{m(S)}e^{-\alpha r_{1k}}.$$

Here the quantities $m(S)$ and $m(S_k)$ are easily calculated from geometric considerations.

This formula is obtained under the assumption that the obstacles are point. In this case, it admits of a simple probabilistic interpretation. If the appearance of a point specifying the target is a random variable uniformly distributed in S, then $\dfrac{m(S_k)}{m(S)}$ is the probability of the target appearing in the area S_k, therefore, the last expression is the mathematical expectation of the local complexity function.

10. Modeling

Simulation in MATLAB for the convenience of visualization was carried out for the plane case. In this case, the pyramid turns into a triangular sector, and the obstacles are represented by the minimum circles covering them. Below Fig. 4 – Fig. **7** show examples of randomly generated scenes for the same sector with obstacles of different radius, for which the local (LC) and integral (IC) complexities are calculated. Target position $x = 32, y = 64$, the size of the mobile object is 3, the positions of obstacles and their radii within 3 - 11 were formed randomly. The number of obstacles in the experiments shown is 10. The calculation time included generating the scene and obstacles, calculating all parameters, displaying the picture on

the screen, calculating the local and integral complexities and was approximately 0.1 sec.

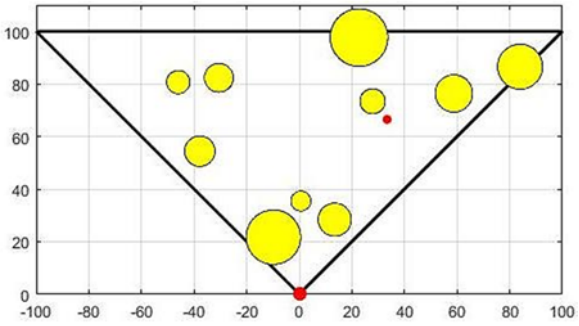


Fig. 4. LC = 1.0000, IC = 0.9791.

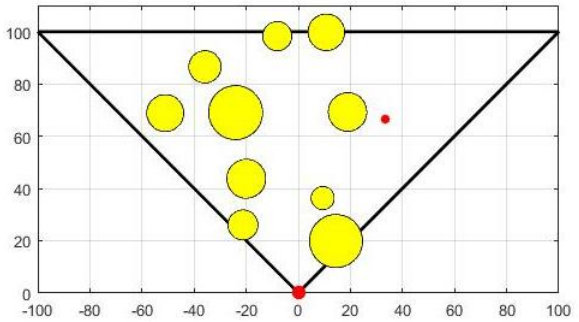


Fig. 5. LC = 0.5371, IC = 0.6282.

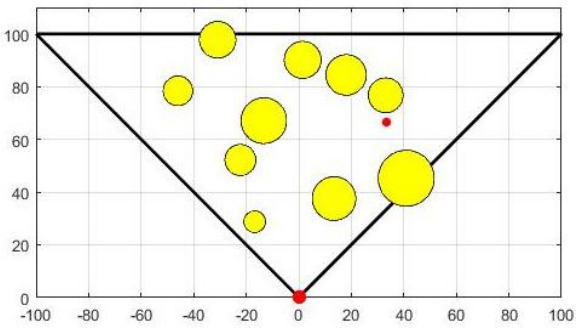


Fig. 6. LC = 0.3670, IC = 0.4503.

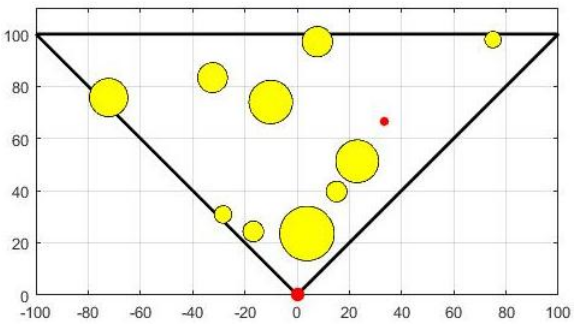


Fig. 7. LC = 0.9889, IC = 0.9805.

11. Conclusions

The paper proposes an approach to determining the complexity of the spatial environment of a mobile object. This concept is based on the safety of movement of an object when planning the path of movement to a given target. The results obtained are constructive and allow real-time calculations. The results of modeling confirmed the constructiveness and computational efficiency of the proposed measures of the complexity of the environment.

The calculation of complexity measures is done quickly enough, so the considered method can be used cyclically, applying it to dynamically changing scenes, i.e., with movable obstacles. In other words, in such conditions, based on sensory information, it is possible to continuously monitor the complexity of the environment and select an algorithm for controlling a mobile object depending on the prevailing conditions. The development of this study can be a generalization of the results for the case when the assessment of the complexity of the environment is based on several indicators that characterize not only the safe attainability of the target, but also the quality of the trajectories of movement.

Acknowledgements

This work was carried out at the Research and Design Bureau for Robotics and Control Systems with the support of the Russian Science Foundation, project no. 22-29-00533.

References

[1]. M. Hoy, A.S. Matveev, A.V. Savkin, Algorithms for collision-free navigation of mobile robots in complex cluttered environments: a survey, *Cambridge University Press*, 2014.

[2]. Xingjian Jing, Behavior dynamics-based motion planning of mobile robots in uncertain dynamic environments, *Robotics and Autonomous Systems*, Vol. 53, Issue 2, 2005, pp. 99-123.

[3]. Path Planning for Vehicles Operating in Uncertain 2D Environments, Ed. V. Pshikhopov, *Elsevier*, *Butterworth-Heinemann*, 2017.

[4]. W. Li, On the Relationship Between Complexity and Entropy for Markov Chains and Regular Languages, *Complex Systems*, Vol. 5, Issue 4, 1991, pp. 381-399.

[5]. M. Feixas, E. del Acebo, Ph. Bekaert, M. Sbert, An Information Theory Framework for the Analysis of Scene Complexity, *EUROGRAPHICS'99*, Vol. 18, Issue 3, 1999.

[6]. L. Niepel, J. Martinka, A. Ferko, P. Elias, On Scene Complexity Definition for Rendering, *Winter School of Computer Graphics and Visualization 95 (WSCG'95)*, Plzen, 1995, pp. 209–217.

[7]. D. Plemenos, M. Sbert, M. Feixas, On Viewpoint Complexity of 3D Scenes, in *Proceedings of the International Conference GraphiCon*, 2004, Moscow, Russia, (http://www.graphicon.ru)

[8]. J. Rigau, M. Feixas, M. Sbert, Visibility Complexity of a Region in Flatland, *EUROGRAPHICS,* 2000.

[9]. A. Karkishchenko, V. Pshikhopov, On Finding the Complexity of an Environment for the Operation of a Mobile Object on a Plane, *Automation and Remote Control*, Vol. 80, Issue 5, 2019, pp. 897-912.

[10]. Siu-Wing Cheng, Tamal K. Dey, Jonathan Shewchuk, Delaunay Mesh Generation, *Chapman and Hall/CRC*, 2013.

[11]. A. Brøndsted, An Introduction to Convex Polytopes, *Springer Verlag*, 1983.

[12]. R. Seidel, The upper bound theorem for polytopes: an easy proof of its asymptotic version, *Computational Geometry*, Vol. 5 Issue 2, 1995, pp. 115–116.

(004)

Identification and Discrete Inversion of Multi-Mass Systems as Part of a Disturbance Observer

C. Schöberlein [1], M. Y. Liu [1], A. Schleinitz [1], H. Schlegel [1] and M. Dix [1]
[1] Institute for Machine Tools and Production Processes, Chemnitz University of Technology
Reichenhainer Straße 70, 09126 Chemnitz, Germany
Tel.: + 4937153130505 fax: + 49371531830505
E-mail: chris.schoeberlein@mb.tu-chemnitz.de

Summary: In the context of Industry 4.0 and the proclaimed digitalization of production towards autonomous and self-monitoring systems, an increasing demand for novel monitoring and diagnostic functions can be derived. Especially in the area of production technology, continuous monitoring and evaluation of the manufacturing processes is of particular interest for developing autonomous production systems. A contribution can be made by machine internal drive systems and their integrated sensors and measurement functions. In this paper, a partial aspect of an observer structure for estimating load and process forces based on motor currents and axis positions of electromechanical drive systems is presented. The key objective is the identification of the required transfer functions of the mechanical transfer behaviour between load application point and servomotor. The methodology is based on the exclusive use of drive internal excitation and signal sources. An identification routine automatically determines the system order based on frequency responses and calculates a corresponding model. The optimization of the model parameters is carried out by applying a Nelder-Mead-optimization algorithm. Subsequently, the optimized models are inverted and compared to the original system. In addition, all model equations are discretized to enable their implementation on discrete sampled computing systems like the machine control.

Keywords: Disturbance observer, Parameter estimation, Transfer function, Discrete inversion, Multi-mass system.

1. Introduction

Process monitoring systems can contribute to increasing the productivity and flexibility of production systems as well as the quality of the manufactured products. For numerous production processes (e.g. metal cutting, deep drawing, burnishing, shear cutting), the machining forces represent a significant quantity for evaluating and monitoring the process itself. Besides an installation of additional sensors, these forces can be estimated by utilizing the already installed position and current sensors of the main and auxiliary drives of the machine.

The majority of common applied disturbance estimation methods is based on a more or less complex model of the drive systems as a disturbance observer. Ohnishi [1] presents a basic structure that subsequently serves as the origin for other observer types. The author models the mechanical subsystem as a first order system with an additional low-pass filter to eliminate resonance effects. In [2] the author considers the mechanical system as an elastically coupled multi-mass oscillator. Nonetheless, a first-order system is ultimately applied, since a directly driven linear axis is used for the experimental investigations.

In order to avoid the modeling of complex multi-mass mechanics, a position measuring system on the load side offers another opportunity. Modern machine tools usually provide such a direct measurement system for reasons of position accuracy. A so-called load disturbance observer is applied by Yamato and Sato to estimate the process forces in milling [3-5]. On the other hand, the simultaneous measurement of motor and load side position signals is also discussed in various publications (e.g. [6, 7]). The investigated system characteristic is limited to second order systems. A simulative comparison of the mentioned methods and an evaluation under variation of model order and its parameters was performed in [8]. In addition to open-loop methods, numerous publications focus on approaches in which the estimated value is continuously fed back into the model. For example, Aslan [9] parameterizes an extended Kalman filter based on frequency response measurements. The main advantage is an appropriate handling of non-minimum-phase multi-mass systems.

However, all model-based methods have in common that a model of the electromechanical system is required. The bandwidth and accuracy of the disturbance estimation increases with the level of detail of the underlying models. The purpose of this paper is to determine appropriate models for an observer approach called transfer function-based disturbance observer (TFDOB), which was already introduced in [10]. Its fundamental structure as well as an exemplary multi-mass system is shown in Fig. 1. In the figure, the parameters C and D denote the stiffness and damping values of the i-th partial oscillator with the corresponding moment of inertia J. The angular velocity ω and position φ as well as the motor torque T_m are the input values for the observer. The index m denotes all motor related parameters, whereas the index l identifies all measured variables on the load side. Note that the number of partial oscillators between motor side and load side (index i) is not

restricted. Eventually, the observer estimates the unmeasurable load torque T_l. An approach for the automatic identification and subsequent validation of the models for friction compensation ($\hat{T}_{f,m}$, $\hat{T}_{f,i}$ and $\hat{T}_{f,l}$) and axis weight correction $\hat{T}_g$ were already investigated in [11]. For the complete observer, the inverted transfer functions in the indirect case $G_{p,m}(z)$ and direct case $G_{p,l}(z)$ for estimating the acceleration torque $\hat{T}_a$ as well as the transfer function of the mechanical part $G_{mech}(z)$ must be determined.

In the following section, the procedure for the automatic identification and inversion of the required discrete models is presented. Subsequently, the experimental validation of the proposed method is carried out on a feed axis of an exemplary machining center. The paper concludes with a summary of the results and an outlook.

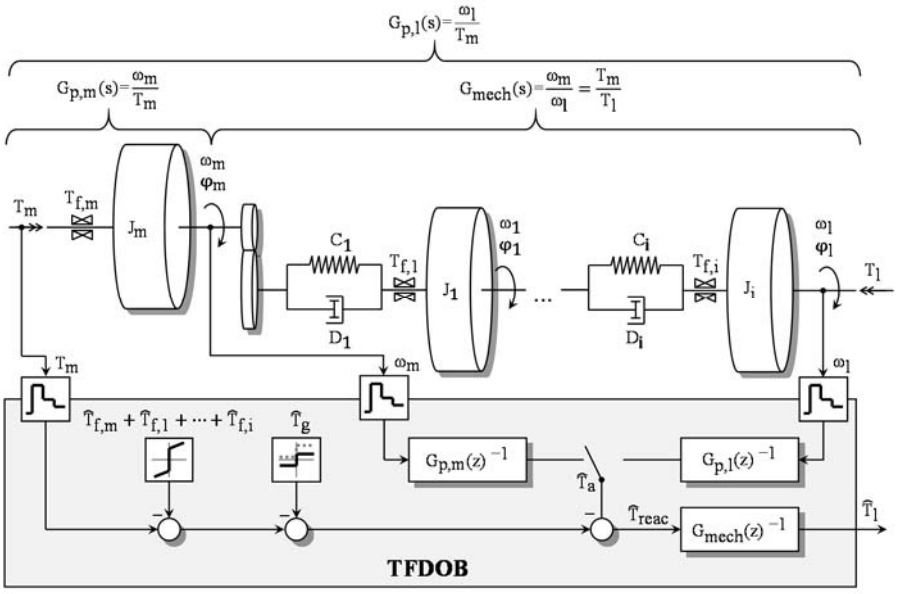


Fig. 1. Signal flow chart of multi-mass system and transfer function-based disturbance observer (TFDOB) for drive-based disturbance estimation.

2. Methodology

The fundamental sequence for determining the discrete inverse model equations is shown in Fig. 2. In the first module, the system is excited on the motor side with a Pseudo-Random Binary Signal (PRBS). The main advantage over other established identification routines, e.g. in [9, 12] is the renunciation of external excitation sources like impulse hammer or shaker. At the same time, the angular velocities on motor side and load side are recorded. Subsequently, the signals are transformed into frequency domain and their magnitude and phase parts are calculated. To avoid distortions of the frequency response in the range of the controller bandwidth, the speed controller should be parameterized as a backup controller during signal recording.

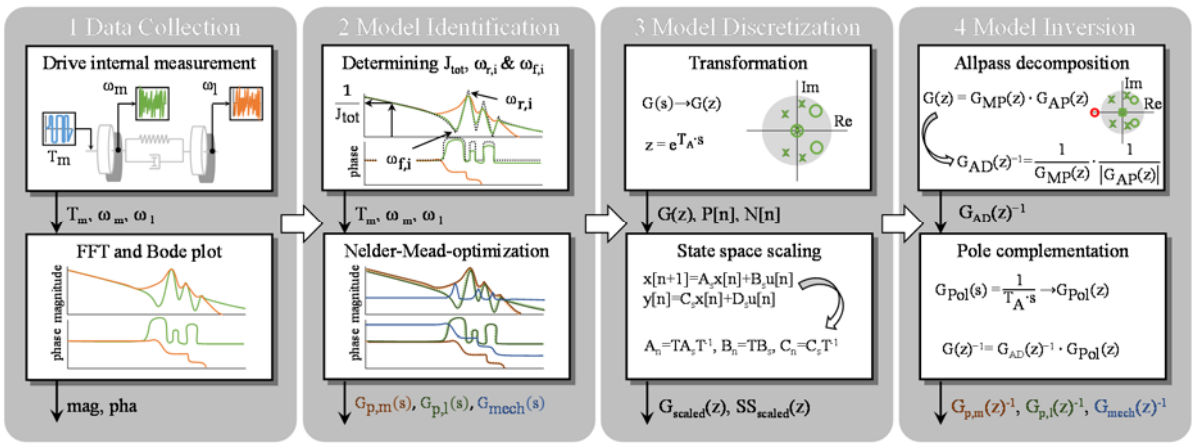


Fig. 2. Flow chart of the discrete model estimation and inversion procedure.

Based on the magnitude response, the model parameters are estimated in module two. The frequency responses for the indirect case $G_{p,m}(z)$ and direct case $G_{p,l}(z)$ are characterized by the total mass moment of inertia in the range of low frequencies according to Eq. (1). In addition, depending on the number of resonance and antiresonance frequencies $\omega_{r,i}$ and $\omega_{f,i}$, a defined number of partial oscillators can be detected. These are calculated using Eq. (2). For the direct case, an adjustment of the partial oscillator characteristic can be made by adapting parameter a {0,1} (cf. [10, 15]).

$$G_J(s)=\frac{1}{J_{tot}\cdot s} \tag{1}$$

$$G_{PO}(s)=\frac{a\cdot\left(\frac{1}{\omega_f}\right)^2\cdot s^2+\frac{2d_f}{\omega_f}\cdot s+1}{\left(\frac{1}{\omega_r}\right)^2\cdot s^2+\frac{2d_r}{\omega_r}\cdot s+1} \tag{2}$$

By multiplying $G_J(s)$ with the product of all partial oscillators $G_{PO,i}(s)$ in Eq. (3), the actual model equations for $G_{p,m}(s)$ and $G_{p,l}(s)$ are derived.

$$G_{p,ml}(s)=G_J(s)\cdot\prod_{i=1}^{n}G_{PO,i}(s) \tag{3}$$

The model of the mechanical transmission behavior $G_{mech}(s)$, which can also be calculated as the quotient of $G_{p,m}(s)$ and $G_{p,l}(s)$, is calculated solely by multiplying all i partial oscillators (Eq. (4)).

$$G_{mech}(s)=\prod_{i=1}^{n}G_{PO,i}(s) \tag{4}$$

As illustrated in the next section, a separate estimation for all three model equations is necessary to achieve correct model parameters. This can be justified by minor differences in frequency and damping values depending on the location of the direct measurement system and the constructive design of the system.

The underlying methodology for the actual determination of the model parameters was already presented in [13, 14] and partially extended. The basic idea is to generate a non-parametric model by connecting several partial oscillators in series and multiply them with the proportion for the total mass moment of inertia. Therefore, in the first part of module two, J_{tot} is determined by calculating the gradient at lower frequencies. At the same time, the resonance and antiresonance frequencies $\omega_{r,i}$ and $\omega_{f,i}$ are determined based the phase characteristic. It is necessary to distinguish between two cases affecting the value of parameter a of Eq. (2). If a previously defined phase threshold value (e.g. 50 °) is exceeded in the negative or positive direction and there is a subsequent reversal of the phase to the original value, the frequencies are close to the crossed threshold values. Therewith, the value of parameter a is set equal to one (cf. Eq. (3)). If, on the other hand, a phase rotation of -180 ° takes place without subsequent reversal, this indicates a partial oscillator according to Eq. (3) with the parameter a set to zero. The associated damping values $d_{r,i}$ and $d_{f,i}$ are initialized with default values of 0.01.

In the second part of module two, the optimization of the model parameters is performed using Nelder-Mead method [15]. The previously determined parameters J_{tot}, $\omega_{r,i}$, $\omega_{f,i}$, $d_{r,i}$ and $d_{f,i}$ are substituted into the corresponding Eq. (3) or (4) and the deviation between measured and modeled magnitude response is minimized. In addition, the method is allowed to vary the frequency values in the range of ± 10 %. The damping values may vary between 0 and 1. Eventually, this results in a continuous-time model for the indirect case and direct case (Eq. (3)) as well as the mechanical part according to Eq. (4).

In module three, the transformation of all continuous-time models into discrete-time models is performed using the z-transformation [16]. The cycle time T_{cyc} of the installed control system is used for the sampling time T_A. Furthermore, there may occur large differences between the largest and smallest exponents when calculating discrete transfer functions. This can ultimately lead to inaccuracies when determining the corresponding poles and zeros. In order to avoid this effect, the transfer functions are initially transformed into the discrete state space. Subsequently, a transformation of the system matrices into an equivalent system is performed using a transformation matrix. The underlying procedure is described in detail in [17].

In module four, the inversion of the model equations is carried out. Especially in case of direct and mechanical transfer functions (Eq. (2) and (3)), the resulting models may have a non-minimum phase character. This means that the transfer function has one or more unstable zeros, which is expressed in the pole-zero diagram by their location outside the unit circle. By inverting the model equation, all zeros become poles and vice versa. Consequently, the inverted transfer function would have the same number of unstable poles, which basically result in an unstable system itself. In this case, the transfer function is decomposed into a minimum phase component and an all-pass component. The basic procedure is explained in detail in [18]. For the subsequent inversion, only the product of the minimum phase component and the magnitude of the all-pass component is used. Although this results in a phase shift of the inverted transfer function (cf. Fig. 5), however, deviations in the phase are usually tolerable for numerous applications. On the other hand, the course of the magnitude response is not affected.

Furthermore, another obstacle arises when inverting the model equations. Due to the fact that the order of the denominator n is always larger than the order of the numerator m for all model equations, a system with differentiating character is obtained by

performing a direct inversion. However, such a system cannot be implemented in reality for reasons of causality [19]. A possible solution was already presented in [10] and applied for a second-order system. The idea is to extend the inverted transfer function with a defined number (n-m) of high-frequency poles until the denominator order corresponds at least to the numerator order. For a continuous system, this is calculated as follows:

$$G_p(s)^{-1} = \frac{1}{G_p(s)} \cdot \frac{1}{(1+T_{AT} \cdot s)^{n-m}} \qquad (5)$$

Note that the multiplication with high-frequency poles also leads to a lowering of the phase response of the inverted transfer functions. With respect to the magnitude, however, an adequate inversion of the model equation is obtained.

3. Experimental Results

The presented approach is validated on a real machine axis of a three-axis machining center DMC 850V of the company DMG Mori. The tests were carried out on the horizontally arranged x-axis of the machine. Its structural design is illustrated in Fig. 3. The rotational motion of the servomotor is converted into a linear movement of the carriage via a belt gear and a ball screw drive. Position sensors are located on the servo motor and on the carriage, whose time derivation serves as the input signals for the identification procedure. By taking into account the spindle pitch h_{sp} and gear ratio i_g the load speed is converted onto the motor shaft. The connection of the PRBS excitation signal to the motor torque T_m is implemented by a drive-internal function generator.

The results at the output of module two (model identification) are illustrated in Fig. 4. All identified and optimized parameters are listed in Table 1. For all three modell equations, a total number of three partial oscillators was detected. Note that all values for resonance and antiresonance frequency were converted from ($rad \cdot s^{-1}$) to (Hz). Regarding the frequency responses, the identified continuous models (orange) match the measured signals (blue) over a broad range of frequencies. Only the cycle time of the controller (T_{cyc} = 8 ms) limits the bandwidth of the recorded signals up to 125 Hz.

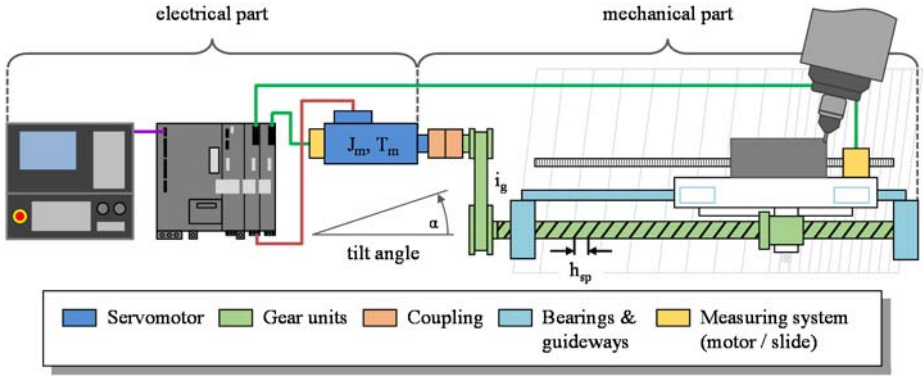


Fig. 3. Structure of the linear feed axis.

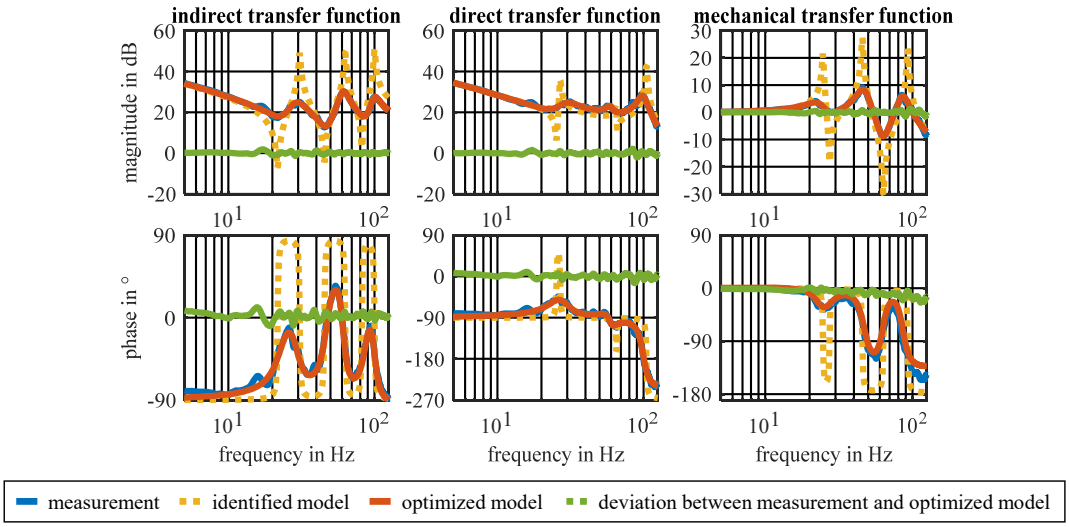


Fig. 4. Frequency response of measured signals and identified and optimized models.

Table 1. Optimized parameters of the identified model equations

System function	Total moment of inertia		Partial oscillator No. 1		Partial oscillator No. 2		Partial oscillator No. 3	
	parameter	value	parameter	value	parameter	value	parameter	value
$G_{p,m}(s)$	J_{tot} in $\frac{kgm^2}{s^2}$	0.0058	ω_f in Hz	23.51	ω_f in Hz	47.04	ω_f in Hz	87.60
			d_f	0.18	d_r	0.11	d_r	0.10
			ω_f in Hz	28.94	ω_f in Hz	61.13	ω_f in Hz	99.82
			d_r	0.14	d_r	0.07	d_r	0.07
$G_{p,l}(s)$	J_{tot} in $\frac{kgm^2}{s^2}$	0.0057	ω_f in Hz	25.18	ω_f in Hz	59.65	ω_f in Hz	99.33
			d_r	0.23	d_r	0.09	d_r	0.03
			ω_f in Hz	28.82	ω_f in Hz	58.95	ω_f in Hz	99.33
			d_r	0.19	d_r	0.06	d_r	0.07
$G_{mech}(s)$	J_{tot} in $\frac{kgm^2}{s^2}$	-	ω_f in Hz	26.18	ω_f in Hz	63.04	ω_f in Hz	85.83
			d_r	0.17	d_r	0.09	d_r	0.26
			ω_f in Hz	23.69	ω_f in Hz	47.09	ω_f in Hz	85.83
			d_r	0.16	d_r	0.10	d_r	0.09

In the next step, all identified and optimized models are transferred to their discrete-time representation and the subsequent inversion takes place. The results can be seen in Fig. 5. The frequency responses of all three discrete models (blue), their discrete inversion (orange) as well as the compensated value (green) are plotted individually. Initially, a phase drop for all models due to the discretization procedure is recognizable. By reducing the sampling time T_A, the bandwidth of the models may be further enhanced. For the direct case in the center part of the figure, the algorithm has detected one unstable zero and consequently performs an all-pass decomposition with one all-pass. Due to the phase-minimum character, this procedure is not required for the indirect case as well as the mechanical part. The magnitude response shows that the all-pass component (purple) has basically no influence on the minimum-phase component (yellow). Only in the phase signal an additional drop is recognizable. Nevertheless, a phase reduction can be observed for all transfer functions. This corresponds to other methods, for example the closed-loop approach in [9]. Apart from that, it can be stated that especially in the magnitude response, an adequate discrete inversion is performed for all three model equations.

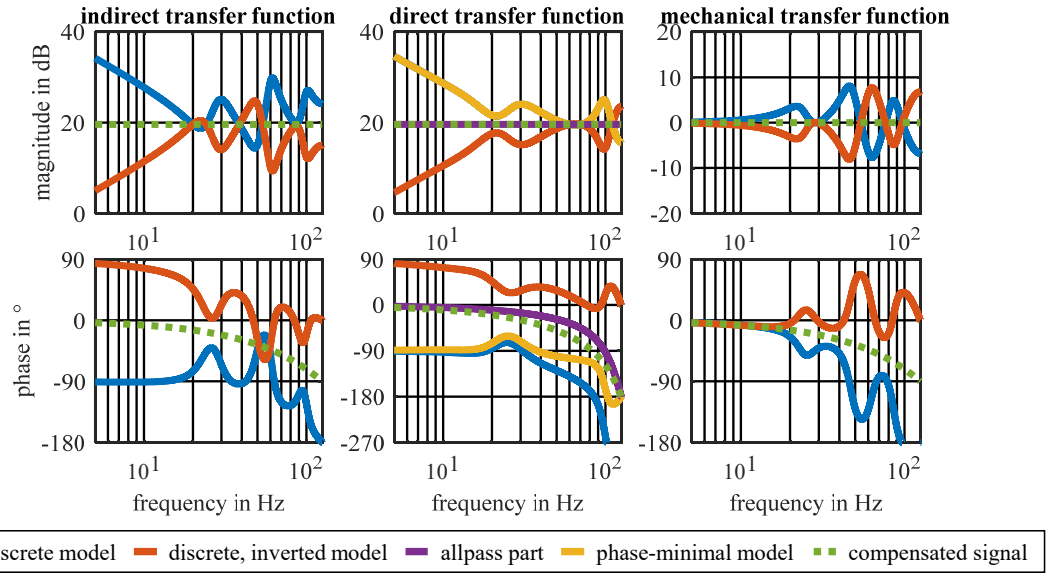


Fig. 5. Frequency response for discrete modeled and inverted transfer functions.

4. Summary and Conclusion

The paper presents a methodology for discrete modeling and inversion of transfer functions for multi-mass systems. The main advantage of the approach is the automatic determination of the system order as well as all necessary parameters. The modeled discrete transfer functions serve as origin for a new observer structure that is not limited to lower-order systems (e.g. two-mass systems). Future studies should aim to validate the complete observer structure under machining conditions. In addition, it should be analyzed to what extent the axis position in the working frame influences the values of individual frequencies in the frequency response.

Acknowledgements

Funded by the Federal German Ministry for Economic Affairs and Energy

References

[1]. Ohnishi K., Matsui N., Hori Y., Estimation, Identification, and Sensorless Control in Motion Control Systems, in *Proceedings of the IEEE*, Vol. 82, Issue 8, 1994, pp. 1253–1265.

[2]. Kempf C. J., Kobayashi S., Disturbance observer and feedforward design for a high-speed direct-drive positioning table, *IEEE Transactions on Control Systems Technology*, Vol. 7, Issue 5, 1999, pp. 513–526.

[3]. Yamato S., Sugiyama A., Suzuki N., Irino N., Imabeppu Y., Kakinuma Y., Enhancement of Cutting Force Observer by Identification of Position and Force-amplitude dependent Model Parameters, in*ternational Journal of Advanced Manufacturing Technologies*, Vol. 104, 2019, pp. 3589–3605.

[4]. Yamato S., Imabeppu Y., Irino N., Suzuki N., Kakinuma Y., Enhancement of Sensor-less Cutting Force Estimation by Tuning Observer Parameters from Cutting Test, in *Procedia Manufacturing*, Vol. 41, 2019, pp. 272–279.

[5]. Sato T., Yamato S., Imabeppu Y., Irino N., Kakinuma, Y., Precise Cutting Force Estimation by Hybrid Estimation of DC/AC Components, in *Proceedings of the 2020 International Symposium on Flexible Automation*, July 8–9, 2020, V001T02A005.

[6]. Yamada Y., Kakinuma Y., Sensorless Cutting Force Estimation for Full-closed Controlled Ball-screw-Driven Stage, *International Journal of Advanced Manufacturing Technology*, Vol. 87, 2016, pp. 3337-3348.

[7]. Yamada Y., Kakinuma Y., Mode decoupled Cutting Force Monitoring by applying Multi Encoder based Disturbance Observer, in *Procedia CIRP*, Vol. 57, 2016, pp. 643–648.

[8]. Schöberlein C., Norberger M., F., Schlegel H., Putz M., Simulation and Disturbance Estimation of Speed-controlled Mechatronic Drive Systems, in *MATEC Web of Conferences*, Vol. 306, 2020, 04001.

[9]. Aslan D., Altintas Y., Prediction of Cutting Forces in Five-axis Milling using Feed Drive Current Measurements, *IEEE/ASME Transactions on Mechatronics*,Vol. 23, Issue 2, 2018, pp. 833–844.

[10]. Schoeberlein C., Schleinitz A., Schlegel H., Putz M., Simulative Investigation of Transfer Function-based Disturbance Observer for Disturbance Estimation on Electromechanical Axes, in *Proceedings of the 17th Int. Conf. on Informatics in Control, Automation and Robotics*, 2020, pp. 651–658.

[11]. Schoeberlein C., Sewohl A., Schlegel H., Dix M., Modeling and Identification of Friction and Weight Forces on Linear Feed Axes as Part of a Disturbance Observer, *International Journal of Mechanical Engineering and Robotics Research*, 2021, in press.

[12]. Brecher C., Rudolph T., In process identification of cutting condition using digital drive signals, in *Proceedings of the CIRP 2nd International Conference Process Machine Interactions*, 2010.

[13]. Münster R., Walther M., Schlegel H., Drossel W., Experimental and Simulation-based Investigation of a Velocity Controller Extension on a Ball Screw System, in *Proceedings of the 14th Mechatronics Forum International Conference MECHATRONICS 2014*, 2014, pp. 226–234.

[14]. Hipp K., Schöberlein C., Schlegel H., Neugebauer R., Simulation based Optimization for Controller Parametrization of Machine Tool Axes – advanced Application, *Journal of Machine Engineering*, Vol. 17, Issue 1, 2017, pp. 57–68.

[15]. Nelder J., Mead R., A Simplex Method for Function Minimization, *The Computer Journal Advance*, Vol. 7, 1965, pp. 308–313.

[16]. Graf U., Applied Laplace Transforms and Z-Transforms for Scientists and Engineers: A Computational Approach Using a Mathematica Package, Switzerland, *Birkhäuser Basel*, 2012.

[17]. Park S., Altintas Y., Dynamic Compensation of Spindle Integrated Force Sensors with Kalman Filter, *ASME Journal of Dynamic Systems Measurement and Control*, Vol. 126, 2004, pp. 443–452.

[18]. Schilling R. J., Harris S. L., Digital Signal Processing Using MATLAB, *USA, Cengage Learning*, 2016.

[19]. Schröder D., Elektrische Antriebe – Regelung von Antriebssystemen, *Springer Vieweg*. Heidelberg, 4th edition, 2015.

(006)

Development and Validation of a Model for Online Estimation of Process Parameters for Adaptive Force Control Algorithms

M. Norberger [1], A. Sewohl [1], S. Sigg [2], H. Schlegel [1] and M. Dix [1]
[1] Institute for Machine Tools and Production Processes, Chemnitz University of Technology, 09126 Chemnitz, Germany
[2] Fraunhofer Institute for Production Systems and Design Technology, 10587 Berlin, Germany
Tel.: + 4937153136970, fax: + 49371531836970
E-mail: manuel.norberger@mb.tu-chmnitz.de

Summary: Production technology is characterized by the use of electromechanical feed axes, for which the concept of cascade control has become established. The concept is based on linear control engineering. It is not suitable for the control of process forces, which is associated with nonlinearities. Here, adaptive control algorithms from the field of higher control engineering represent a promising approach. The basis for adaptation is the estimation of process parameters. In this paper, the development of a model for online parameter estimation is presented. Key components are data pre-processing, parameter estimation based on a recursive least squares (RLS)-algorithm, and data post-processing. The functionality of the parameter estimation is demonstrated in simulation and validated by means of experiments on a test setup with modern industrial motion control. In addition, various influencing factors are examined and their effects are evaluated.

Keywords: Force control, Adaptive control, Parameter estimation, Motion control.

1. Introduction

In the area of production engineering, there are ongoing efforts to improve manufacturing strategies and processes in terms of stability, quality and efficiency. One possibility for ensuring stable process conditions and reducing rejected parts is closed loop control of quality determining parameters [1]. However, their control is not trivial and requires precise knowledge as well as correspondingly real-time-capable sensor technology. At this point, the process force is a suitable and very relevant variable. It is often the limiting factor for the design of the processes and the choice of parameters. As a controlled variable, the process force is predestined to ensure stability and safety of many processes [1, 2]. It contains important information regarding the process state and allows conclusions to be drawn about deviations in the manufacturing process, the machine, the tool, the workpiece or the material. However, there are many challenges and requirements associated with force control. The process itself is part of the controlled system, so that deviations of the controlled system and nonlinearities occur more frequently. With classic PID controllers, this results in poor performance or even instability. PID-control forms the basis of the established cascade control for electromechanical feed axes, which is also known as servo control [3]. Control of process forces in production machines with electromechanical feed axes is still a developing field and offers space for potential improvement. The use of higher control concepts is recommended for controlling non-linear systems. In particular, adaptive algorithms that can react to deviations of the controlled system represent a promising approach to meet the challenges. The most accurate possible a priori information of the system behavior are used for the design and parameterization of the control of technical systems [4], [5]. For this purpose, e.g. construction data but also non-invasive identification methods can be used [6]. In practice, some information may not be available a priori. Basically, this can be caused on the one hand by insufficient knowledge of process parameters in the case of a time-invariant process behaviour and on the other hand by a time-variant process behavior [4]. This can be remedied by the estimation of variable process parameters, which forms the basis for adaptation. The performance of adaptive control depends significantly on the accuracy and real-time capability of the estimation model.

This publication focuses on the development of a model for online parameter estimation. Key components are data pre-processing, parameter estimation based on a RLS-algorithm, and data post-processing. The estimation is performed in real time based on measured actual values and data from the control system. The functionality of the parameter estimation is demonstrated in simulation and validated by means of experiments on a test setup with modern industrial motion control.

In the next chapter, the basic concept of parameter estimation is described. Different possibilities are discussed and selected on the basis of the use case. The chosen experimental test-setup for the validation is presented in the third chapter. Subsequently, the structure of the parameter estimation unit as well as the conducted experiments are explained and evaluated in chapter four. Finally, the results and conclusions are summarized in the last chapter.

2. Parameter Estimation

To achieve a high control performance, the online estimation of process parameters must be convergent. Otherwise, poor control behavior up to instability is possible [7]. Online parameter estimation, both for feed axes and in general, is a very wide field. For general information and an overview of different algorithms see for example [4, 7, 8] and for online parameter estimation on feed axes see for example [6], [9, 10]. Basically, there are recursive and non-recursive parameter estimation methods. With regard to the memory requirements and the computing time as well as from the point of view of the cyclic availability of new measurement data in the process image of a programmable logic control (PLC), recursive methods should be preferred [6, 8]. The basic method for recursive parameter estimation is RLS [6], which is a recursive computational prescription of the well-known least squares method. If the process parameters are time-invariant, a distinction can be made between the cases of an immediate parameter change and a continuous parameter change [8]. In the case of an immediate change, the variables of the algorithm can be reset. In the case of a continuous change, it is useful to implement a forgetting factor λ [11].

The task of the parameter estimation model for an adaptive force control is to estimate the effective stiffness K_E online as quickly as possible and with as minimal noise as possible. K_E is defined as the change of the process force per change of the actual position. Due to the significant noise of the force measurement and the fact that both ΔF and Δx can be zero, it is not useful to calculate K_E as a numerical derivative. Analogous to [13], the use of an RLS algorithm proves to be more appropriate. With the approach:

$$\Delta F[t] = \widehat{K}_E[t] * \Delta x[t] + \varepsilon[t] \tag{1}$$

the calculation rule for the recursive least squares method results in:

$$\varepsilon[t] = \Delta F[t] - \Delta x[t] * \widehat{K}_E[t-1] \tag{2}$$

$$K[t] = \frac{P(t-1) * \Delta x[t]}{\lambda[t-1] + \Delta x^2[t] * P[t-1]} \tag{3}$$

$$\lambda[t] = 1 - \gamma * \varepsilon^2[t] - (1 - K[t] * \Delta x) \tag{4}$$

$$P[t] = \frac{(1 - K[t] * \Delta x[t]) * P[t-1]}{\lambda[t]} \tag{5}$$

$$\widehat{K}_E[t] = \widehat{K}_E[t-1] + K[t] * \varepsilon[t] \tag{6}$$

where ε is the prediction error, K is the gain, P is the covariance and $\widehat{K}_E$ is the estimation of the effective stiffness. The parameter estimation is further affected by the initialisation through the start value $\widehat{K}_{E,init} = \widehat{K}_E[0]$ and the start covariance $P_{init} = P[0]$. However, the influence of these two initialisation values gets lost quickly, so that they are not considered in more detail in the further analysis. $\widehat{K}_{E,init}$ is set to $300\,\frac{N}{mm}$ and P_{init} to 1000. A high initialisation value of the covariance leads to a fast estimation of the stiffness when the parameter estimation unit is switched on. The forgetting factor λ achieves a successive reduction of the weighting of old measurement data. The memory of the RLS estimator can be approximated as a function of λ to:

$$N = \frac{2}{1 - \lambda} \tag{7}$$

where N is the number of measured values in the estimator [8]. For $\lambda = 0.96$ the estimation is therefore based approximately on the last 50 measured values, for $\lambda = 0.5$ only the last 4 measured values are taken into account. Small values for λ thus lead to a fast adaptation of the parameter estimation to changing parameters, but bear the risk of a fast estimator blow-up and can thus lead to a strongly noisy parameter estimation [8]. The RLS-algorithm can be influenced by means of the forgetting factor. Typical values for λ are in the range of $0.95 \ldots 0.99$ [4], [8]. When using a variable forgetting factor, the control is done via the adjustment factor γ.

3. Test Setup

For the experiments and validation, a test-setup of an electromechanical feed axis was selected, which is designed for loads up to $10kN$. The basic structure of the test-setup corresponds to a portal construction. However, only one drive is used to generate the movement. Position control is implemented in cascaded structure. It is also possible to switch to a force control on the same level. The mechanical construction and control engineering structure are described in more detail in [12]. When performing load tests it has been shown that the force controller can become unstable. These instabilities result from deviations and non-linearities in stiffness, which is essential for the design of the force control. The parameter estimation method described in this paper is able to detect unexpected changes in stiffness. A modular and exchangeable spring package was designed for the experiments as shown in Fig. 1. This allows systematic replication and investigation of changes and deviations in the system stiffness.

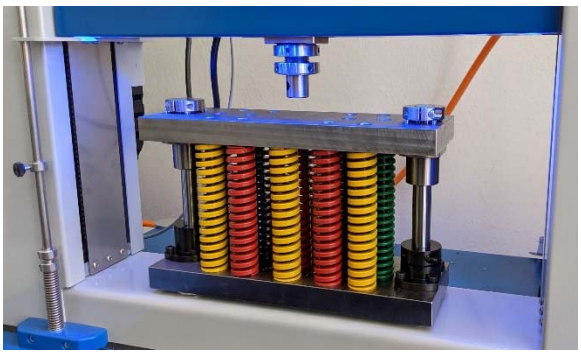


Fig. 1. Modular spring package for investigating changes in stiffness in experiments with force control.

Variable load characteristics can be initiated with high reproducibility by a movement of the axis against the spring package. The use of spacer elements can provoke high steps in the stiffness. The stiffnesses of the springs are known, which gives a reference value for parameter estimation. A manufacturing tolerance of 10 % per spring must be taken into account.

4. Implementation and Validation

4.1. Structure of the Parameter Estimation Model

As shown in Fig. 2 the parameter estimation model consists of three components. The data pre-processing, the actual parameter estimation with the help of the RLS-algorithm and data post-processing. Core component is the parameter estimation, which is explained in chapter 2. The basic structure of this algorithm is described in more detail in [4], [7] and [8]. Once the algorithm has been chosen, it can be adapted by using the forgetting factor, depending on the specific implementation. The design of the data pre-processing and the choice of suitable design parameters is considerably more extensive than the RLS-algorithm. The purpose of data pre-processing is to improve the signal-to-noise ratio and to provide the RLS-algorithm with signals with a high excitation. Therefore, suitable data sets of Δx and ΔF for the stiffness estimation must be generated. The discrete data of the measured force $F_{AV}[t_i]$, the actual position $x_{AV}[t_i]$, the actual speed $v_{AV}[t_i]$ and the speed setpoint $v_{SP}[t_i]$ are made available to the parameter estimation unit as an input signal from the motion control in the base cycle of the PLC $t_i = 1ms$. A sign check of the feed rate is performed against the L previous measurement data in the input data memory, which has the cycle time t_j. This allows the detection of standstill phases and measurement noise and serves for data pre-filtering. Subsequently, the data is low-pass filtered and stored in the input data memory. The sequence of filtering in the data preprocessing is shown in Fig. 3(a). With the initialization of the algorithm, the data points $x_0 = x_{AV}[t_i]$ and $F_0 = F_{AV}[t_i]$ are defined as reference points $[x_{n0}, F_{n0}]$ and new data are provided cyclically. Averaging is performed as long as the following abort conditions are not met:

$$| x_{n0} - x_0 | > \varepsilon_x \; or \; | F_{n0} - F_0 | > \varepsilon_F \qquad (8)$$

When the position limit ε_x or the force limit ε_F is exceeded, the process is terminated and the mean value of the interval is transferred to the ring buffer of the input data memory. In addition, force and position values in the following cycle $[t_i + 1]$, are defined as the new reference data points and averaging starts again. This corresponds to a digital FIR low pass filtering whose filter length is controlled by the abort condition (8). Thereby it is ensured that a model excitation is present and no invalid data sets with $\Delta x = 0$ are used. The generation of the input data for the RLS-algorithm with the cycle time t_k is based on the mean values $x[t_j]$ and $F[t_j]$ and it is shown graphically in Fig. 3(b). Overlapping data points are used for a balance between high ΔF-excitation and fast response to stiffness changes. This overlap is achieved by means of the ring buffer of length L in the input data memory. It corresponds to the difference between the newest and oldest mean value, so that $L - 2$ values are skipped. In this way, suitable RLS input data $\Delta x[t_k]$ and $\Delta F[t_k]$ can be generated if the measurement data are strictly monotonically decreasing or increasing over sufficiently long time periods. Typically, this applies to the position and the correlated force. In a final step, after parameter estimation, data post-processing can take place by low-pass filtering.

4.2. Design of the Parameter Estimation Model

First, the effects of individual influencing factors in the data pre-processing are investigated. In the following, a suitable design parameter set for the parameter estimation unit will be determined. Synthetic data are used for this purpose, since they can be used to generate exact stiffness curves with known parameters.

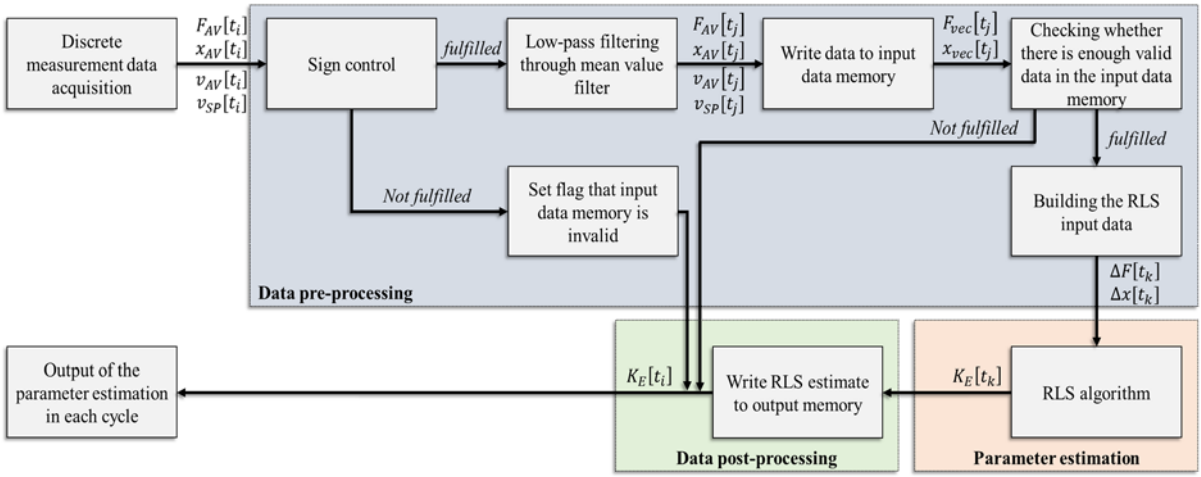


Fig. 2. Structure of the presented method for estimation of the stiffness.

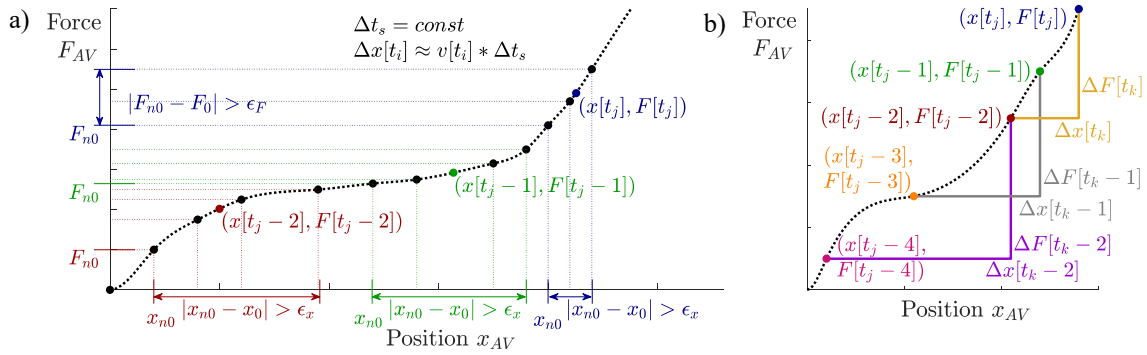


Fig. 3. Sequence of averaging in data pre-processing from measured data (a) and generation of the input data for the RLS-algorithm with an input data memory length of $L = 3$ (b).

The setpoint and actual force F_{SP} and F_{AV}, the actual position x_{AV} as well as the setpoint and actual velocity v_{SP} and v_{AV} are required in the estimation model. These signals have to be provided simulatively with a correct sampling frequency, a correct quantization and a correct noise by the machine model. Furthermore, F_{AV} is calculated in the test setup from the moving average of the current and the last nine measured values. The measured values show noise and are quantized to 1 mN. This is also taken into account in the model. To simulate the behavior of the parameter estimation unit, a realistic modeling of this signal is particularly relevant. In order to determine the variance of the force signal, the machine is loaded from $0N$ to $5kN$ at a typical stiffness value $K_E = 500\frac{N}{mm}$. The force set points of $1kN$, $2kN$, $3kN$, $4kN$ and $5kN$ are held for $2s$ each. From the steady state within these holding periods, the variances were determined. The mean variance of the quantized measurement signal is $\sigma_F^2 = 0.513N^2$. The position signal at standstill shows noise with a variance of $\sigma_x^2 = 1.9585 * 10^{-14}\ mm^2$. Furthermore, the actual velocity shows a noise with a variance of $\sigma_v^2 = 3.4985 * 10^{-8}\ \left(\frac{mm}{s}\right)^2$. To simulate this noise using Matlab Simulink, a stochastic process of mean-free, Gaussian-distributed pseudo-random numbers with a sampling time of $1ms$ is added to the actual values. The test data for the investigation of influencing factors consist of movements at a feed rate of $0.2\frac{mm}{s}$, $0.4\frac{mm}{s}$ and $1\frac{mm}{s}$ into a spring with a stiffness of $100\frac{N}{mm}$. The stiffness is set to $500\frac{N}{mm}$ after one second. Then the feed rate is kept constant for $500\ ms$. The sampling frequency is $1\ kHz$. The results are illustrated in Fig. 4. It can be seen that varying the filter length L of the ring buffer has little effect and the estimation behavior improves with higher values. A similar effect can be observed for the position limit ε_x.

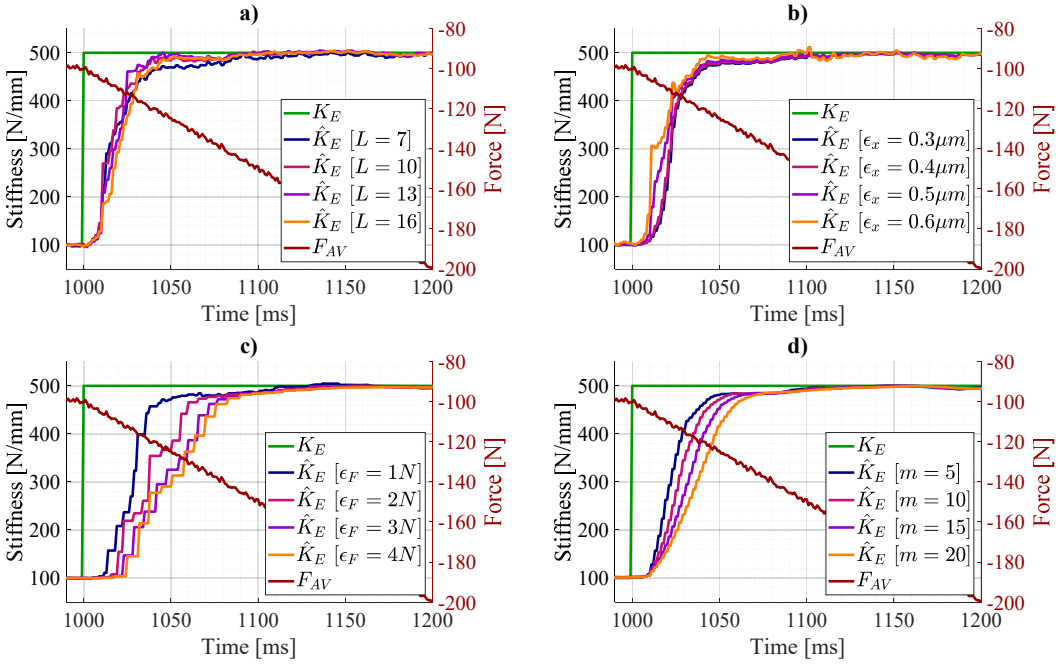


Fig. 4. Influence of different settings of the parameter estimation unit on the behaviour of the estimation. a) Variation of the length L of the input data memory; b) Variation of the position limit ε_x; c) Variation of the force limit ε_F; d) Variation of the average filter length m.

The parameter estimation model reacts faster and there is a better approximation to the nominal value for smaller force limits ε_F. This can also be observed for the average filter length m. The estimation of $\widehat{K}_E$ agrees after 40-50ms in nearly all cases with the setpoint jump of K_E to 90%.

4.3 Validation Steps

The aim of the validation of the parameter estimation unit is to analyse whether the stiffness $\widehat{K}_E$ estimated on the basis of the input data agrees with the real stiffness of the test setup K_E. The validation is carried out in three steps and thus three different procedures: (I) in Matlab using synthetic data with a precisely known stiffness; (II) in Matlab using data from the real test machine recorded in advance with the conventional force controller; (III) online in the Beckhoff motion control system using real-time data from the experiments. For cases (I) and (II) the parameter estimation unit is used on a development PC in Matlab Simulink. For case (III) it is implemented as a TcCOM object (Matlab Simulink model in real-time environment of TwinCat) on the machine controller. In step (I), individual influencing factors were first varied and examined. The tests and results are described in chapter 4.2 and essentially served to constrain the possible settings of the parameter estimation unit.
The determined values thus provide an orientation. However, these parameters also influence each other and are additionally dependent on the feed rate. For step (II), a suitable combination was selected in the simulation with an empirical variation of the parameters. The final parameter set for data pre-processing is summarized in Table 1. The simulation results with recorded measurement data from the machine are illustrated in Fig. 5. A ramp-shaped force profile with standstill phases is used. The influence of the adjustment factor γ is investigated here. Larger values lead to a faster estimation of the stiffness (see Fig. 5(b)), but this also increases noise effects (see Fig. 5(c)).

Table 1. Determination of the design parameter set for the RLS algorithm.

ε_x	ε_F	L	m
0.323 μm	1.294N	18	5

A good compromise is achieved with $\gamma = 0.03$. It can also be seen that the algorithm detects standstill phases and provides reliable values for a constant stiffness, stiffness jumps and even when the direction of motion is reversed (see Fig. 5(a)).

In step (III), the model was verified with real-time data on the test setup during ongoing operation with one constant stiffness. Here the spring package was loaded with a force of 3kN and then unloaded again. The force-position curve and the stiffness estimated with the model from the actual values $\widehat{K}_E$ are shown in Fig. 6. The actual stiffness K_E is determined from Δx_P and ΔF_P in the range from the start position P_s to the end position P_e. The results are summarized in Table 2.

The stiffness profile in the area of 4-6mm results from deviations of the spring lengths and the non-linear range until all springs have contact. It can be seen that the estimated stiffness $\widehat{K}_E$ settles to a constant value with slight fluctuations after the contact area and the learning phase of the estimation model.

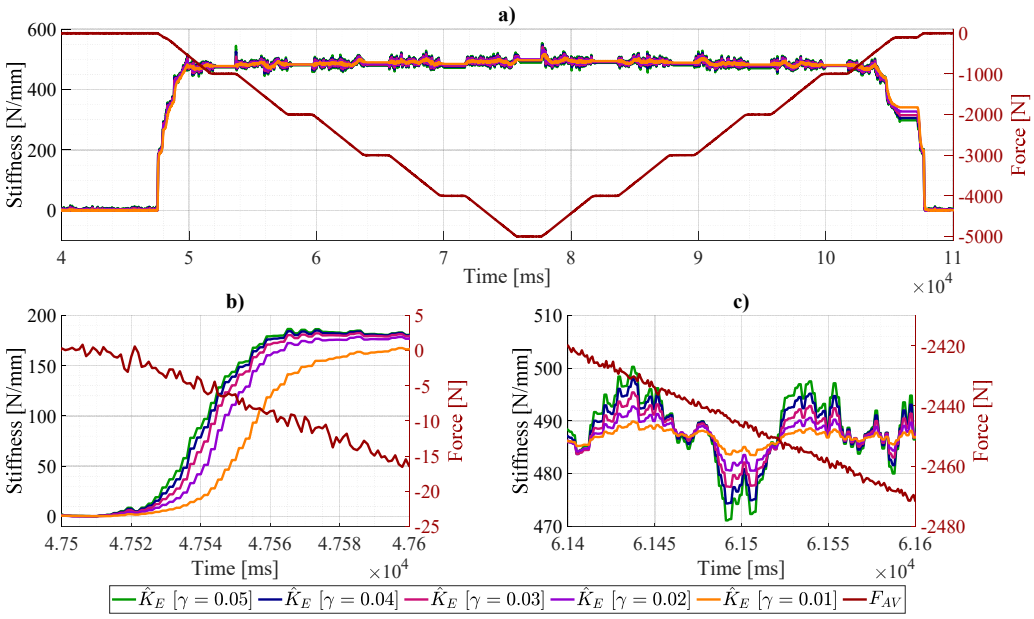


Fig. 5. Comparison of the parameter estimation $\widehat{K}_E$ for different adjustment parameters γ using real measurement data of the machine (simulation with pre-recorded measurement data). a) Overview of the entire stiffness estimation; b) Parameter estimation at the first stiffness jump K_E; c) Noise of the parameter estimation $\widehat{K}_E$.

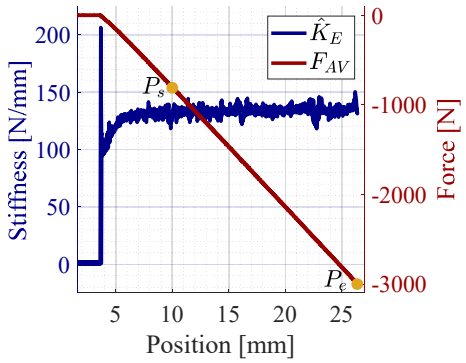


Fig. 6. Parameter estimation on the test setup.

A comparison between the estimated stiffness $\widehat{K}_E$ and the actually calculated stiffness K_E using the values in Table 2 shows very good agreement. Open-loop tests with several abrupt changes in stiffness were also carried out on the test setup. Here, as in the simulation, the estimation model provided reliable values for stiffness steps.

Table 2. Data for verification of the parameter estimation unit with real measured values on the test setup.

Data		**Plateau *P***
Start position, End position, Δx_P	mm	$P_s = 10, P_e = 26.37,$ $\Delta x_P = 16.37$
Start force, End force, ΔF_P	N	$F_s = -809, F_e = -3000, \Delta F_P = -2191$
Stiffness K_E determined from Δx_P and ΔF_P	$\frac{N}{mm}$	$K_E = 133.84$
Approximated estimation $\widehat{K}_E$ of the parameter estimation unit	$\frac{N}{mm}$	$\widehat{K}_E = 130 \dots 140$

5. Summary and Conclusion

In this paper, a model for estimating the process stiffness was presented. It is composed of the three components data pre-processing, parameter estimation, which is based on an RLS-algorithm, and data post-processing. The verification of the model was done in three steps by simulations and experiments. The influence of different design parameters was also investigated. It has been found that the estimated stiffness $\widehat{K}_E$ agrees sufficiently accurate with the actual stiffness K_E for both real and synthetic data. In addition, the developed model provides reliable results even in the face of various challenges, such as stiffness steps, reversal of the direction of motion and standstill phases. With the methodology described in this publication and the developed parameter estimation unit, the basis for the design of an adaptive force control has been created. However, it is noted that the estimated stiffness $\widehat{K}_E$ is delayed from the real stiffness between 30-50*ms* depending on the design parameters of the data pre-processing, the feed rate and the cycle time of the control. The implications for adaptive control will be considered in future research. Furthermore, an optimization of the empirically set design parameters is also intended in further work.

References

[1]. J. M. Allwood, et al., Closed-loop control of product properties in metal forming, *CIRP Annals – Manufacturing Technology*, Vol. 65, 2016, pp. 573-596.

[2]. X. Yao, et al., Machining force control with intelligent compensation, *The International Journal of Advanced Manufacturing Technology*, Vol. 69, No. 5-8, 2013, pp. 1701-1715.

[3]. W. Leonhard, Control of Electrical Drives - Power Systems, 3rd edition, Heidelberg, Germany: *Springer*, Berlin Heidelberg, 2012.

[4] I. D. Landau, et al., Adaptive control - Algorithms, analysis and applications, Communications and control engineering, 2nd edition, London, *Springer*, 2011.

[5] K.-P. Schulze and K.-J. Rehberg, Entwurf von adaptiven Systemen - Eine Darstellung für Ingenieure, Vol. 1, Berlin, *VEB Verlag Technik*, 1988.

[6] R. Neugebauer, et al., Non-invasive parameter identification by using the least squares method, *Archive of Mechanical Engineering*, Vol. 58, 2011, pp. 185-194.

[7] R. Isermann, Identification of dynamic systems: An introduction with applications, *Springer*, Berlin, 2011.

[8] K. J. Åström and B. Wittenmark, Adaptive Control, 2nd edition, Dover Books on Electrical Engineering. Newburyport, *Dover Publications*, 2013.

[9] S. Beineke, Online-Schätzung von mechanischen Parametern, Kennlinien und Zustandsgrößen geregelter elektrischer Antriebe, Dissertation, Universität Paderborn, Fortschritt-Berichte VDI Reihe 8, Meß-, Steuerungs- und Regelungstechnik, Düsseldorf, *VDI Verlag*, 2000.

[10] R. Neugebauer, et al., Time-Based Method for the Combined Identification of Velocity-Loop Parameters, *Archive of Mechanical Engineering*, Vol. 58, 2011, pp. 175-184.

[11] D. W. Clarke and Ch. J. Hinton, Adaptive control of materials-testing machines, *Automatica*, Vol. 33, Issue 6, 1997, pp. 1119–1131.

[12] A. Sewohl, et al., Performance analysis of the force control for an electromechanical feed axis with industrial motion control, in *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics 2020*, Paris, France, Vol. 1, 2020, pp. 667-674.

[13] D. W. Clarke, Adaptive control of servohydraulic materials-testing machines: a comparison between black- and grey-box models, *Annual Reviews in Control* 25, 2001, pp. 77–88.

(007)

Simulation of Automated Handling in Textile Manufacturing of US Military Apparel to Improve Efficiency and Quality

Z. B. Rosenberg, J. A. Joines and J. S. Jur
Department of Textile Engineering Chemistry and Science, North Carolina State University, Raleigh, NC, USA
E-mail: zbrosenb@ncsu.edu

Summary: In this paper, we simulate implementing automated handling for transferring fabric part pieces throughout the construction process of US military-relevant textile products. The current infrastructure of the textile industry would benefit from integrating new automated technology in the textile production line to replace manual operations. Using Simio software, we simulated the current production process for the cummerbund in the US Army's Modular Scalable Vest (MSV). The current production is done through batch production with manual transferring of the batches of parts to the next workstation and the loading/ unloading of part pieces into sewing workstations. Using SIMIO simulation software, we can model the implications on time and labor of replacing the manual operations with automated robotic handling to enable continuous flow manufacturing. Incorporating automation in the manufacturing of the cummerbund increases efficiency, throughput, and quality of production.

Keywords: Automated handling, Materials handling, Textile manufacturing, Digital simulation, Industry 4.0.

1. Introduction

There is growing demand to deliver textile products, such as apparel, to the customer faster. Automation can enable manufacturing on-demand and shorter lead times by using robotics to construct products. Using robotics replaces labor, allowing for continuous production and increased throughput. Additionally, automated equipment provides more consistent construction than manual assembly, generating improved uniformity.

The assembly of textile products remains largely dependent on manual operations [1]. Minimal automated processes have been inserted into the garment construction process due to the challenge of developing proper textile handling equipment as well as the economic feasibility of such an investment. The textile industry has remained economically viable without automation because skilled, inexpensive labor is widely available in areas such as Central America and Southeast Asia, making it difficult to justify manufacturing in high labor cost countries [2]. Implementing automated processes provides the prospect to manufacture textile products in high labor cost countries and to increase quality and consistency in the final product.

Typical production facilities for textile products have multiple sewing workstations that perform a single assembly step for the product. Production is done in a batch process, where a "batch" of products is completed at each step before moving to the next production step [3]. Batch processing is more effective with the current infrastructure since the operator does not need to move individual pieces between construction steps and instead can move a grouping, making production more time-efficient. Batch production also allows for flexibility in the case of a piece of equipment going down, allowing other workstations to continue, and preventing the entire production line from stopping.

Some automated equipment has been inserted into production floors; however, such equipment still depends on operators (human-assisted automation), making it challenging to implement floor plans with continuous flow production and on-demand manufacturing [4]. The automated textile equipment being employed most broadly consists of sewing machines that automatically move fabric pieces through the sewing operation. However, the output of the automated sewing machines is limited by the time needed manually to load and unload the machine quickly and accurately. The operator can incorrectly load the sewing machine where the textile parts are not accurately placed, causing a wrinkle or misalignment in the component and leading to a defective product. Because of the complexities in handling fabric, this equipment is primarily available for sewn flat pieces or sewing simple, repeated folds such as the pleat in the back of a button-down shirt. Using automated handling for the loading process could improve the construction quality of the products and increase the throughput of the manufacturing process.

The assembly process of a cummerbund, which is part of the US Army's Modular Scalable Vest (MSV) already utilizes five types of human-assisted automated equipment. Additionally, the MSV fabric part pieces can be handled with a robotic arm because of its high stiffness and dense woven structure. We have modeled the production process using Simio to simulate the implications of including additional automated procedures in the assembly. Because of the integration of automated handling, continuous flow manufacturing becomes a viable construction method to improve production efficiency.

2. Methods

In our research, we used the discrete event simulation software SIMIO to simulate the implications of adding automated handling and procedures in the assembly of a cummerbund. The manufacturer has provided data about their current production that is used to drive and validate the simulation. Various production strategies of inputting a robotic arm to replace an operator were modeled to determine the most effective implementation of robotic handling.

2.1. Manufacturing Process of MSV Cummerbund

The current assembly of the MSV cummerbund consists of 12 construction steps. These steps all occur at different workstations manned by operators of various skill levels depending on the operation. The assembly is done through batch production in batches of 20 pieces, meaning 20 pieces are assembled before the batch can move to the next workstation. The manufacturer provided the average times to complete each step and the time required to transfer the batch to the next workstation. These assembly steps and times can be found in Table 1. The current assembly process incorporates five pieces of human-assisted automated sewing equipment, which still requires an operator to load, position, and unload the part pieces into the machine and transfer the completed batch to the next assembly station. This equipment reduces the necessary skill level of the operator. The human-assisted automated operations are indicated with an (A).

Table 1. Assembly steps of cummerbund with current production and transfer times. Automated processes indicated with an (A).

Cummerbund Assembly Step	Production Time (seconds)	Transfer Time of Bundle (seconds)
A	63.5	15
B	14.6	45
C (A)	50.9	20
D (A)	66.0	11
E (A)	33.0	14
F (A)	60.0	15
G	86.8	15
H (A)	56.7	10
I	58.7	8
J	36.0	8
K	98.6	8
Total Time	10 mins 24.8 secs	2 mins 49 secs

2.2. Simio Model of Cummerbund Assembly

The factory floor was modeled in Simio to simulate the current production process. We ran twenty replications of the model and averaged the results to determine the daily production and timing of the cummerbund assembly. In this paper, we explored four alternate production process variations of the factory floor with various levels of automation. The models included implementing automated handling with batch production, automated handling with continuous manufacturing, automated handling with continuous manufacturing and redistributed labor, and all automated processing.

2.3. Timing of Automated Handling

The theoretical production time was calculated for the automated processes with implemented automated handling. Employing robotic handling decreases the production time required for each automated step by removing the time it takes for the operator to grasp the part pieces, position them accurately, and then move the equipment into place and then place the completed part in the pile for the next step. The theoretical production time was determined through video analysis of the current construction process. The theoretical production time was defined as the time for the machine to start and finish the sewing process, removing the time of handling, grasping, placing, and transferring the part pieces. These times are shown in Table 2. The time for handling, placing, and transferring the part piece was specified in the SIMIO model as taking five to ten seconds for each automated assembly operation. This time accounts for both the timing of the robotic arm operations as well as encompassing the extra time in case a robotic arm encounters issues or were to malfunction or need repair.

Table 2. Production time of automated assembly steps of cummerbund when using robotic handling. Automated processes indicated with an (A).

Cummerbund Assembly Step	Production Time (seconds)
C (A)	28
D (A)	27
E (A)	10
F (A)	6
H (A)	7

The transfer time for the automated processing with automated handling is dependent on how the cummerbund is manufactured, whether it is batch or continuous. In batch production, the time to transfer remains the same because the completed parts are stacked into piles and still need to be moved as a batch to the next workstation, despite the automated loading and unloading of the fabric. When the production changes to continuous manufacturing, the transfer time becomes negligible because the sequential production steps would need to be positioned closely together for the robotic arm to reach between the steps. The robotic arm transfers the part from the completed assembly workstation into the next assembly operation. The model accounts for the transfer time as the five to ten

seconds added to the previously mentioned robotic assembly operations to represent the loading and unloading of parts.

3. Results and Discussion

3.1. Simio Model of Current Production of Cummerbund Assembly

The current factory layout and results from the simulation are shown in Fig. 1 and Table 3, respectively. In this simulation model, each assembly step was given a 10 % variance in the production time using a Beta Pert distribution. The variance is added to account for the variability of a human operator having some faster and some slower production times. The variance of a human operator can range depending on the operator's experience, the difficulty of the operation, and fatigue.

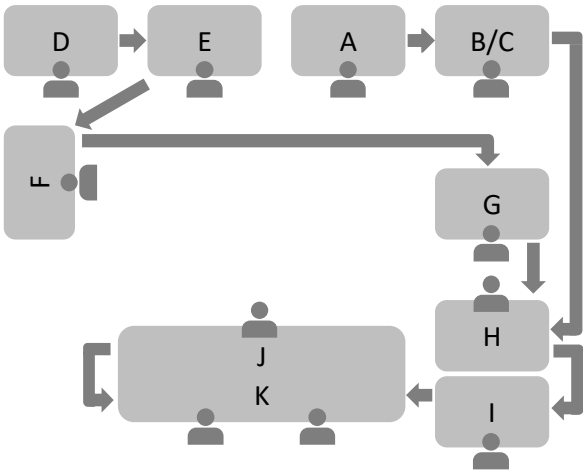


Fig. 1. Current production layout.

Table 3. Production output of current factory layout in simulation.

Current Production	Output
Number of Operators	11
Number of Robots	0
Hours in Work Day	8
Average Time in System	660.43 minutes
Number Produced per Hour	39.72 units/hour
Average Numbers	317.79 units/day

The simulation of the current production accounted for an eight-hour workday. For this production line, the manufacturer employs eleven operators, or one shifts worth, due to the labor shortage of seamstresses in the United States. Currently, the manufacturer is not able to hire two shifts worth of sewing operators due to the scarcity. The current production line produces 317.79 cummerbunds a day, or 39.72 units per hour. On average, the cummerbund is in the system for 660.43 minutes. This is the time it takes for the fabric components to enter the operation and those exact components to be in a completed cummerbund. The long time in the system is attributed to the bottlenecks within the assembly. The modeled output aligns with the manufacturer's current production capabilities that were provided. Even optimizing the system by adding an additional operator and station at two bottleneck operations, the number produced per day can only increase 419.8 units/day with the average time in the system 150.8 minutes.

3.2. Scenario 1: Automated Handling with Batch Production

Robotic arms were added to the model to simulate automated loading and unloading of the fabric pieces into the automated equipment shown in Fig. 2.

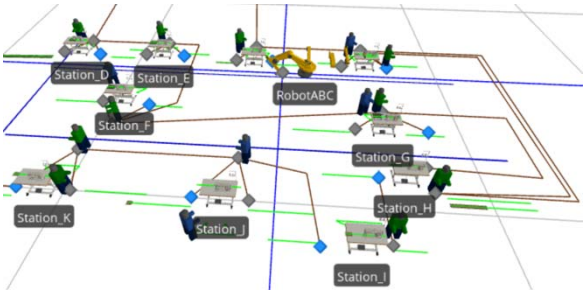


Fig. 2. Simio model of current factory facility with an integrated robotic arm for handling fabric.

The implementation of robotic arms reduced the average production times to the times provided in Table 2. The variance was reduced to 5 % for the automated processing steps to reflect the consistency and reliability that automated handling offers. The production remained in batch production, where 20 units had to be completed before transferring to the next workstation. The automated handling enabled an increased workday of sixteen-hours to account for the reduced labor that could be redistributed to two shifts. This is assuming the manufacturer is able to employ one additional operator, or reallocate an operator from a different production line. The results of automated handling within the workstation are shown in **Table 4.**

Table 4. Production output of current production with inserted robotic handling to load/unload automated equipment.

Automated Batch Production	Output
Number of Operators	6
Number of Robots	5
Hours in Work Day	16
Average Time in System	628.38 minutes
Number Produced per Hour	39.17 units/hour
Average Numbers Produced	626.67 units/day

The daily production of cummerbunds increased to 626.67 units, nearly doubling the current production, and a minor improvement in the time in the system to 628.38 minutes, signifying a slight reduction in the bottlenecks of the system. However, the cummerbunds produced per hour marginally decreased to 39.17 units per hour. This is a result of the fabric needing to be handled twice, both by the robotic arm for the automated assembly operation as well as by the operator to transfer the batch. Even though the timing of the assembly steps is reduced with automated handling, having to transfer each part piece twice minimizes the effectiveness of robotic handling. Nonetheless, the daily production nearly doubled because robotic handling enabled a sixteen-hour workday, improving daily output.

3.3. Scenario 2: Automated Handling with Continuous Manufacturing

The model with implemented robotic handling was then modified for continuous manufacturing. In the continuous manufacturing model, each unit moves individually through the assembly process. Continuous manufacturing provides a more suitable method of production with robotic handling. The transferring and placing of the part pieces into the assembly operations are combined, and the part pieces can move continuously throughout the production. For continuous flow, the machines have to be located directly next to one another so the robotic arm can reach both workstations for transferring the fabric part. The continuous manufacturing simulation layout is shown in Fig. 3.

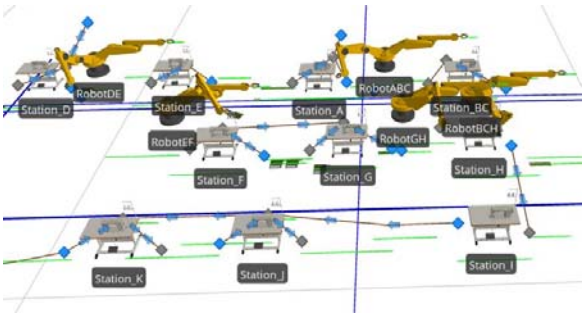


Fig. 3. Simio model of the facility with integrated robotic arms at the automated workstations for handling and transferring fabric.

Automated handling provides more consistency and reliability than manual operations. This facilitates controlled timing of the production steps to enable a continuous process flow. Batch production is more suitable for manual operations because it allows flexibility in the timing of the production. If one workstation is delayed, whether the equipment goes down or the operator needs a break, the other workstations can continue to function. In continuous manufacturing, there is less leeway for inconsistent timing because each construction step depends on the previous step to produce in time. Continuous or lean manufacturing requires a reliable supply of materials to keep the assembly process running.

Switching to automated material handling offers consistent production to enable continuous manufacturing. The cummerbund assembly has five automated processes that all occur during the early stages of the construction process. The results of the continuous automated assembly are shown in Table 5.

Table 5. Production output of continuous production with inserted robotic handling and automated loading/ unloading at automated workstations.

Automated Continuous Production	Output
Number of Operators	6
Number of Robots	5
Hours in Work Day	16
Average Time in System	1282.55 minutes
Number Produced per Hour	34.97 units/hour
Average Numbers Produced	559.54 units/day

With robotic handling and continuous manufacturing, the daily production produces 559.54 units per day and 34.97 units per hour. This is an improvement from the current production but a reduction from the automated handling in batch production. This diminished value of introducing automated handling is represented in the increase of the average time in the system, which increases to 1282.55 minutes owing to the bottlenecks in the system. In this production process, the automated processes all occur during the early stages of production. The later steps remain manual and have longer manufacturing times. Because each operation is dependent on receiving the completed material from the prior operation, it must wait for that operation to finish. Since the assembly operations all take a different amount of time to complete, the shorter operations are held up by the longer operations. The bottlenecks in continuous manufacturing have a higher impact than batch production because the operations are more dependent on one another for efficient production.

3.4. Scenario 3: Automated Handling and Redistributed Labor with Continuous Manufacturing

The output for continuous manufacturing is lower than the output for batch production, which is attributed to the bottlenecks in the system. However, continuous manufacturing reduces the overall handling time by combining the assembly operation handling with the transfer handling. We redistributed three operators from the original operations and placed them at the bottleneck process or longer operation steps to

improve continuous production. We modeled having two operators at operations G, I, and J. However, with employing more workers on the production assembly, the assembly operation returned to one shift, or an eight-hour workday. This is assumed due to the likelihood of not being able to hire or reallocate seven additional operators. The results are shown in Table 6.

Table 6. Production output of continuous production with inserted robotic handling at automated workstation and an additional operator at operations G, I, and J.

Automated Continuous Production with Redistributed Labor	Output
Number of Operators	9
Number of Robots	5
Hours in Work Day	8
Average Time in System	971.83 minutes
Number Produced per Hour	47.28
Average Numbers Produced	378.28

Redistributing the labor showed improvements in production output and a reduction in bottlenecks. The cummerbunds produced per hour increased to 47.28 units and the average time in systems decreased to 971.83 minutes. However, due to the eight-hour workday, the daily output fell to 378.23 units a day, an improvement from the current output, but a regression from both automated assemblies with sixteen-hour workdays.

3.5. Scenario 4: All Automated Processing

The overall goal is to automate the textile assembly process, which enables more consistency in production, reduced labor, faster production times, and the ability to manufacture on demand because the equipment can run 24/7 and requires less lead time to hire and train production workers.

Automating the full assembly would require the development of automated equipment for the additional operations. We determined the timing of the theoretical automated equipment that would need to be produced based on the video analysis process done in section 2.3. The updated assembly times are shown in Table 7.

Operation J is removed when the entire process is automated because it is no longer necessary with automated equipment. Operation K is not able to be automated since it is the final inspection, however, we assume that the time will be reduced by approximately a third, due to the consistency and accuracy of automated equipment.

The all automated process was modeled with the times shown in Table 7 and 10 seconds for the robotic handling to place, handle, and transfer the parts at all the assembly operations. Each production step was given a 5 % variability, and the workday was modified to 24 hours of production since the assembly operations are conducted by primarily robotic equipment that can operate continuously without breaks. There is one operator at K, which can be filled by three shifts to account for the 24-hour workday. The results of continuous automated production are shown in Table 8.

Table 7. Production time of automated assembly steps of cummerbund when using robotic handling and automated equipment.

Cummerbund Assembly Step	Production Time (seconds)
A (A)	5
B (A)	5
C (A)	28
D (A)	27
E (A)	10
F (A)	6
G (A)	15
H (A)	7
I (A)	20
K	33
Total Time	221.6 seconds

Table 8. Production output of continuous production with robotic handling and automated assembly at operations A-I.

All Automated Processing	Output
Number of Operators	1
Number of Robots	9
Hours in Work Day	24
Average Time in System	4.23 minutes
Number Produced per Hour	97.2 units/hour
Average Numbers Produced	2332.8 units/day

Daily production increases to 2,332.8 units with the implementation of all automated equipment. The bottlenecks are alleviated, representative by the time in system decreasing to 4.23 minutes, similar to the sum of the assembly operations, signifying efficient production. The units produced per hour more than doubles to 97.2 units. Additionally, by implementing automated handling, we can infer the quality of the product will improve from the consistency of robots compared to a human operator. However, actualizing an autonomous assembly of a textile product would require significant endeavors and investments in automated equipment.

3.6. Comparison of Automated Insertion Scenarios

Robotic arms were inserted to fully automate the current human-assisted automated assembly steps in the production of an MSV cummerbund. A summarization of the output for the different types of scenarios discussed in this paper is shown in Table 9.

Introducing robotic handling in the assembly of the MSV cummerbund reduces the number of operators required for the assembly process. However, robotic handling does not necessarily increase the speed of production, but instead allows labor to be redistributed,

enabling two shifts of operators, doubling the production hours. Batch production serves as a more suitable production process when there are longer human operations still in the assembly process. If available, labor should be redistributed to the longer operations to decrease the bottleneck effects. However, if redistributing the operators reduces production to one shift, the daily output will decrease despite an increase in the hourly production. An all automated production process has the most efficient assembly and operates for twenty-four hours thus significantly increasing the daily output.

Table 9. Comparison of the production output of continuous production with robotic handling and for the current manufacturing process and the four alternate scenarios.

Scenario	Current	1	2	3	4
Operators	11	6	6	9	1
Robots	0	5	5	5	9
Units/Hour	39.72	39.17	34.97	47.28	97.2
Units/Day	317.79	626.67	559.54	378.28	2332.8

4. Conclusions

Simio software was used to simulate the implications of inserting automated handling in the assembly of an MSV cummerbund. When automated handling is employed at the automated equipment operations in the current batch production process, the daily production of cummerbunds nearly doubles to 626.67 units from improved efficiency and introducing a sixteen-hour work day. Using automated handling in continuous flow production is less effective producing 559.54 units per day. Reallocating the saved labor can improve the bottlenecks in continuous manufacturing by increasing hourly production from 34.97 units per hour to 47.28 units per hour, but due to the increased labor, an eight-work hour day must be employed, reducing daily production to 378.28 units per day. With the creation of automated equipment for the remaining four assembly steps, autonomous continuous flow production is possible in producing 97.2 units per hour and 2332.8 units per day due to the capability to run twenty-four hours. Automated handling of textiles is essential to automating the assembly of textile products. Automating the handling of textiles improves efficiency for production by reducing assembly times and enabling longer production hours. Automation in textile handling is most beneficial when the entire process can be automated, alleviating the bottlenecks that come with manual operations in a continuous production line. However, automated equipment, including machinery and robotic arms, requires a significant investment, so it essential to assess the value of adding automation to the production process before investing.

Acknowledgments

This work was funded through the University of North Carolina Systems Office and the United States Army Combat Capabilities Development Command.

References

[1]. International Labour Office- Geneva, The future of work in textiles, clothing, leather and footwear, *International Labour Office, Sectoral Policies Department. – Geneva: ILO*, 2019. Working Paper: No. 326.

[2]. R. Handfield, H. Sun, L. Rothberg, Assessing Supply Chain Risk for Apparel Production in Low Cost Countries Using Newsfeed Analysis, *Supply Chain Management*, Vol. 25, No. 6, 2020, pp. 803–21.

[3]. K. Kulkarni and R. Adivarekar, Developments in Textile Continuous Processing Machineries, in Advances in Functional Finishing of Textiles, *Springer Singapore,* 2020.

[4]. M. Suh, Automated Cutting and Sewing for Industry 4.0 at ITMA 2019, *Journal of Textile and Apparel, Technology and Management,* Special Issue: ITMA, 2019.

(008)

Multi-Robot Cooperative SLAM Using Panoramas

J. Y. Feng and Z. XuanYuan
[1] Beijing Normal University - Hong Kong Baptist University United International College, ZhuHai, China
Tel.: +86-131-43116416
E-mail: 530300865@qq.com, zhexuanyuan@uic.edu.cn

Summary: Simultaneously Mapping And Localization (SLAM) systems are essential in robotics systems. Using a single robot to map a large scene is time-consuming. Therefore, a map can be constructed using a group of robots working together. Visual-based loop detection in cooperative SLAM is essential, which can help to efficiently and accurately merge and construct global maps among robots working independently. However, most visual SLAM algorithms focus on loop detection along the same trajectory direction. These algorithms cannot handle multi-robot cooperative systems well because, in such systems, multi-robot trajectory direction may often form loops along opposing or perpendicular directions. This paper proposes a multi-robot cooperative SLAM system based on panoramic images. Each robot uses a camera as a sensor and a local map algorithm to construct a local map. All local maps are merged into a global map when loops are detected. The panoramic images provide scene information in all directions for loop detection. Experiments show that the loop detection accuracy improves in the proposed system, and the time for constructing a global map is significantly reduced using a multi-robot cooperative system.

Keywords: Visual SLAM, Cooperative mapping, Panoramic image.

1. Introduction

In different application scenarios such as autonomous driving and robot navigation, it is usually necessary to do mapping and self-localization in large scenes, such as a whole city scale. In this case, running SLAM with a single robot is time-consuming. In addition, in map building, the pose errors will accumulate as the traveling distance increases. Maps can be divided into several sub-maps and built using a group of robots, which can then merge into a global consistent map by sharing information when loops are detected. In this case, the distance of each robot is shorter, the running time is shorter, and the accumulated error is also reduced.

Fig.1 shows the comparison of sample images in the same and different trajectory directions from the KITTI dataset [1]. The appearance of two images in the same trajectory directions is similar, but the appearance of two images in different trajectory directions is quite different. This leads to the fact that although monocular visual SLAM can complete map building efficiently and accurately, it can only perform loop detection along the same trajectory direction, but not in opposing or perpendicular directions, which is not a practical and efficient assumption for multi-robot cooperative systems because the path planning for efficient multi-robot cooperation may result in rendezvous (loopback) of robots along with different trajectory directions and viewpoints. SLAM Algorithm that is robust to viewpoint variation is crucial for multi-robot cooperative systems.

This paper proposes a cooperative visual SLAM system based on panoramic images, using multiple robots to collect scene maps and build global maps. With panoramic images as input, 360-degree scene information can be obtained. On this basis, it can effectively solve the problem of loop detection along with all trajectory directions. The main contributions of this work include the following:

1. A loop detection algorithm for a multi-robot SLAM system based on panoramas is proposed to enhance the accuracy and efficiency of the cooperative mapping tasks.
2. A multi-robot cooperative mapping system has been designed and implemented to test the efficacy of the loop detection algorithm on different datasets.

Fig. 1. Comparison of images in the same and different trajectory directions in KITTI Dataset [1].

2. Related Work

In the following, we independently conclude with a review of visual SLAM, fisheye-based visual SLAM, visual-based cooperative SLAM, and place recognition.

2.1. Visual SLAM

The early visual SLAM algorithms are mainly based on filtering [2], [3]. The filtering-based method is inefficient and easy to accumulate linearized errors.

Afterward, keyframe-based approaches were proposed. It only uses the selected frame as the keyframe to estimate the map. PTAM [4] was the first keyframe-based SLAM algorithm proposed to separate the two functions of tracking and mapping in the SLAM system as two threads. ORB- SLAM2 [5] is one of the most popular visual SLAM. It adopts the multi-threaded framework of PTAM, but the ORB[6] features are extracted. Loop detection thread is added, which uses an efficient search algorithm BoW(Bag-of-Words) [7].

2.2. Fisheye Based Visual SLAM

Because a single camera's FoV (Field of View) is limited, extracting feature points and detecting the correct loop in some indoor and outdoor scenes is not easy. The fisheye camera has a wide FoV and can collect more scene information. The challenge of using the fisheye camera is how to solve the fisheye camera's distortion effectively. In [8], cubemap is used to improve the distortion of the fisheye camera. The method uses five planes: up, down, left, right and front, and then projects the spherical image separately. [9] is a multi-fisheye camera SLAM system. This method requires the calibration of a polynomial model of the fisheye camera. The camera parameters are used for connection transformation between multiple cameras. [10] uses a multi-fisheye camera to obtain 360-view images and input them into the SLAM system for mapping.

2.3. Visual Based Cooperative SLAM

Generally, multiple-robot cooperative SLAM can be a centralized system [11]. In a centralized system, multiple robots equipped with sensors, communication components, and processors are used to complete the mapping task in a centralized system. Each robot can run visual SLAM independently. The computation for some tasks, such as map merging, is performed by a specific robot on the team or by an external agent [12]. The specific robot or the external agent communicates with other robots as the system's center. Other robots are independent and do not communicate with each other. The central server receives data sent by other robots and provides required data and feedback. [13], [14] are cooperative systems that use this structure.

2.4. Place Recognition

Place recognition [15] is of great significance to visual SLAM. One method is to extract local features. Hundreds of local features may be extracted from each image. BoW [7] clusters discrete extracted features and uses many images to train offline to generate a vocabulary containing multiple words. A typical vocabulary usually contains thousands of words, but FAB-Map 2.0 [16] can use a vocabulary as large as 10,000 words.

Some neural network-based methods have been applied to place recognition. NetVLAD [17] is modified based on VLAD [18] and directly embedded in the trainable CNN(convolutional neural network) architecture. [19] used the Edge Boxes object proposal method combined with a mid-level CNN[20] to identify and extract landmarks to complete place recognition.

In the past few years, feature-based methods have become one of the main technologies in place recognition. Its low computing cost and effectiveness can be applied to low processing and storage capacity devices [21]. The application scenarios of the multi-robot cooperation visual SLAM system are complex and huge. It increases the difficulty of pre-training of the neural network-based methods. The neural network-based approach requires large computation. In a centralized system, the central server needs to process a large amount of data from the client robots, and the use of neural network-based methods will increase its burden.

3. Methodology

3.1. System Framework

This cooperative SLAM system uses a visual odometer based on a panoramic camera as a sensor and a global place recognition based on panoramic images. The use of panoramic images solves the problem of loop detection in different trajectory directions in a multi-robot system.

As shown in Fig. 2, it is the workflow of the multi-robot cooperative SLAM system. The system contains a central server and multiple robots. Suppose robots in the system are equipped with a panoramic camera and wireless communication interfaces. Each robot starts from its initial position and independently constructs its local map. The robot's initial position is random, which depends on the user. The robots do not know the position and coordinate system of other robots. The visual SLAM algorithm used by each robot is similar as the algorithm described in [10]. The difference is that we have added a data transmission part in the loop detection. The central server communicates with multiple robots in the system. Each robot will send its map data to the central server through its wireless communication channel in robot movement. The central server maintains a globally consistent map and a corresponding sub-map for each robot. After the central server receives the data of each robot, it will reconstruct the keyframes and map points consistent with each robot and add them to the sub-map corresponding to each robot. And the central server will do place recognition among the sub-maps. When it is detected that different robots have come to common areas, the sub-maps corresponding to these robots will be merged to get a globally consistent map.

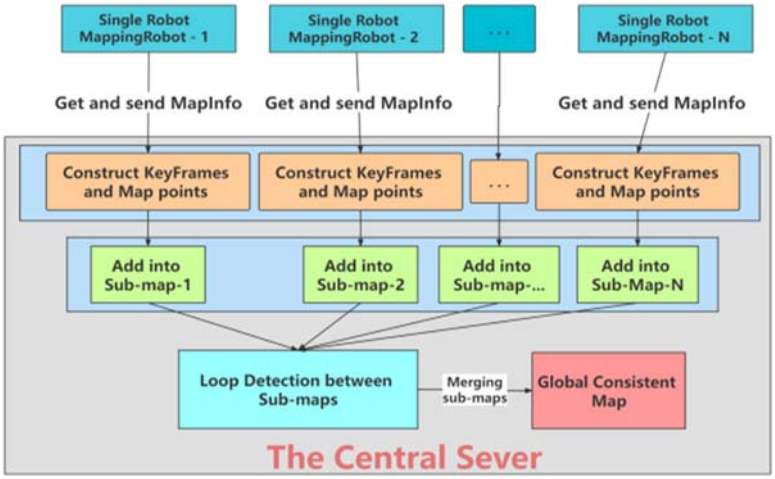


Fig. 2. The framework of the cooperative SLAM

3.2. Single Robot Mapping

Each robot is equipped with a LadyBug3 camera as a sensor. The panoramic images taken by the LadyBug3 camera are fed to local panoramic SLAM which has been modified based on [10]. Every robot will generate and maintain its own map. See an overview Fig. 3, there are four main threads in the local panoramic SLAM, tracking, local mapping, loop closing, and global optimization. The first three threads typically run in parallel. Global optimization is enabled after a loop is detected.

The tacking and the local mapping work together to complete the visual odometer function, which uses continuous panoramic images to produce a sparse map. This map shows the robot's trajectory and is used to achieve localization. The loop detection is in charge of place recognition. It detects whether the robot is at the same place it has been before. If the loop is detected, the local map near the current keyframes is optimized using the two keyframes of place recognition matches. Next, the global BA (Bundle Adjustment) thread is enabled. It will apply the optimization to all the keyframe poses, and map point coordinates in the map. These processes are used to reduce the accumulated errors in the robot movement.

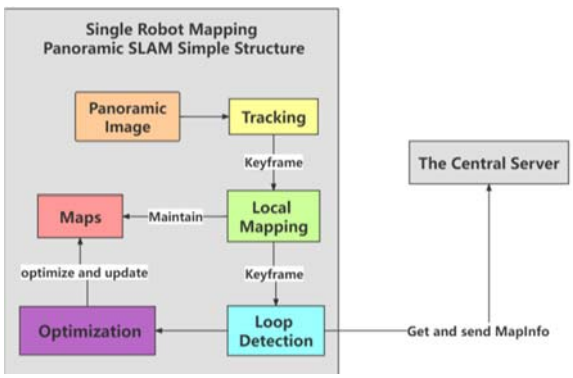


Fig. 3. The framework of the single robot mapping.

3.3. Communication Model

We transmit data mainly about keyframes and map points and their corresponding relationships. We use keyframes instead of all captured frames. For example, a dataset containing 498 frames, a map constructed by a single robot only contains 181 keyframes. Only transmitting keyframes greatly reduces the data that needs to be transmitted and ensures the map information's integrity.

We use LCM (Lightweight Communications and Marshalling) [22] to complete data transmission. LCM can encapsulate the data to be transmitted into a whole, facilitating data transmission and management.

Our goal is to continuously rebuild the map of each robot using the transmitted data in the central server, so there is no tracking module in the central server. The loop detection module only uses the information of keyframes and map points and does not update those information. Therefore, we chose to fetch the relevant data in the loop thread of each robot and transmit it.

In the transmission process, packet loss is inevitable due to the bottleneck of network communication. The central server can associate newly received data with the current map to reduce the impact of packet loss.

3.4. Loop Detection and Map Merging

3.4.1. Loop Detection on Panoramic Images

The loop detection in panoramic SLAM uses the method mentioned described in [15]. This method extracts the image's LDB (Local Difference Binary) descriptor. It extracts a binary string P_k for each sub-panorama I_k in a panoramic image I. The binary string of a panoramic image is calculated by concatenating the binary strings of n sub-panoramas. Calculating the similarity between panoramic images i and j can be done by associating the elements of P_i , P_j and generating a distance metrics M.

The loop detection in panoramic SLAM sets a fixed threshold to filter the appearance of keyframes. If the

distance metrics M calculated for the two keyframes is less than the threshold, then they are similar enough to be considered as a loop. This loop candidate pairs are then further verified using geometry-based validation.

We set 20 different thresholds for the loop detection of monocular images and panoramic images in the Kashiwa dataset [23], respectively. The process of loop detection does not include geometry-based validation. Fig. 4 describes the comparison of recall obtained from panoramic and monocular images. It can be seen that the recall obtained by using panoramic images is higher, which means more loops can be detected by using panoramic images.

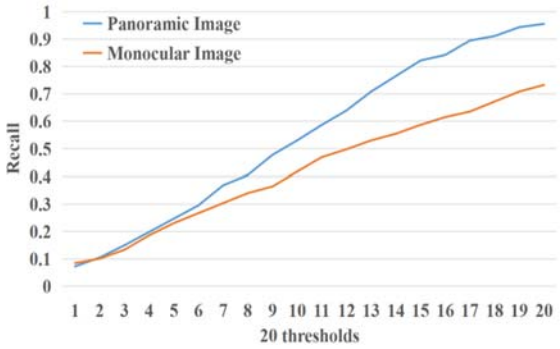


Fig. 4. Comparison of recall between panoramic images and monocular images.

Our experiments showed that the loop detection in the original panoramic SLAM can detect loops in the same trajectory direction, but it is sometimes challenging to detect loops in different trajectory directions. The reason is that the fixed threshold set is unsuitable for complex scenarios. As shown in Fig. 5, we calculated the distance between different keyframes and found that their loop frames fluctuates greatly in terms of distance. The smallest distance is less than 55,000, and the largest distance is greater than 90,000.

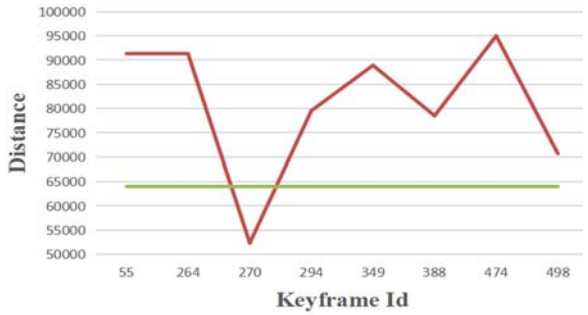


Fig. 5. The red line represents the value of distance calculated by several keyframes and their corresponding loop keyframes. The green line represents a fixed threshold of 64,000.

As shown in Fig. 6, it is the calculated distance ranges between a keyframe (id=270) and its covisibility keyframes, loop keyframes, and non-loop keyframes. We can get that the calculated distance range of the loop keyframe is closer to the covisibility keyframes than the non-loop keyframes varies greatly.

Based on these observations, we believed that applying a fixed threshold to all keyframes for loop detection is not optimal to deal with complex scenes. Therefore, we propose a distance filtering method based on an adapted threshold.

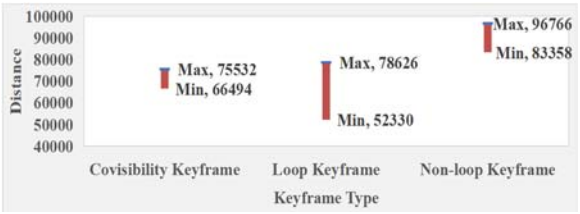


Fig. 6. The comparison of distance ranges calculated from the keyframe (id = 270) with its covisibility keyframes, loop keyframes, and non-loop keyframes in the ground truth.

3.4.2. The Adaptive Threshold Method

As mentioned before, we believe that the threshold for filtering distance should be related to a keyframe's covisibility keyframe. To verify this idea, we set up the following experimental steps:

1. Obtain a certain number of keyframes near the current keyframe, for example, 10.

2. Calculate the distance metrics M between the current keyframe and the keyframes in neighbors, respectively. Then store all the distance matrices M into the set C.

$$C_k = \{M_{k,1}, M_{k,2}, M_{k,3}, \ldots, M_{k,10}, \ldots M_{k,j}\} \quad (1)$$

3. To compare precision and recall, we set 20 gradually increasing thresholds.

$$T_{1\sim10} = C_{min} \times m \times \frac{C_{avg} - C_{min}}{10}$$
$$T_{11\sim20} = C_{avg} \times m \times \frac{C_{max} - C_{avg}}{10}$$
$$m = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} \quad (2)$$

According to Fig. 7, we found that before the 12th threshold, as the threshold increases, the recall increases relatively large, and the amount of frames to be verified does not increase much. After the 12th threshold, as the threshold increases, the increase in the recall is relatively small, but the amount of frames to be verified increases a lot. Therefore, we believe that the 12th threshold is the most appropriate.

As shown in Fig. 8, it compares the performance of adaptive threshold and fixed threshold. The blue line represents the calculated adaptive threshold T of all keyframes with loops. The orange line represents the fixed threshold of 64000. From Fig. 5 and Fig. 8, we found that the optimal adaptive threshold is varies

greatly thus further proved the fixed threshold method to be sub-optimal.

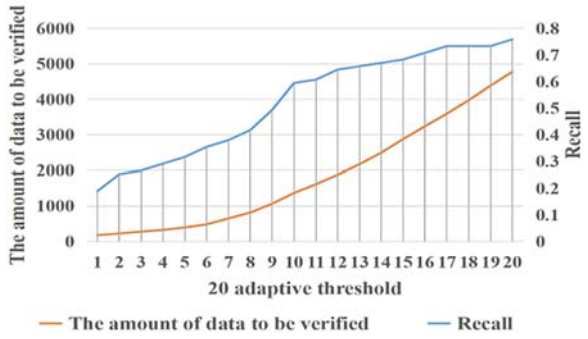


Fig. 7. Comparison between the amount of data to be verified and recall of 20 threshold.

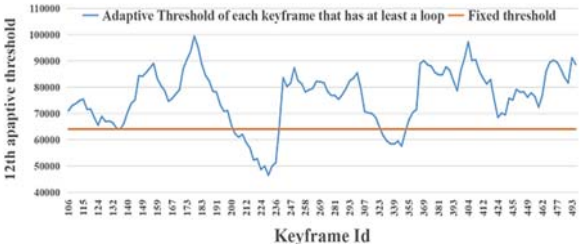


Fig. 8. The comparison of adaptive threshold and fixed threshold. The blue line represents the calculated adaptive threshold of the keyframes with loops. The orange line represents the fixed threshold.

3.4.3. Map Merging

Loop detection is in charge of detecting loop matched pairs between different sub-maps. After detecting a loop, a 4×4 matrix of the loop matched pairs are obtained by RANSAC iterations, like (3). It represents a similar transformation between the pose of the current keyframe and the pose of the detected loop keyframe. Using M_{RT} we can obtain the relationship of rotation and translation between the poses of these two keyframes.

$$M_{RT} = \begin{bmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{bmatrix} \quad (3)$$

Then, we can use the (3) to merge the sub-map corresponding to the two robots by a rigid transformation, like (4). It transforms every keyframe and map point in two local coordinate systems to a global coordinate system.

$$M_2 = M_{RT} \times M_1, \quad (4)$$

where M_1 and M_2 are the poses of the keyframes and the coordinates of the map points in the two coordinate systems respectively.

The global consistent map can be obtained by merging the sub-maps of the two robots. Once the sub-maps of the robots are merged into the global consistent map, the above RT matrix is applied to add the subsequent keyframes and map points of the robot into the global map. In this way, the global consistent map can be built continuously.

4. Evaluation

4.1. Datasets

We evaluated our system based on the Kashiwa dataset [23]. The dataset contains a total of 498 frames. Each frame was captured using a panoramic camera. This dataset contains loops in the same and different directions.

4.2. Loop Detection Results

A total of 79 loops matched pairs are in the ground truth in the Kashiwa dataset. The number of loops matched pairs that can be detected using the adaptive threshold, and the fixed threshold is shown in Table 1. The number of keyframes that can find loop keyframes with the fixed threshold is very small. More loops can be found using adaptive thresholds.

Table 1. Comparison of detection results between adaptive threshold and fixed threshold.

	Detected loop	Recall	Precision
Adaptive threshold	51	64.56 %	100 %
Fixed threshold	13	16.46 %	100 %

4.3. Cooperative SLAM System Performance

Fig. 9 shows the designed cooperative SLAM system based on two robots and merging two sub-maps in the same trajectory direction (a) and different trajectory direction (b). The red arrow indicates the direction of movement of robot1. The yellow arrow indicates the direction of movement of robot2. The purple circle represents the loop detected between robot1 and robot2. The central server uses the transmitted data to reconstruct the sub-map (1) and (2) corresponding to each robot. Then loop detection with adaptive threshold will detect loops in each sub-map. In the experiments, both of the loops in the same trajectory direction and different trajectory directions are detected. Then, the two corresponding sub-maps are merged, and the global consistent maps are constructed (a)-(3) and (b)-(3).

Table 3 shows the comparison of time-cost between single robot mapping and cooperative mapping. The mapping of a single robot used the frames from 1 to 280, and it consumed 161.915s. In the cooperative system, robot 1 used the frames from 1 to 140 and consumed 84.9719s. Robot 2 used the frames from 141

to 280 and consumed 76.751s. The central server received the data from robot-1 and robot-2 and reconstructed the map, run loop detection, and merged sub-maps. It consumed a total of 10.084s. It can be seen that the central server without feature extraction and tracking has very high computational efficiency. It is more efficient to use multiple robots to map in the same scene.

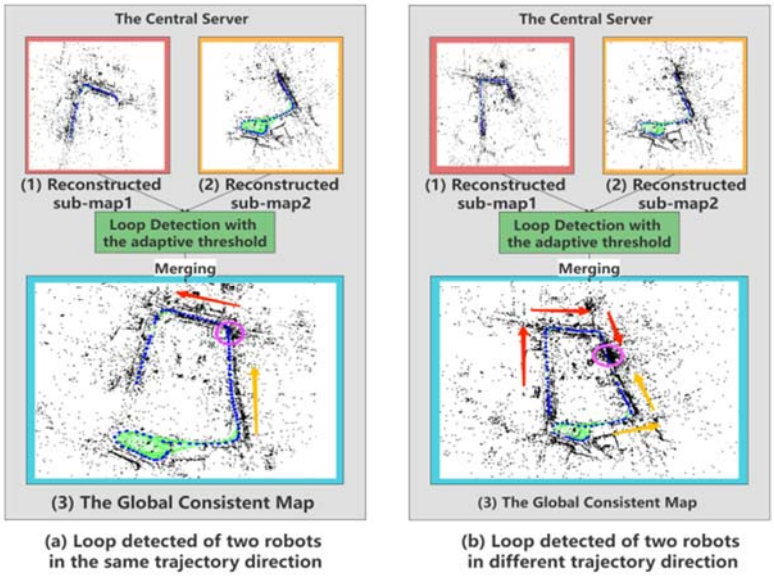


Fig. 9. Cooperative SLAM system based on two robots and the merging of two sub-maps in the same trajectory direction (a), and different trajectory directions (b).

In terms of accuracy, we use APE (Absolute Pose Error) as the metric. Table 2 shows the APE comparison between single robot and multi robots. We can see that compared with single robot mapping, cooperative mapping can reduce the cumulative error and increase the accuracy of mapping.

Table. 2. Various parameters of APE(m).

	rmse	**mean**	**std**	**sse**
Single Robot	0.83	0.69	0.46	45.23
Multi Robots	0.24	0.21	0.11	3.28

Table. 3. The comparison of time-cost between single robot mapping and cooperative mapping.

Frame range	1-280	1-140	141-280	1-140+ 141~280
Mapping	Single robot	Robot1	Robot2	The central server
Time-cost(s)	161.91	84.97	76.75	10.08

5. Conclusions

We propose a multi-robot cooperative system based on panoramic SLAM. Panoramic images are used to accurately detect whether independent robots meet each other in Omni-directions and then merge the map. We improved the method by keyframes, and map points were added. Our system can associate keyframes and map points with existing maps even when packets are lost in network communication. We also improved the loop detection part of panoramic SLAM. We propose three adaptive thresholds instead of fixed thresholds to filter candidate loop keyframes. Experiments show that the filtering effect of the adaptive threshold is better than that of the fixed threshold. The adaptive threshold can be used to find more loop matching pairs and correct loops. Both of them improve the efficiency and robustness of loop detection. Experiments show that our system can complete multi-robot transmission and map merging well in the same and different trajectory directions.

References

[1]. A. Geiger, P. Lenz, C. Stiller and R. Urtasun, Vision meets robotics: The KITTI dataset, *The Intl. J. of Robotics Research*, 2013.

[2]. Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29, 6, 2007, pp. 1052-1067.

[3]. E. Eade and T. Drummond, Scalable monocular SLAM, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* Vol. 1, New York City, USA, June 2006, pp. 469–476.

[4]. Klein, G. and Murray, D., Parallel tracking and mapping for small AR workspaces, in *Proceedings of*

the International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, 2007, pp. 225-234.

[5]. Raúl Mur-Artal, and Juan D. Tardós, ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras, *IEEE Transactions on Robotics*, 2016.

[6]. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, ORB: an efficient alternative to SIFT or SURF, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV),* Barcelona, Spain, November 2011, pp. 2564-2571.

[7]. D. Galvez-López and J. D. Tardós, Bags of binary words for fast ´ place recognition in image sequences, *IEEE Transactions on Robotics*, Vol. 28, No. 5, 2012, pp. 1188-1197.

[8]. Wang Y., Cai S., Li S.-J., Liu Y., Guo Y., Li T., Cheng M.-M., CubemapSLAM: A Piecewise-Pinhole Monocular Fisheye SLAM System, in *Proceedings of the 14th Asian Conference on Computer Vision*, Perth, Australia, December 2–6, 2018, pp. 34-49.

[9]. Urban, S., Hinz, S., MultiCol-SLAM-A modular real-time multi-camera SLAM system, 2016. arXiv:1610.07336.

[10]. Ji, S. et al., Panoramic SLAM from a multiple fisheye camera rig, *ISPRS Journal of Photogrammetry and Remote Sensing,* 159, 2019, pp. 169-183.

[11]. Leung K. Y. K., Cooperative Localization and Mapping in Sparsely-Communicating Robot Networks. Ph. D. Dissertation, *University of Toronto*, Toronto, ON, Canada, 2012.

[12]. Saeedi, Sajad, Trentini, et al., Multiple - Robot Simultaneous Localization and Mapping: A Review, *Journal of Field Robotics*, Volume 33, Issue 1, 2016, pp. 3-46.

[13]. Forster S. Lynen, L. Kneip and D. Scaramuzza, Collaborative monocular slam with multiple micro aerial vehicles, in *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013, 14002391.

[14]. P. Schmuck and M. Chli, Multi-UAV collaborative monocular slam, in *Proceedings of the IEEE Int. Conf. on Robotics and Automation*, 2017, 17058387.

[15]. Arroyo R. et al., Bidirectional Loop Closure Detection on Panoramas for Visual Navigation, in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2014, 14466969.

[16]. M. Cummins and P. Newman, Appearance-only SLAM at large scale with FAB-MAP 2.0, *Int. J. Robot. Res.,* Vol. 30, No. 9, 2011, pp. 1100-1123.

[17]. R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, NetVLAD: CNN architecture for weakly supervised place recognition, in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition,* 2016, pp. 5297-5307.

[18]. H. Jégou, M. Douze, C. Schmid, and P. Pérez, Aggregating local descriptors into a compact image representation, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.

[19]. N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, Place recognition with ConvNet land-marks: Viewpoint-robust, condition-robust, training-free, in *Proceedings of the Robotics: Science and Systems Conf.*, Rome, Italy, Jul. 2015.

[20]. Krizhevsky, I. Sutskever and G. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the Annual Conference on Neural Information Processing Systems,* 2012, pp. 1090-1105.

[21]. Garg S., N. Suen De Rhauf and M. Milford, Don't Look Back: Robustifying Place Categorization for Viewpoint- and Condition-Invariant Place Recognition, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA' 2018),* 2018.

[22]. Huang, A. S., E. Olson, and D. C. Moore, LCM: Lightweight Communications and Marshalling, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[23]. Kashiwa (http://gpcv.whu.edu.cn/data/panslam.html)

(009)

The Autonomous Pollination Drone

D. Hulens [1*], W. Van Ranst [1*], Y. Cao [2] and T. Goedemé [1]
[1] EAVISE, KU Leuven, Saint-Katelijne-Waver, Belgium
[2] Magics Technologies NV, Geel, Belgium
E-mail: dries.hulens@kuleuven.be, wiebe.vanranst@kuleuven.be, ying.cao@magics.tech, toon.goedeme@kuleuven.be

Summary: In this paper we will supplement the declining bee population by the development of a small drone that is able to autonomously pollinate flowers. The proposed solution uses a DJI Tello drone carrying a Maix Bit processing board capable of running all processing on-board. Additionally, the drone is equipped with a small color camera and a distance sensor to detect and approach the flower. We developed a two stage algorithm that is able to detect the flower, steer the drone towards the flower and makes the drone touch the flower to pollinate it.

Keywords: Pollination drone, Two-stage approach, Neural network.

1. Introduction

World food consumption in projected to increase in the coming decades [1], this together with a declining insect population [2] and the roles bees play in the pollination of crops [3], is a big concern for world food security. To have a successful crop, a farmer might rely on the surrounding ecosystem, artificially introduce a bee colony to pollinate their crops or rely on some other biological pollination method. However, due to collapse of surrounding ecosystems, regulatory difficulties such as protection from invasive species, or the circumstances of an artificial factory environment, interest in human made pollination methods has grown. Currently, pollination robots are already being used [4, 5].

These methods use mobile robot arms together with some kind of pollination brush to go from flower to flower, distributing pollen from plant to plant. Using such a robot arm however does have some limitations: the way a farm is laid out might for instance not easily allow a robot to pass between different crops, the type of produce might not lend itself to be touched by a robot arm, or the field might be too big or steep to build infrastructure for such a mobile ground robot. Nowadays, some crops are even planted above each other in vertical farms, which makes it impossible for a wheeled robot to pollinate [6].

All of this means that there is a lack and a certain need for a more generic and versatile pollination method.

In this paper we developed a mini drone equipped with a camera and on-board processing to autonomously pollinate flowers using a two-stage deep learning approach. As a preliminary use case, we focus on sunflowers.

The three main novelties of this paper are:
1. Embedded on-board real-time computer vision.
2. Trained with a partial synthesized dataset.
3. A two-stage flower approaching visual servoing (successfully demonstrated).

In the remainder of this paper, we investigate how we create such a pollination drone.

In section 2 we lay out the related work on robot pollination methods, autonomous drone technology and the computer vision techniques we use to steer the drone. In section 3 we go in depth into the inner workings of our pollination drone. In section 4 we explain the experiments we did to evaluate the effectiveness of our pollinator drone and discuss our promising results. We conclude in section 5.

2. Related Work

In this section we will first investigate existing artificial pollination solutions and then go into hardware platforms that can be used for on board computer vision based navigation. As mentioned earlier, world food consumption is projected to increase significantly in the coming decades [1]. To address this challenge, more efficient ways of producing food should be developed to feed this growing demand. A possible solution using technological means is to aid nature with the pollination of crops and plants. In recent research, robots were already developed to automatically pollinate crops and flowers by using a robotic arm on a base equipped with wheels [7, 5]. For the detection of the flower, they use Inception-V3 [8] together with color segmentation. This is a common approach for detecting flowers as seen in [9] where they combine a CNN and SVM to predict an even more accurate segmentation of the flowers. The downside of a robotic arm on a wheeled platform is its size and maneuverability. In [10] a drone was used to pollinate the flowers. This drone has multiple advantages over a wheeled robot, but was way to big (50cm x 50cm) to do the extremely precise job of pollinating flowers. In this work we developed a small drone (10cm x 10cm) with on-board processing power which can estimate the position and angle of the flower. Furthermore, this

drone can autonomously fly towards the flower and precisely pollinate the flower.

We investigated many hardware platforms that would fit our criteria for on board computation of Deep Learning based computer vision tasks. Our requirements for the platform were the following:

1. The platform should be light weight;
2. The platform should be low-power;
3. Easy to deploy and integrate with an existing drone;
4. Able to run state-of-the-art computer vision algorithms with a sufficiently high frame rate.

For the most part this rules out higher power mini-computer devices such as the NVIDIA Jetson series. The STM32F746 or other ARM Cortex M7 based platforms provide a deep learning library (CMSIS-NN) to run deep learning applications on the CPU architecture. The Ambiq Apollo3 Blue development board is able to run Tensorflow Lite models mostly targeted at voice and gesture recognition not computer vision. The Greenwave GAP8 is a RISC-V processor optimized to run deep learning models in a multi threaded way. The Kendryte K210 processor is also based on a RISC-V core, but in addition to that also contains a deep learning accelerator that is able to run YOLO [13] models at up to 20 fps. Because of its performance and relative light weight, we decided that the Kendryte processor best fits our use case.

From a machine learning point of view, we need models that are able to do two types of tasks: detection and classification (direct steering). These areas are a very active field of research, and many pipelines are available, each with their own trade-offs. Discussing all current techniques would lead us too far from the scope of this paper. Instead, we will discuss the techniques that are available on the chosen platform here. The Kendryte processor has support for the Mobilenet [11] family of backbones. This is a backbone especially optimized for low-power mobile devices, and is able to save a tremendous amount of computing power by using depthwise convolutions. Since the publication of the Mobilenetv1 paper, many additions to the network have been proposed as well [12]. However, due to the platform we choose, we are limited to Mobilenetv1. The classification head uses a simple cross entropy loss. In the detection stage we use a modified detection network, together with a Mobilenet to do fast inference. The detector makes bounding box predictions with only a single pass over the network, as opposed to region based detectors like Faster-RCNNN [14] and Masked-RCNN [15]. Due to their nature, generally speaking, single stage detectors are faster than region based ones.

The authors in [16] did a complete survey on the current state of object detection algorithms.

3. Autonomous Drone Pollination

In this section we go further into detail about the inner workings of our pollination drone. We create an artificial sunflower dataset containing images of real, as well as artificial and virtual sunflowers. We use the low power Kendryte K210 hardware platform which is able to run quantized Mobilenet based networks mounted on a commercially available DJI Tello Talent drone. A two stage flower approaching technique is developed to successfully touch and pollinate the flower. In the first stage, we detect the flower from a distance between 8 and 0.8 m using a convolutional detection network. In the second stage (0.8 - 0 m) we use an end-to-end network to predict the center of the flower in the camera feed. To guide the drone towards the center of the flower, a PID control loop is used to convert the location of the flower to steering commands for the drone. In this section we explain all of these components one by one.

3.1. Flower Dataset

To train the computer vision neural networks, we created an artificial dataset based on images of real sunflowers, artificial (synthetic) sunflowers and virtually rendered sunflowers. An important aspect we take into account is that for the drone to successfully pollinate a sunflower, it should approach the sunflower in a line perpendicular to the sunflower. In order to be able to do this, we need to estimate the angle of gaze orientation of the sunflower to enable us to correct for this when we are planning the drone's trajectory. This means that in each image the angle of the flower w.r.t. the camera should also be annotated.

For the virtual flowers, we simply render them at different frontal angles. For the real and synthetic sunflowers, we made a recorder device that is able to rotate the flower automatically while taking pictures of the flower. The generated images are then pasted on images from the Places dataset, more specific on images from Japanese gardens [17] to use as background. A total of 5660 annotated images were generated this way. An example of these images can be seen in Fig 1.

Fig. 1. Left: 3D generated sunflowers for end-to-end approach. Right: Generated images for direct visual servoing approach.

3.2. Hardware Platform

Since the precise steering the drone has to perform, processing of the video feed should happen on-board such that the delay between receiving images and correcting the drone's position is minimal. This implies the need for a small and light-weight processing board capable of performing real-time image processing. We

choose to use a Maix Bit development-board which contains a Kendryte K210 (RISC-V) processor. This board measures 53 mm x 25 mm and weights 25 g. Furthermore, this board is equipped with a camera, resulting in a minimal image receiving delay.

The camera feed is directly read by the Maix Bit on which we either run the end-to-end network or the detection network. The Maix Bit communicates with the Tello drone via its on-board ESP32 microcontroller. In addition to the camera, we also have a ToF depth sensor (VL53L1X) to measure the distance from the flower in the final approach. A detailed overview of our setup is shown in Fig. 2. Fig. 3 shows the drone carrying the Maix Bit.

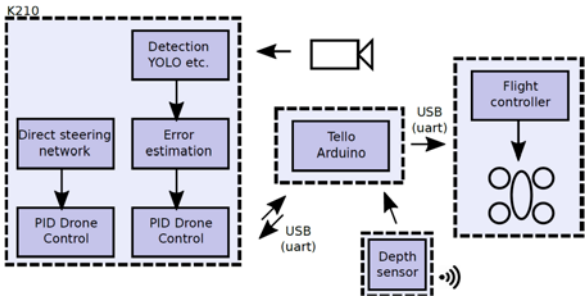


Fig. 2. Overview of our hardware setup.

Fig. 3. Our pollination drone carrying the Maix Bit.

3.3. Hybrid End-to-end and Detection Approach

As explained before, the vision part of our approach plan is divided into two stages. In the first stage we run a CNN for flower pose estimation: position, size (indication of the distance) and orientation. This model makes the drone fly towards a position close to the flower (approximately 80cm), directly facing it. In the second stage, we trained an image-based visual servoing network in an end-to-end approach such that it directly outputs steering commands towards the sunflower's position. This model is used for the final approach, the pollination touchdown.

3.3.1. Detection Stage

In the first stage (distance of 8m to 0.8m) we use a Mobilenet [11] detector trained on our own flower dataset. We altered the CNN architecture to additionally output the angle of the flower. This predicted angle is used to steer the drone towards a 0° angle w.r.t. the flower. The position of the flower is used to steer the drone such that the flower is always in the center of the drone's view. In this stage the size of the detected flower is used as a distance measurement. When the drone approaches the flower and the size of the flower exceeds 60 % of the frame-height, the flower becomes too big to be reliably detectable and we switch to the second visual servoing stage.

3.3.2. Direct Visual Servoing

In this stage we use an end-to-end network based again on a Mobilenetv1 architecture trained for classification that directly outputs steering commands (up, down, left, right or center). This network is trained on zoomed-in images of a flower. Here, we use a ToF distance sensor to measure the distance between the drone and the flower with high accuracy. The pollination stick in front of the flower measures 8cm, when the distance becomes smaller than 8cm we assume the stick touched the flower, and we fly backwards to restart the first stage and find a new random flower.

3.4. PID Control Loop

To transfer the position of the flower into steering commands for the drone, four different PID loops are used, one for each of the drone's axes. The goal of the PID loops is to center the flower in the drone's view and steer the drone such that the angle between the drone and the flower reaches zero degrees (ideal for pollination). To position the drone such that the flower is in the center of the screen, the Y-coordinate of the flower is used to move the drone up or down (Altitude) while the X-coordinate is used to steer the drone to the left or right (Roll). The angle of the flower is used to control the drone's rotation around its Z-axis (Yaw). When the Yaw is changed, consequently, the flower will also move to the left or right in the image, which will be again compensated by controlling the Roll. To approach the flower, the size of the detection is used to move the drone forwards or backwards (Pitch). When the distance becomes smaller than 0.8 meter, we switch to an end-to-end stage where the time of flight distance sensor is used to control the Pitch instead.

4. Experiments and Results

In the first experiment, we evaluated the average precision of the detection results of our first model. We reached a mAP of 0.36 on our test dataset (note that the different angle classes in our dataset make our dataset more challenging, when doing real-world test this accuracy proved sufficient).

The second experiment we conducted was evaluating the success-rate of pollinating a sunflower. We repeated this experiment 20 times were the drone

took off from a distance of 5m, flew to the flower, touched it and flew back. We reached a success-rate of 85 %. In the other 15 % of non-success, a false detection was the cause of failure.

The maximum frame rate of which images could be processed on the Maix Bit is approximately 30 fps for the direct steering model and 20 fps for the detection network. We also tested the minimum frame rate needed to smoothly steer the drone towards the flower since a slower frame rate results in less power consumption. The minimum framerate needed to smoothly approach a flower is measured at 12 fps. A demo of our working pollination drone can be seen in [18] and Fig. 4 shows a successful pollination sequence.

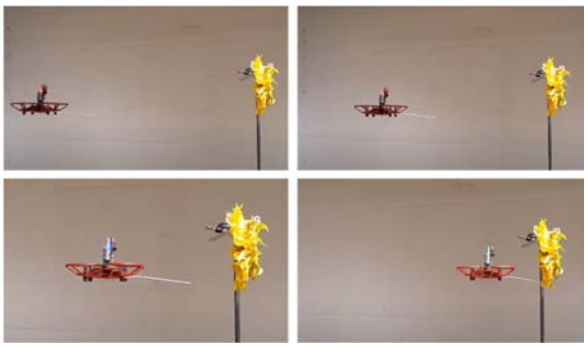


Fig. 4. Successful pollination sequence.

5. Conclusion

In this paper we developed a pollination drone that detects flowers, estimates the angle and flies towards the flower in a two stage approach. Different experiments were performed to evaluate all the parts of this approach, yielding promising results. We managed to run all processing on-board, resulting in a fully autonomous drone with a minimal delay between taking images and the control of the different degrees of freedom. Real world test demonstrate a success rate of 85 %.

Furthermore, in this work we demonstrate that, using currently available hardware, it is possible to autonomously fly a drone towards a flower and maneuver it in a way similar to what a bee would do to pollinate a flower. It remains to be seen how well methods like these would scale up to replace actual bee populations, and what other environmental impacts a dwindling insect population can have on the survival of humanity, and the surrounding ecosystems it relies on. A discussion should be had on how much we want to rely on technological solutions for our own food security, and what role humanity should play in interfering in existing ecosystems.

Acknowledgements

This research received funding from the Flemish government (Flanders AI program).

References

[1]. Michiel van Dijk et al., A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050, *Nature Food*, 2, 2021, pp. 494–501.
[2]. David L Wagner, Insect declines in the Anthropocene, *Annual Review of Entomology*, 65, 2020, pp. 457-480.
[3]. Shaden A. M. Khalifa et al., Overview of bee pollination and its economic value for crop production, *Insects*, 12, 8, 2021, p. 688.
[4]. Ting Yuan et al., An autonomous pollination robot for hormone treatment of tomato flower in greenhouse, in *Proceedings of the International Conference on Systems and Informatics (ICSAI),* 2016, pp. 108-113.
[5]. Nicholas Ohi et al., Design of an Autonomous Precision Pollination Robot, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 2018, pp. 7711 – 7718.
[6]. Fatemeh Kalantari et al., A review of vertical farming technology: A guide for implementation of building integrated agriculture in cities, *Advanced Engineering Forum*, 24, 2017, pp. 76-91.
[7]. Jared Strader et al., Flower Interaction Subsystem for a Precision Pollination Robot, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* 2019.
[8]. Christian Szegedy et al., Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2818 - 2826.
[9]. Philipe A. Dias, Amy Tabb, and Henry Medeiros., Apple flower detection using deep convolutional networks, *Computers in Industry*, Vol. 99, August 2018, pp. 17-28.
[10]. Peter C. Guglielmino et al., Autonomous Drone Pollination, *Worcester Polytechnic Institute,* 2021.
[11]. Andrew G. Howard et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017, arXiv:1704.04861.
[12]. Andrew Howard et al., Searching for mobilenetv3, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019.
[13]. Joseph Redmon and Ali Farhadi., YOLO9000: better, faster, stronger, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2017.
[14]. Shaoqing Ren et al., Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 39, 2016, pp. 1137-1149.
[15]. Kaiming He et al., Mask r-cnn, *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 2980 – 2988.
[16]. Li Liu et al., Deep learning for generic object detection: A survey, *International Journal of Computer Vision,* 128, 2020, pp. 261–318.
[17]. Bolei Zhou et al., Places: A 10 million image database for scene recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40, 6, 2018, pp. 1452 - 1464.
[18]. https://youtu.be/quX5HhVyR3g, *Youtube,* 2021.

(010)

"They got my keys!": On the Issue of Key Disclosure and Data Protection in Value Chains

A. Mosteiro-Sanchez [1,2], M. Barcelo [1], J. Astorga [2] and A. Urbieta [1]
[1] Ikerlan Technology Research Center (BRTA), Arrasate/Mondragón, Spain
[2] University of the Basque Country, Bilbao, Spain
E-mail: amosteiro@ikerlan.es

Summary: Value chains exchange massive volumes of data. Interconnection between companies makes them vulnerable to value chain attacks and data breaches. Information is encrypted before uploading it to shared databases, and its hash is stored in distributed data solutions like Distributed Ledger Technology (DLTs) to ensure data auditability. Key secrecy is vital for a security system based on an encryption algorithm. However, keeping encryption keys secret is vital to the system's security. Thus, accidental disclosures compromise the system security, as does the lack of a periodic key renovation system. This paper presents a solution based on Ciphertext-Policy Attribute-Based Encryption (CP-ABE). Our proposal enforces key renewal after each security event and guarantees it via time-based encryption, performed by Smart Contracts. We thereby ensure users refresh their keys periodically, attackers do not access new information once the key disclosure is discovered, and that original information does not need to be re-encrypted.

Keywords: Value chain, CP-ABE, DLT, Smart contract, Industry 4.0, Key disclosure.

1. Introduction

Smart manufacturing and Industry 4.0 imply massive information exchange across value chains. Information exchange increases operational efficiency and improves value chain management. Furthermore, it bolsters decision-making and company competitiveness.

However, interconnections between members of the value chain make the security of the entire chain reliant on the security of each member. A security breach in one member can have repercussions for the rest. Data breaches resulting from cyber-attacks on value chains [1] have a high monetary and reputational cost for the affected companies [2]. Thus, establishing a secure value chain infrastructure is essential for a thriving Industry 4.0 development.

Encryption schemes can guarantee industrial data confidentiality in value chains. However, encryption algorithms have a high computational cost, and it is essential to balance performance and security. This balance is relevant when data is generated and encrypted by IIoT devices with limited capabilities. Regarding security, encryption schemes require the generation of cryptographic keys. These keys must be renewed periodically to ensure the system's security, and thus a key revocation and renewal system must be implemented. Key disclosures compromise sensitive information since any attacker in possession of the disclosed key can read the encrypted data. Therefore, security systems designed for value chains must ensure that keys are renewed after a security event. Concerning performance, IIoT devices benefit from solutions that reduce the computational load while maintaining the security properties achieved with encryption. Value chain scenarios with multiple participants requiring the same information benefit from solutions that allow one-to-many encryption.

Several schemes have been proposed to achieve the aforementioned one-to-many encryption. Fiat and Naor outlined the first version of broadcast encryption [3]. This cipher has been further enhanced by Boneh *et al*. [4]. Following the same idea of one-to-many, Waters proposed Ciphertext-Policy Attribute-Based Encryption (CP-ABE) [5]. CP-ABE protects information according to access policies, which are defined by attributes. For instance, a piece of data can be encrypted according to an access policy such as *"CompanyA_Engineer OR External Auditor"*. An authority provides users with secret keys to decrypt the messages. This authority generates users' secret keys according to users' attributes in the system. This way, only users whose secret keys fulfil the access policy can read the encrypted data.

After successful data protection, information must be distributed to the different value chain members. For this purpose, distributed data sharing systems such as Distributed Ledger Technology (DLT) have proven their suitability for data sharing in supply chains [6]. The information is stored in a distributed database, while its hash is stored in a DLT. This allows all value chain members to keep track of data modifications. This way, data control and trust are shared among all supply chain members. Data sharing via DLTs can be combined with encryption systems, including one-to-many encryption systems. For this purpose, the information must be encrypted before being stored in the shared database, and the hash recorded in the DLT must match the encryption.

To recover the original information, partners need a valid decryption key. However, because of the particularities of value chains, new users coexist with old users, and access privileges may change. Similarly, secret keys being disclosed pose a threat to system confidentiality. There needs to exist a key management system resilient against key misuse—*i.e.,* keys

obtained by nefarious means, accidentally disclosed keys or legitimate keys that should have been renewed.

Traditional key revocation systems require lists identifying secret keys that cannot be used again. However, this implies that at some point, the encryptor and decryptor must negotiate new keys to continue the information exchange. In addition, data protected with the old keys have to be re-encrypted. Hence, the key revocation system designed for a value chain must provide a solution where IIoT devices do not need to renegotiate keys or re-encrypt information.

This paper proposes a time-based encryption system based on CP-ABE to protect the system from key misuse. Users must own secret keys generated after a particular timestamp that reflects the last detected security event. This way, users with invalid keys or old keys are automatically denied access to information and have to authenticate themselves again. The solution foregoes the revoked key lists used by traditional revocation systems. In addition, the IIoT devices do not need to reencrypt the information, nor do they have to renegotiate secret keys with value chain partners.

This paper has been structured as follows. Section 2 of this paper defines the identified issue with traditional key revocation methods. Section 3 describes the proposed solution, and Section 4 summarizes the conclusions of the proposal.

2. The Issue of Key Disclosure

Key disclosure poses a threat to information confidentiality. To ensure one-to-many encryption, value chain members protect the data at the source and upload the resulting ciphertexts (CTs) to a shared database. The hash of the CT is stored in the DLT so that every chain member can verify its integrity. The auditability provided by DLTs makes it possible to detect modifications to the information and trace which value chain partner has made them. However, regarding CT retrieval, anyone who has a key created during the system's lifetime can retrieve the original message. As shown in Fig. 1, if there are no key management systems, data consumers can use old and disclosed keys to access data.

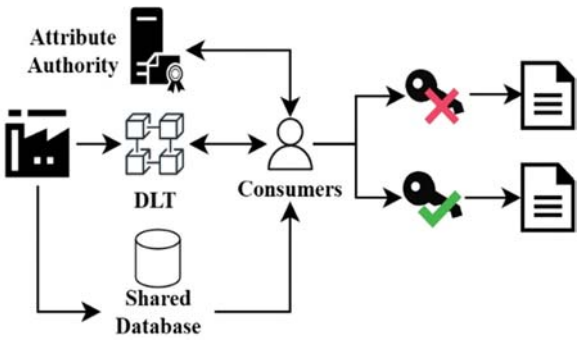


Fig. 1. With no key revocation, no longer valid keys (in red) can still obtain the information.

Encryption ensures information confidentiality but does not procure user identification. Therefore, even if the data breach is discovered, there is no mechanism to prevent attackers from reusing the stolen keys. Current key revocation systems rely on lists to record revoked keys. Each time a data consumer uses their key, it is compared to those in the list. Although this prevents the use of disclosed keys, it implies list management. In addition, the keys are linked to the CTs, which must be re-encrypted. A new symmetric key must also be generated and securely transmitted in symmetric encryption systems. In the case of asymmetric encryption algorithms, the encryptor and decryptor have to renegotiate the key pair.

Every key disclosure implies the risk of attackers using stolen keys to access the system. Therefore, there is a need for a revocation system compatible with one-to-many encryption systems and which does not have to re-encrypt every CT generated during the system lifetime. The system must also not renegotiate keys between the IIoT device that encrypted the information and the chain member that retrieved it. To this end, applying a new encryption layer whenever data is retrieved can limit key abuse. In this paper we propose to link this encryption to a security timestamp reflecting the last known security event. This guarantees that the system rejects expired or leaked keys without maintaining a revocation list. The solution, based on CP-ABE, is explained in the next section.

3. Proposed Solution

3.1. CP-ABE-based Time Encryption

The proposed solution implements an encryption system that protects information against key misuse. The solution ensures regular refreshing of partners' keys and prevents attackers from using stolen keys. It also exempts IIoT devices from data re-encryption and key renegotiation. To explain how this is achieved, first, we introduce the concept of security events.

With the term "security event" we define any situation that affects the system's security state—for example, a periodic key renewal or a security incident. An event would also be restoring the system after detecting a security breach. Our solution must account for different events and maintain one-to-many encryption. Thus, we base our solution on the discretionary data access provided by CP-ABE.

As explained, CP-ABE protects information according to access policies and guarantees that only partners whose attributes comply with the policy can read encrypted data. Thus, it achieves one-to-many encryption without the need to identify each user. Any CT to which we add a new CP-ABE encryption layer can obtain this property. Our previous work [7] explored the concept of encryption layers and can be applied for time-based encryption.

In our time-based CP-ABE solution, the first time value chain partners want to retrieve information from

the shared database, they request a secret key (*SK)* to an authority. The authority authenticates the partner, timestamps the request, and considers the timestamp (TK_{TMSP}) as the partner's attribute. Thus, the system generates each user's secret key based on the TK_{TMSP} of their key request. Previously, the system has generated another timestamp reflecting the last system restoration after a security event (E_{TMSP}). Using CP-ABE properties, the requested information can be encrypted at the time of the request according to a policy that requires a secret key more recent than the security event. This way, users with old keys or known disclosed keys cannot access the information. Instead, they have to identify themselves again and authenticate their identity to the authority to obtain a new key.

3.2. Message Exchange

This section discusses how the proposed system deploys time-based encryption by combining it with the DLTs mentioned in Section 2. The proposed time-based encryption is represented in Fig. 2. It shows the inclusion of the time encryption module, so the consumer interacts directly with it instead of with the database and the DLT.

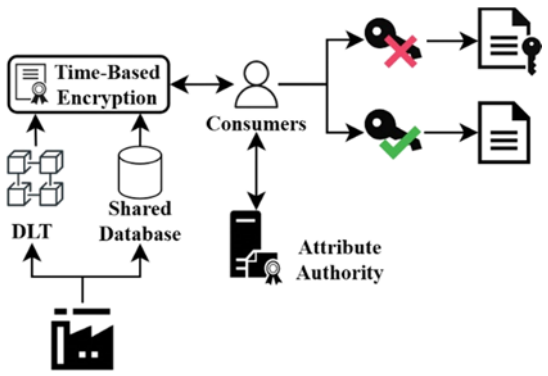


Fig. 2. Time encryption denies decryption to data consumers with no longer valid keys (in red).

The Attribute Authority (AA) generates some Public Parameters (PP) and a Master Secret Key (MSK) during system setup. The PPs are made available to all value chain members, while the MSK is known only to the authority. Value chain members store the CTs in a shared database, while the hash of the CT is stored in the DLT. Meanwhile, data consumers receive the secret keys (SKs) based on TK_{TMSP} from the AA. The authority uses the MSK to generate these SKs taking TK_{TMSP} as the attribute.

Consumers interact with a Smart Contract (SC) to retrieve the CTs. By its nature, SC can only work with the information stored in their associated DLT. Thus, for the SC to retrieve data stored outside the DLT, it needs the help of an oracle. Oracles [8] are services capable of providing information for smart contracts. They act as the layer between data outside the DLT and SCs. With the help of the oracle, the SC can retrieve the CT, the PPs, and the system restoration timestamp (E_{TMSP}). The SC uses E_{TMSP} to define an access policy such that $AP = (TK_{TMSP} > E_{TMSP})$. The SC uses this AP and the PPs to encrypt the CT requested by the consumer. The newly encrypted CT_{time} is then sent to the consumer. Consumers who do not comply with the AP defined by the SC cannot obtain the original information. The encryption and decryption processes are detailed below and are depicted in Fig. 3.

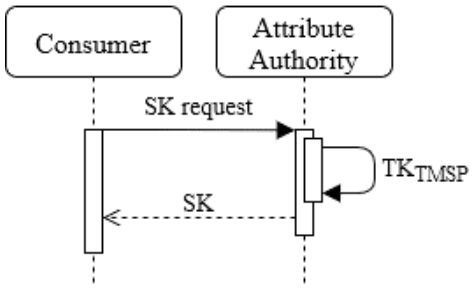


Fig. 1. Secret key request and generation.

1. Consumers request a secret key according to the AA.
2. The AA records the request timestamp, $TK_{TMSP.}$
3. The AA generates the secret key using TK_{TMSP} as the attribute.
4. The consumer receives the requested secret key.

After obtaining the SK, consumers interact with the SC to retrieve the CTs. In Fig. 2, we present how the Smart Contract performs the Time-Encryption before returning data to the consumer who requested it.

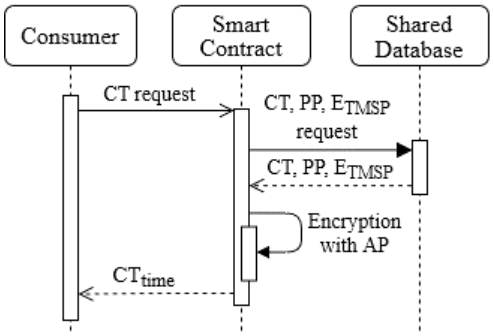


Fig. 2. Time Encryption.

1. Consumers request a CT from the Smart Contract.
2. The Smart Contract uses an oracle to retrieve the PP, the $E_{TMSP,}$ and the required CT from the shared database.
3. The SC encrypts the CT according to an access policy that requires a TK_{TMSP} newer than $E_{TMSP.}$

$$AP = TK_{TMSP} > E_{TMSP}$$
$$E_{CP-ABE}(CT, AP) \rightarrow CT_{time}$$

4. The Smart Contract sends CT_{time} back to the consumers.
5. Consumers decrypt CT_{time} using their key. If TK_{TMSP} fulfils the condition expressed in the *AP*, they access the information. Meanwhile, they recover ⊥ if the key is older than the last event.

Thus, consumers must request a new key from the AA whenever they recover ⊥. This strategy ensures that consumers keep their keys up to date, and the authentication system prevents attackers from obtaining valid keys. Furthermore, IIoT devices are exempt from further interactions beyond the original encryption, as they do not have to renegotiate keys or re-encrypt information. And since at no point is the original CT decrypted by anyone other than the consumer, data confidentiality is guaranteed during the whole process.

4. Conclusions

Large-scale information exchange in value chains increases the competitiveness of the involved companies. However, it also makes them vulnerable to attacks and data breaches. Distributed storage solutions combining databases and DLTs enable chain members to trace data tampering. This storage can be combined with one-to-many encryption solutions to reduce the risk of data breaches. However, users' secret key management and revocation remains a challenge that needs to be addressed in an efficient and scalable manner.

For the system to be secure, it is necessary to guarantee periodic key renovation and establish mechanisms preventing the use of compromised keys. For this purpose, this paper proposes the application of time-based encryption. The solution is based on CP-ABE and is achieved using a Smart Contract combined with oracles. For this purpose, the authorities provide consumers with CP-ABE decryption keys that match the key generation timestamp (TK_{TMSP}). Smart Contracts use the oracles to retrieve the CTs, public parameters and the last security event timestamp, E_{TMSP}. They use these parameters to encrypt de CTs according to access policy $AP = (TK_{TMSP} > E_{TMSP})$. This new ciphertext is sent to the consumers.

Since decryption requires a key generated after the security event, the system guarantees that only users with updated keys can access the information. As for the attackers, they need to steal a new key each time the system is restored. Since they must bypass the authentication system, data is protected against disclosed key misuse. Finally, IIoT devices are not overloaded since they do not have to re-encrypt information or renegotiate keys.

Acknowledgments

This work was financially supported by the European commission through ECSEL-JU 2018 program under the COMP4DRONES project (grant agreement Nº 826610), with national financing from France, Spain, Italy, Netherlands, Austria, Czech, Belgium, and Latvia. It was also partially supported by the *Ayudas Cervera para Centros Tecnológicos* grant of the Spanish Center for the Development of Industrial Technology (CDTI) under the project EGIDA (CER-20191012), and by the Basque Country Government under the ELKARTEK program, project TRUSTIND - Creating Trust in the Industrial Digital Transformation (KK-2020/00054).

References

[1]. ITRC, Q1 2021 Data Breach Analysis, 2021.
[2]. IBM, The 2020 Cost of a Data Breach, 2020.
[3]. A. Fiat and M. Naor, Broadcast Encryption, in *Proceedings of the 13th Annual International Cryptology Conference Advances in Cryptology - CRYPT0 '93*, California, USA, 1993, pp. 480-491.
[4]. D. Boneh, C. Gentry and B. Waters, Collusion Resistant Broadcast Encryption with Short Ciphertexts and Private Keys, in *Proceedings of the 25th Annual International Cryptology Conference Advances in Cryptology -- CRYPTO 2005*, Santa Barbara, California, USA, 2005, pp. 258-275.
[5]. B. Waters, Ciphertext-Policy Attribute-Based Encryption: An Expressive, Efficient, and Provably Secure Realization, *in in Proceedings of the 14th International Conference on Practice and Theory in Public Key Cryptography*, Taormina, Italy, 2011, pp. 53-70.
[6]. T. M. Fernández-Caramés, O. Blanco-Novoa, I. Froiz-Míguez and P. Fraga-Lamas, Towards an Autonomous Industry 4.0 Warehouse: A UAV and Blockchain-Based System for Inventory and Traceability Applications in Big Data-Driven Supply Chain Management, *Sensors*, Vol. 19, No. 10, 2019, p. 2394.
[7]. A. Mosteiro-Sanchez, M. Barcelo, J. Astorga and A. Urbieta, Multi-Layered CP-ABE Scheme for Flexible Policy Update in Industry 4.0, in *Proceedings of the 10th Mediterranean Conference on Embedded Computing (MECO'2021),* Budva, Montenegro, 2021, pp. 1-4.
[8]. C. Smith, Ethereum: Oracles, 3 January 2022. (https://ethereum.org/en/developers/docs/oracles/)

(011)

Virtual Commissioning of an Automotive Station for Door Assembly Operation

R. Balderas Hill, J. Lugo Calles, J. Tsague, T. Master and N. Lassabe
Capgemini Engineering, France
E-mail: {rafael.balderas-hill, jesus-hiram.lugo-calles, junior.tsague, tobiah.master, nicolas.lassabe}
@capgemini.com

Summary: Industry 4.0 has established new standards in manufacturing and quality processes, which are mainly based on the type and efficiency of the different components of an industrial system. Virtual commissioning (VC) is an approach that allows the user to propose a commissioning solution using computer models of industrial systems to speed up and improve the traditional processes. This paper presents a VC solution for an industrial automotive assembly line process using toolchains of standard and off-the-shelf software and hardware components, including physical PLC and HMI. The virtual assembly line environment was developed using digital mock-ups of industrial robots, smart conveyors, controllers and sensors. The toolchains share most of their components, virtual and physical, and can be classified as pure software or software and hardware applications. The connection between the elements of the toolchain is done using industrial protocols like Modbus TCP and APIs, ensuring complete data transfer. After several tests, the results have shown that by performing the simulated process with the VC platform, it is possible to verify automation hardware and software, with a representative digital robotic cell, thanks to the applied simulation technique.

Keywords: virtual commissioning, virtual engineering, automotive industry.

1. Introduction

Nowadays, simulation systems are recommended for engineering and decision-making tasks. [1] Simulation models from different fields are combined, using multi-domain simulation tools, to create a complete model of the system specifications and behaviors, considering the different physical interactions and communication protocols. [2, 3].

A digital factory, also known as Smart Factory or Cyber-Physical Production Systems, is the generic term for a network of digital models. It considers as well methods and tools integrated in the life cycle phases of a production system, characterized by their scalable and modular structure based on the idea of concurrent engineering and computer integrated manufacturing. [1, 2, 4-7] This approach eases the integration and replacement of production lines whilst being flexible to disruptions and failures. This type of simulation generates model-based copies, known as digital twins, that can be defined as virtual representations of a physical assets enabled through data for real-time prediction, optimization, monitoring, controlling, and improved decision making [8].

Virtual Commissioning (VC) focuses on the optimization and validation of automatic systems, increasing quality and efficiency in production engineering whilst decreasing required time, through using Smart Factories and commission them in a simulation environment. [4, 9] As mentioned in [1], VC can be considered as the quality gate of the mechatronic, robotics and automation engineering results, and the first step on the development process that ensures their interoperability.

Increasing the production of complex and/or customized products with short life-cycles, like in the automotive industry, requires a lot of engineering and planning effort.[1, 3, 10-12] VC is the best option for these scenarios, reducing ramp-up time, resulting in shorter product's time to market.

This technology allows to verify the functionality of systems, through the testing of off-line programs for specific devices, and avoid unexpected expenses due to inadequate component selection or damage during testing. [13] VC can be seen as a part of the modern approach to Product Lifecycle Management (PLM), being mostly used for product design inception through its manufacture. [14]

Research has been done reagind the standardization of VC to identify its level of complexity, detailing functions, with the main goal to ease its implementations and business model. [15] This standard categorizes VC into 5 levels, 1 to 5, according to the functionality and accuracy of the solutions, being level 5 the highest. This paper presents first the description of a level 5 VC Toolchain and the automotive use case; secondly, the simulation and results are studied; lastly, conclusions are discussed.

2. Virtual Commissioning Toolchain

As it was previously introduced, VC consists in modeling functional and 3D kinematized models to do Verification and Validation (V&V) of automation hardware and software. The advantage of doing such virtual engineering techniques, is that the V&V of automation solutions can be done offline by connecting the simulation platforms to the operational technologies, e.g., PLCs. Leading to not impact the production in the assembly lines, since the production process does not require to be interrupted. Depending

on the phases within the workflow from engineering until real commissioning, one might need to implement different simulation technologies. Because of the previous, we show an implementation of three simulation approaches permitting to validate requirements during different project phases. The targeted architecture can be shown in Fig.1. The modeling and simulation tools are based on:

- *ControlBuild*: For functional and behavioral modeling of the mechatronic components of the automotive assembly station
- *Delmia*: For modeling 3D kinematized components.

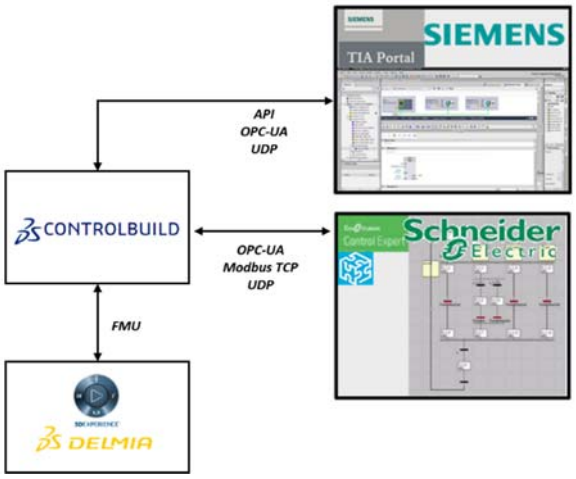


Fig. 1. Targeted Virtual Commissioning toolchains.

The operational technologies are based on Siemens and Schneider editors, with their appropriated programming environments, respectively *TIA Portal*, and *Control Expert*. To apply such global architecture into the automotive use case, in the next section we show the application of the simulation platform for a door assembly process of an automotive use case, extensively described in [10]. Additionally, the three types of simulation approaches, i.e., Model-in-the-Loop (MiL), Software-in-the-Loop (SiL), and Hardware-in-the-Loop (HiL), are shown, thus validating high and low-level automation and robotics requirements.

3. Simulation and Testing Specification

The automotive assembly cell consists of a smart conveyor (used to carry the car body through the cell), a couple of robots (LRo and RRo), buffers (LB and RB), positioning sensors (WA1 and WA2), safety gate sensors (LSG and RSG), and safety barrier sensors (SLBX). All of those components are arranged in a mirror manner, as shown in Fig. 2. Control and power electronic components, like PLC, drives, reles, robot controllers, industrial cabinets, etc., physical models are omitted in the 3D environment as their location in the real world is expected to be in the control components area of the shopfloor.

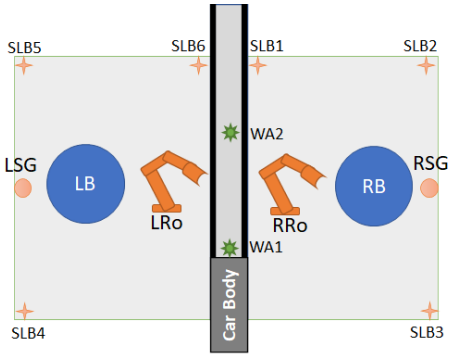


Fig. 2. Automotive cell 2-D layout.

A brief recall of the processes performed in this use case is shown in Fig. 3 and listed below:

- **Start cycle:** Absences of faults are acknowledged in the station, afterwards, a signal is sent to start the production process.
- **Production mode:** The production sequence is launched. Signals are sent in parallel to the different devices in the cell to begin their processes.
- **Switching car:** The conveyor brings the car body into the assembly area. The conveyor linear velocity and position are controlled throught the use of position sensors and a motor drive unit.
- **Check buffer:** Right and left buffers rotate to bring the body sides to the position where the two pick-and-place robots can perform the assembly. The angular positions of the buffers are measured by absolute encoders.
- **Side car positioning:** The robots assemble the body sides to the chassis. When the task is finished, the car continues along the conveyor to the following process, outside of the current robotic cell.
- **Safety flag system:** Safety sensors, emergency buttons, and circuit breaker systems are monitored during the whole process. This task possess the highest priority of the whole operation, whenever a single alarm and/or fault is detected, the process enters in safety mode and stops the production.

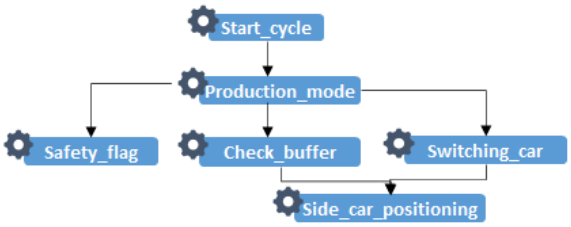


Fig. 3. Door assembly operation flowchart for a single cycle.

To V&V the automation process, we focused on the use of two validation environments: Siemens and Schneider automation software. According to Fig. 4, there are two toolchains permitting to perform both SiL, and HiL techniques. On the upper right of the toolchain in Fig. 4, the system to be commissioned is a Siemens virtual PLC 1515F-2 PN, and a virtual HMI KTP700F. These automation components are connected to the simulation platform via shared memory, permitting to exchange data bidirectionally, to V&V the programs and routines computed in the virtual PLC and HMI through SiL. On the bottom right of the Fig. 4, a physical HMI is connected to the simulation platform. In this case a virtual PLC Modicon 580 and a physical HMI Harmony STU855 from Schneider Electric are used. The communication between the simulation platform and the execution layer based on Schneider automation software is done via Modbus TCP. From the system setup, it has been accomplished the end-to-end communication and digital continuity from the simulation to the execution layer, permitting thus to improve the commissioning process for an automotive scenario.

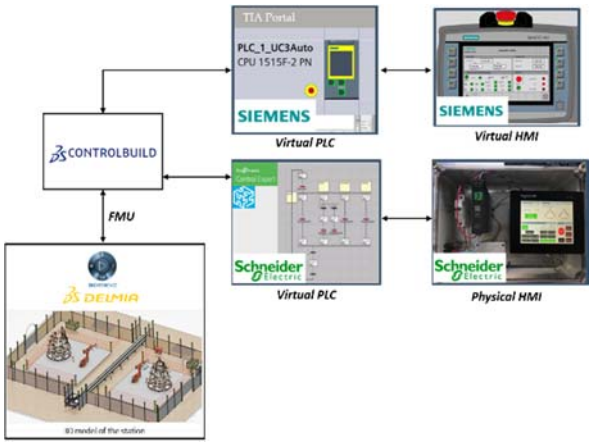


Fig. 4. System to be commissioned through MiL, SiL, and HiL.

In terms of network and infrastructure architecture, Fig. 5 shows a schematic of the communication protocols among the several components. The communication between the simulation PC and the physical HMI is done via a physical connection to the local network of both components establishing an exchange through Modbus TCP. The virtual PLC runs in the simulation PC, which is virtually connected to ControlBuild via Modbus TCP. Finally, the communication between ControlBuild, and Delmia, for the visualization of the 3D kinematized components is done via FMU.

Fig. 6 permits to visualize the mechanism of variable exchange performed between the PLC's (From TIA Portal and ControlExpert) and ControlBuild. It is worth noticing that the communication is done via a coupling driver respectively, using the PLCSimAdv API of Siemens, for TIA Portal, and Modbus TCP for ControlExpert. Additionally, in order to associate the variables between the PLC and ControlBuild, output and input mapping tables are created in order to properly match the variables, ensuring thus, a bidirectional communication. It is worth noticing that safety and control signals, as well as the information generated by simulated sensors in Delmia and Controlbuild provide the required data for closing the control loop with the PLCs (physical or simulated), meaning that the proper mapping of these variables along the toolchain is in high regard and data loss must be minimized.

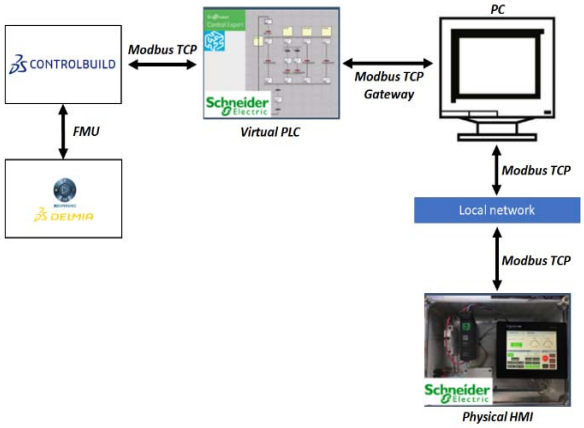


Fig. 5. Network setup for HiL with physical HMI.

4. Conclusions

VC implementations allow the user to develop, simulate, and test different industrial scenarios. These simulations can be as complex and detailed as the tests required, always providing enough information for the V&V process. The automotive use case studied in this research was tested with SiL and HiL toolchains, demonstrating the feasibility to simulate the operation with different control components without losing any of its features. It is worth noticing that the 3D simulation environment requires a computer with enough graphical and processing resources to perform a fluid simulation, being this the biggest bottleneck of the toolchain. Regarding the control devices and software, due to the fact that the process was well analyzed and interpreted, the logic behind the solution was implemented in an optimal manner, allowing the virtual PLCs to run it without any setback.

Acknowledgements

This work was conducted and funded with the support of the Research Tax Credit (CIR in French).

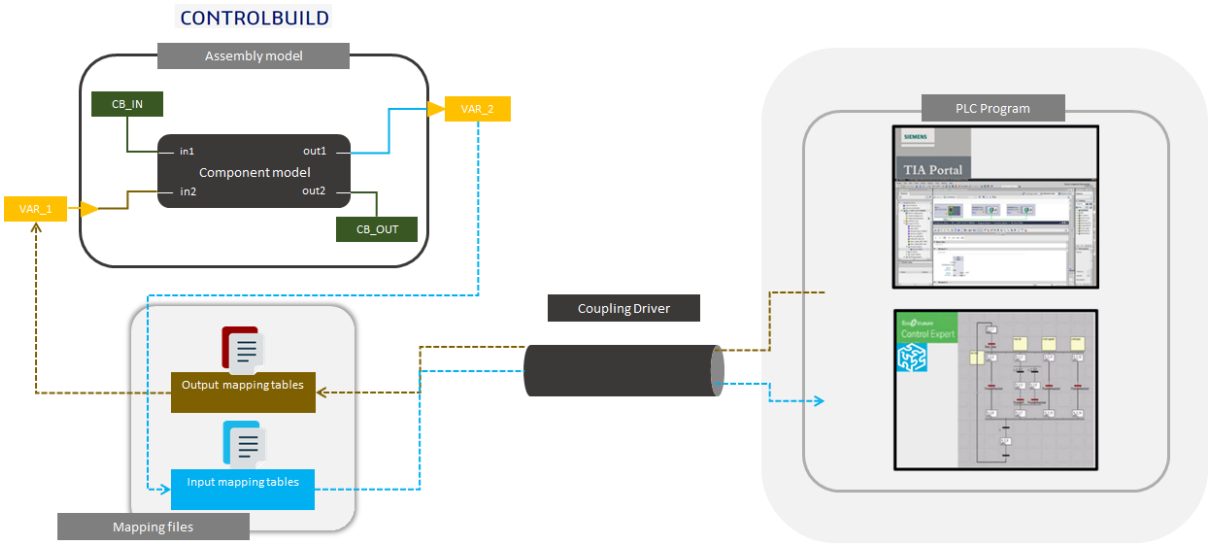


Fig. 6. Communication and mapping.

References

[1]. S. Weyer, T. Meyer, M Ohmer, D. Gorecky, D, Zühlke, Future Modeling and Simulation of CPS-Based Factories: An Example from Automotive Industry. *IFAC-PapersOnLine*, Vol. 49, Issue 31, 2016, pp. 97-102.

[2]. C. Scheifele, A. Verl, O. Riedel, Real-time co-simulation for the virtual commissioning of production systems, *Procedia CIRP*, Vol. 79, 2019, pp. 397-402.

[3]. S. Süß et al., Test methodology for virtual commissioning based on behaviour simulation of production systems, in *Proceedings of the 21st IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* 2016, pp. 1-9.

[4]. T. Lechler, E. Fischer, M. Metzner, A. Mayr, J. Franke, Virtual Commissioning – Scientific review and exploratory use cases in advanced production systems, *Procedia CIRP*, Vol. 81, 2019, pp. 1125-1130.

[5]. S. T. Mortensen, O. Madsen, A Virtual Commissioning Learning Platform, *Procedia Manufacturing*, Vol. 23, 2018, pp. 93-98.

[6]. T. Breckle, M. Kiesel, J. Kiefer, N. Beisheim, The evolving digital factory – new chances for a consistent information flow, in *Proceedings of the Conference on Intelligent Computation in Manufacturing Engineering*, Naples, Italy, 2018, p. 252.

[7]. D. Sobrino, R. Ružarovský, R. Holubek, K. Velíšek, Into the early steps of Virtual Commissioning in Tecnomatix Plant Simulation using S7-PLCSIM Advanced and STEP 7 TIA Portal, in *Proceedings of the Modern Technologies in Manufacturing (MTeM 2019) MATEC Web Conf.*, Vol. 299, 2019, 02005.

[8]. A. Rasheed, O. San, T. Kvamsdal, Digital twin: Values, challenges and enablers from a modeling perspective, *IEEE Access*, 8, 2020, pp. 21980-22012.

[9]. A. Fernández, M. A. Eguía and L. E. Echeverría, Virtual commissioning of a robotic cell: an educational case study, in *Proceedings of the 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* 2019, pp. 820-825.

[10]. A. Kampker, S. Wessel, N. Lutz, M. Reibetanz and M. Hehl, Virtual Commissioning for Scalable Production Systems in the Automotive Industry: Model for evaluating benefit and effort of virtual commissioning, in *Proceedings of the 9th International Conference on Industrial Technology and Management (ICITM),* 2020, pp. 107-111.

[11]. R. Balderas-Hill, J. Delbos, S. Trebosc, J. Tsague, G. Feroldi, J. Martin, T. Master, N. Lassabe, Improving interoperability of Virtual Commissioning toolchains by using OPC-UA-based technologies, in *Proceedings of the 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* September 2021, Vasteras, Sweden, pp. 01-07.

[12]. S. Makris, G. Michalos, G. Chryssolouris, Virtual Commissioning of an Assembly Cell with Cooperating Robots, *Advances in Decision Sciences*, Vol. 2012, 2012, Article ID 428060.

[13]. Ružarovský, Roman, Holubek, Radovan and D. Sobrino, Virtual Commissioning of a Robotic Cell Prior to its Implementation Into a Real Flexible Production System, *Research Papers Faculty of Materials Science and Technology Slovak University of Technology,* Vol. 26, no. 42, 2018, pp. 93-101.

[14]. A. Philippot, B. Riera, V. Kunreddy, S. Debernard. Advanced Tools for the Control Engineer in Industry 4.0, in *Proceedings of the IEEE International Conference on Industrial Cyber-Physical Systems (ICPS),* 2018, Saint Petersburg, Russia, pp. 555 – 560.

[15]. Albo, Anton & Falkman, Petter. A standardization approach to Virtual Commissioning strategies in complex production environments, *Procedia Manufacturing*. 51, 2020, pp. 1251-1258.

(012)

A Model Driven and Hardware Agnostic Approach of Virtual Commissioning

S. Marchand, H. Alhousseini, R. Bres, F. Dumas, M. Lachaise, L. Poulet de Grimouard and M. Stieglitz
Capgemini Engineering, ER&D, 350 Avenue JRG GAUTIER de la LAUZIERE,
13593 Aix-en-provence, FRANCE
Tel.: +33 4 84 93 44 76
E-mail: sylvain.marchand@capgemini.com

Summary: With the possibility to interconnect almost every systems, the Industry 4.0 has brought a lots of possibilities to improve lead-time and cost management while dealing with the design or the evolution of assembly lines. New tools and concepts are developed, such as Virtual Commisionning or digital twins, to improve the process to upgrade an existing assembly line or create a new one. These tools allows to design, simulate and validate changes without implying costly and time-consuming tests on real infrastructures. This paper describe an approach, and its implementation, to facilitate the use of Virtual Commisionning tools. It uses standard OPC UA communication and information modeling to implement a model driven and hardware agnostic approach to be the boundary between the virtual and the physical world. The solution can reuse models, such as AutomationML, from the design phase, can be used with simulation tool to develop and test the implementation in a virtual environment, and be used in the target hardware infrastructure for integration tests.

Keywords: Cyber-physical Systems, Model transformation, Design principles in Industry 4.0, Smart Factories, Smart Manufacturing and Technologies, Digital Production and Virtual Engineering.

1. Introduction

Smart factories and Industry 4.0 have enhanced a lot connectivity in factories. The goal is not only to connect production systems together, but to share all their knowledge with other systems. Using the data, failure can be anticipated with predictive maintenance, changes can be tested on virtual environments, etc. All those new possibilities lead to substantial decrease in risks and costs compared to traditional development, testing, and commissioning phase.

From an engineering point of view, the data collected is a goldmine. It can be used to reproduce real-world problems in digital twin and continuously improve processes.

However, some issues are yet to be solved to fully take advantage of all these data. First of all industrial communication protocols are numerous, among which PROFIBUS, MODBUS, Fieldbus, HART, ASi, LonWorks, DeviceNet, ControlNet, CAN Bus, and Industrial Ethernet are the most famous according to [1, 2]. Moreover, those protocols are often purely industrial protocols, not designed to go out of the factory, which causes side effects, such as poor security handling.

Then modeling has to be handled in a unified way, so that every system speaks the same language. Once again, the variety of languages to model assets is wide, with for example UML, SysML, AutomationML [3], OWL, etc.

These facts tend to be solved by the maturation of OPC UA (Open Platform Communication Unified Architecture). Indeed, OPC UA standard defines in its specification [4] 4 axes:

- The message model to interact between applications.
- The communication model to transfer the data between end-points.
- The information model to represent structure, behaviour and semantics.
- The conformance model to guarantee interoperability between systems.

These four pillars are the base of the OPC UA interoperability standard. They define the standard way to exchange data between applications, and the way to universally model it.

Moreover, OPC UA has been proven to be a secure protocol [5] and it is more and more integrated as a standard way to acquire data by software editors.

1.1. Data Acquisition Using OPC UA Aggregation Server

As stated by its specification [4], OPC UA is designed to be a platform independent, reliable and secure protocol which can be used from plant-floor PLC (Programmable Logic Controller) to enterprise servers. Its conformance units and profiles organization [6], allows it to be scalable from an embedded sensor to an enterprise server. For these reasons, OPC UA has been widely adopted by device vendors (PLC manufacturers, robots manufacturer, etc.) as well as industrial software platforms: SCADA (Supervisory Control And Data Acquisition), MES (Manufacturing Execution System), ERP (Enterprise Resource Planning), etc. OPC UA is now part of the RAMI4.0 (Reference Architecture Model for Industry

4.0), as the preferred protocol for its communication layer [7].

Modern factories can now use OPC UA as communication protocols at every level of the process. Recent PLC are OPC UA servers, gateways exist to transform classic industrial communication protocol to OPC UA, and software can be fed with data using their OPC UA clients. For those reasons, data acquisition using OPC UA looks like a foregone conclusion.

OPC UA aggregation servers have been studied for a few years now, and several sample exist to serve as a starting point for data collection. As described in [8] several challenges must be faced to successfully setup an OPC UA aggregation server. The steps identified for a generic approach of aggregation are:

1. Type aggregation: consists in getting all the types from an underlying server. To do so, the Types node of the underlying server is browsed recursively, and the result is merged with already known types in the aggregation server.
2. Instance aggregation: consists in getting all instances in the underlying server. Nodes aggregated from source servers are then evaluated with specified rules to be put in the right place in the aggregation server's address space.
3. Service mapping: In the proposed architecture, the service calls on the aggregation server are forwarded to the source server on which the node concerned by the call is located.

This generic approach lays the foundations of the construction of an aggregation server, and is used in implementations, such as the sample application of the OPC foundation. However, it has some drawbacks:

- In a vertical multi-level aggregation, the propagation of the service call from the client to the base source server might be long and cause latency or timeout.
- There is no local version of nodes. Every request must be forwarded to the underlying source server. If a group of nodes must be read by several clients, it would generate a lot of unnecessary Read requests to the underlying server. That can be a problem on some industrial networks, which should not be saturated.

To deal with these problems, a second kind of aggregation (Data Warehouse mode) has been implemented, based on the OPC foundation's sample aggregation server. Basic steps defined previously still remain true, the server starts by the discovery of types contained in underlying servers, then it gets instances, and services are mapped with underlying server's services when it is necessary. Nodes are copied in the aggregation server's address space, and subscriptions to corresponding nodes are created. Doing so, some services, such as Read do not need to be mapped, and propagated. For some services like method call, there is no other way than propagating the call to obtain the desired behaviour.

Another limit to be tackled with is the redundancy of namespaces. Some namespaces are defined by vendors, and duplicated in each of their devices. For instance, if a shop floor is equipped with 3 robots from the same manufacturer, they will all define the type representing this robot. The first approach would be to differentiate types depending on their source server. Using this approach, there is no risk to mix different versions of the same type, or 2 types named the same. The second one would be to merge common types, because the type of the 3 robots is the same, it should be defined and handled so. Both approaches are possible based on configuration of the aggregator.

1.2. Data Modeling

IEC62541 defines OPC UA's meta model [4, 9]. This meta model defines the way to represent data using an object oriented model, see Fig. 1. In OPC UA model, objects are composed of objects, variables and methods. The model describes types as well as instances.

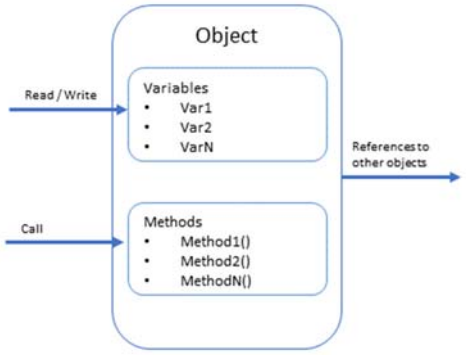


Fig. 1. OPC UA Object model.

The set of existing types and instances are represented in a server as a graph, with nodes (types and instances) bound by references, see Fig. 2. Nodes have attributes, that define what they are, and references that link them to other nodes.

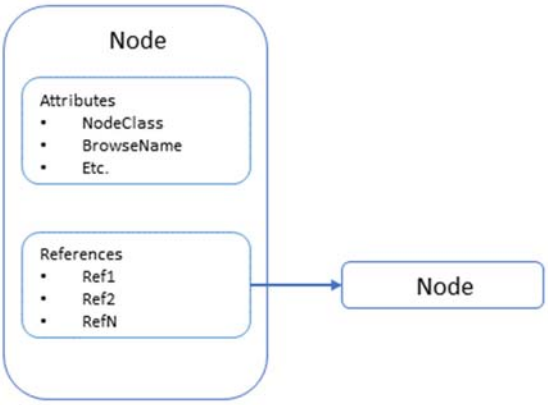


Fig. 2. OPC UA Node model.

An OPC UA server can load models using 2 standard methods. The first consists in loading a standard Nodeset2 file that describe the address space of a server. The second is based on code. UA server

solutions generally comes with a code generator that converts models to usable code (specific to the server). Both ways are complementary and have pros and cons.

Modeling is an important point of OPC UA specification and its meta model allows users to model any of their industrial assets. On top of the meta model defined by the OPC UA specification, some working groups have formed to standardize the use of OPC UA modeling in their fields. Those initiative gave birth to companion specifications. The list of these specifications [10] includes several information models aiming at the description of machines such as OPC UA for Machinery, OPC UA for Robotics, OPC UA for MachineTools, etc. Some of these models are dedicated to modeling language adaptations such as OPC UA for I4 Asset Administration Shell, ISA-95 Common Object Model, OPC UA for AutomationML, etc.

AutomationML (AML) is a data exchange standard [11] used for production system engineering. It gives a way to exchange data between applications to model the system and represents it using an object oriented paradigm. The use of AML format by OPC UA server has been standardized [12, 13]. It defines a way to transform AML file in OPC UA model loadable by an OPC UA server.

However, the wide variety of existing models, whether a standard translation exist or not, are only models and do not define how the data and the model are bond.

2. Architecture

To try to go a step further and solve issues mentionned in Part 1, using a generic approach, the I4SSM (Industry 4.0 Semantic and Security Management) has been developed. The main piece of the solution consists in an OPC UA server, which implements an architecture with 3 functions, described in Fig. 3:

- Aggregation.
- OPC UA operation server.
- Semantic layer.

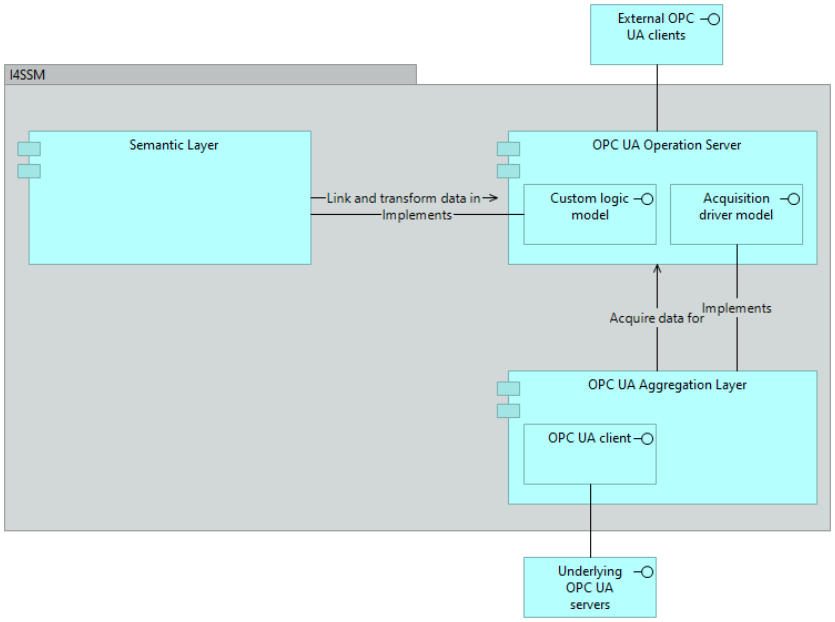


Fig. 3. I4SSM architecture.

2.1. OPC UA Operation Server

The operation server is a standard OPC UA server which exposes and manages interfaces to configure the aggregation module and the semantic layer.

The operation server handles OPC UA client calls and handles it with the help of other components if needed.

Acquisition of data is managed by the aggregation module, which discovers and updates nodes from underlying servers. These nodes are exposed to clients using the operation server. The link between both modules is done via an internal event and messages mechanics.

The operation server also manages the loading of models. They can be loaded using 2 formats: Nodeset2 or compiled libraries.

The main difference with state of the art servers [14] relies in the fact that an interface and reflection mechanism is used here so that a compiled model can be used without having to rebuild the whole server.

2.2. Aggregation Layer

The aggregation layer of I4SSM is designed to be generic. The implementation of the aggregation module is based on the conclusions drawn from

Part 1.1. This article focuses on the Data warehouse mode as default implementation, which creates a local equivalent to nodes discovered from underlying data sources. Moreover, the aggregation module exposes an interface to develop specific acquisition drivers, in case a non OPC UA connection to data source is required.

As for previous implementation [8], the aggregation server is an OPC UA client, and uses Browse standard service to discover its underlying source server address space. It begins with the discovery of types and continues with instances. Discovered types are added to aggregation server's types, i.e. they are added in the aggregation's server Type tree. Then the server creates a copy of underlying servers instances, and organizes it in the aggregator's address space by putting them in folders corresponding to their source server, as shown in Fig. 4.

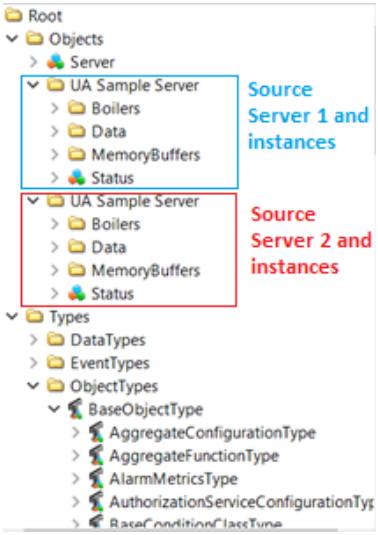


Fig. 4. Address space organization.

The discovered nodes are assigned a new and unique NodeId in the aggregation server. A mapping of the local nodes with remote nodes (in the underlying server) is maintained, so that the remote node can be accessed for service calls that cannot be fulfilled locally. Subscriptions are created by the client embedded in the aggregator to update local nodes. The Sampling interval of monitored items and Publish interval of the subscription are configurable.

Once the discovery is completed, Read, Browse and subscription can be made on the aggregator address space by a client, without interacting with the underlying server. However, services like Write and Call must be forwarded to the underlying server to be effective. Indeed the writing of a parameter in an underlying server must be done in the real server to be effective.

Because plants and shopfloor are dealing with potentially important data volume, scaling up is a key point to manage Data Model as well as upload and download information to the shopfloor. To address this particular aspect, a collaboration with Intel has enabled the use of different hardware configuration enhanced with software or hardware add-ons and accelerators. Some common actions have been implemented on a test environment facilitating test execution. This environment enabled the possibility to test the solution's scaling, both horizontally (more aggregated servers) and vertically (more layers of aggregation servers) to ensure the possibility to use the solution with an architecture as close as possible from real-life applications. The test of the functionalities of I4SSM both separately and in parallel has given feedbacks on the implementation of the solution. For instance Table 1 shows improvement of performances between campaign 1 and 2, concerning the number of servers that can be aggregated. As a result, the feedback on the first 2 tests campaigns has reinforced the need to include in the project life cycle the stress tests to anticipate performances issues and find bottlenecks in order to facilitate code optimization by design.

Table 1. Success rate of aggregation tests for N source server with 5k variables each.

5K VARIABLES		
Source servers	**1st campaign**	**2nd campaign**
1	100 %	100 %
2	100 %	100 %
3	60 %	100 %
4	0 %	100 %
5	0 %	100 %
10	0 %	100 %
15	0 %	100 %

2.3. Semantic Layer

The Semantic layer plays a major role in I4SSM architecture. It implements the link between models loaded using the operation server and data acquired using the aggregation layer.

To execute this task, in a generic, and as standard as possible way, the OPC UA meta model is used to define what a SemanticLink is. The address space of the server is created with a SemanticLinkSet object which lists all semantic links instanciated in the server and has methods to add or remove links, as shown in Fig. 5.

The model for semantic link definition is extensible, to do so a base abstract type defines common parameters of all semantic link types. This base type is defined in Table 2. A semantic link makes a link between at least 2 variables, a source and a target. The relation can be one to one, or many to many. A link can define simple links, such as "this target variable represents this data", it can also do some operation on the fly to specify that "this target variable is the same as this source data with a different unit". It can also be used to perform basic arithmetic operations, such as addition, power of ten, or affine function. Moreover, an interface is available for a user to extend the set of links, by adding its own types of

link and implementing its associated behavior. This mechanism is implemented so that a simple configuration of the server is needed to add a type of link to the server, no rebuild of the server is required.

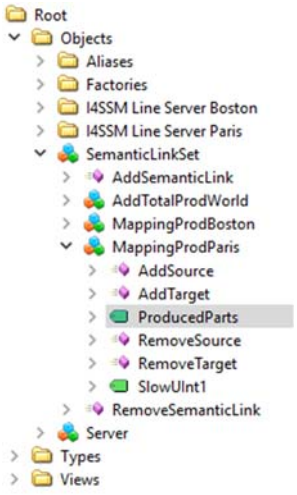


Fig. 5. SemanticLinkSet in I4SSM address space.

Table 2. SemanticLinkType definition.

SEMANTICLINKTYPE ATTRIBUTES			
Attribute		**Value**	
BrowseName		SemanticLinkType	
IsAbstract		True	
SEMANTICLINKTYPE REFERENCES			
References	**Node Class**	**BrowseName**	**Modelling Rule**
SubtypeOf	ObjectType	BaseObjectType	
HasComponent	Method	AddSource	Optional
HasComponent	Method	AddTarget	Optional
HasComponent	Method	RemoveSource	Optional
HasComponent	Method	RemoveTarget	Optional

The modeling and implementation using OPC UA meta model offers 2 major advantages: a standard way to design and model semantic links, and the possibility to use a simple OPC UA client to add, remove and modify semantic links in the server. Indeed, any OPC UA client that can perform method call is able to manage semantic links. That way, scripting can be used to massify the creation of links, to fit better with real use case in which thousands of semantic links are needed to map data from a production line to its high level model.

2.4. Fault Recovery and Persistency

Fault recovery and persistency are major concerns for I4SSM. First of all, after having done the mapping of thousands of nodes, a user does expect that it will still be up and working in case of a server restart. However, this matter is not trivial to solve because of the automatic discovery of underlying servers address spaces.

Nodes discovered in underlying sources are mapped with a local node, which NodeId is uniquely generated in the server. The mapping table and the NodeIds must be persisted, so that it can be retrieved after a restart or a crash. It is important, so that existing semantic links can find their sources and targets back. The same is also true for external client applications that often store NodeIds and base their internal mappings on that.

To be able to set the server in its prior shutdown state, the following items of interest have been identified:

- Nodes added with the AddNode service.
- Nodes updated with the Write service.
- References added with the AddReferences service.
- User roles.
- Sematic links.
- Aggregated nodes' NodeId mappings.

These elements are identified as key elements that need to be persisted. Nodes added to the server must be recreated, as well as the references. Updates on nodes must be persisted to be able to retrieve parameters as they were before the stop of the server. For example a parameter that defines an alarm threshold should not be rewritten by the user when the server restarts. The same idea is also true with user roles configuration.

3. Use case: Virtual Commissioning and Automotive Cell

3.1. Automotive Cell Components and Process

The Virtual Commissioning simulation platform, described in [15], is composed of two software allowing to perform functional models of mechatronic components of the station (ControlBuild), and a 3D visualization environment for the kinematized components (3DX Delmia). Additionally, these two simulation tools are connected via OPC UA to the I4SSM server introduced previously. Fig. 6 shows an overview of the proposed toolchain.

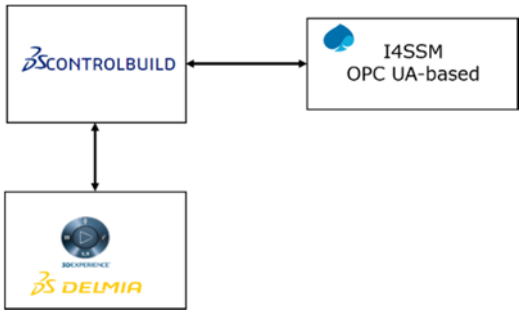


Fig. 6. Toolchain.

The automotive cell consists in several components allowing to emulate an assembly process for the body sides of a car. The assembly station is composed of mechatronic components, such as robots, sensors and safety components, as shown in Fig. 7, and listed as follows:

- Car body
- Car conveyor
- Rotational buffers
- 6-Degree-of-Freedom robots
- Presence sensors
- Safety components, such as safety gates and light barriers.

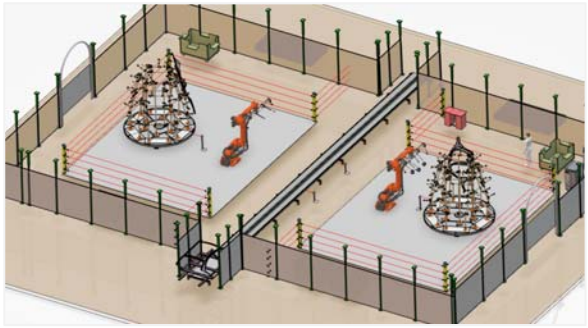


Fig. 7. Automotive station.

Additionally, the variables governing the full assembly process are described by various datatypes, thus allowing to show the cyber-model translation capabilities in what follows. The various types data translated into an OPC UA model contains therefore information related to sensors, robots positions and velocities, allowing the station to operate adequately.

3.2. Model Translation and Mapping

The model loaded in I4SSM is translated from a format exported by ControlBuild to a Nodeset2 format. Doing so, ControlBuild can be interfaced with the OPC UA server using its OPC UA client and connect its internal variables to variables in I4SSM.

ControlBuild can then be used as a simulation and validation tool without having to bother about underlying implementation. That is to say that the simulation can be run with I4SSM only, with virtual Programmable Logic Controller (PLC), or using real PLC. Since ControlBuild is mapped with the loaded model, nothing has to be done at its level to switch from a full virtual test to a test implying real hardware.

If PLCs (whether virtual or physical) are used, I4SSM aggregation layer copies their address space. That way, PLCs can be from different vendors, implying potentially different address spaces, without consequence for ControlBuild. Using built-in SimpleMappingSemanticLinkType offered by I4SSM, the model suiting ControlBuild can be linked with data discovered automatically from underlying sources.

This use case demonstrates the possibility to create a digital twin of an automotive assembly line, using OPC UA and taking advantage of a unified model. In this scenario, Delmia is responsible for the 3D visualization of the line, ControlBuild manages its simulation. I4SSM is responsible for implementing the hardware abstraction layer, by the exposition of a model that the simulation tool can understand, and by linking this model to the real data. With that abstraction, simulation can be performed and visualized in 3D, moreover tests can be done using virtual or real PLCs, without changing the configuration of the simulator. Using this, engineers can virtually simulate a change of parameters, validate a new PLC program using virtual PLC and even test a modification on the physical environment.

4. Conclusions

This paper presents a way to deal with the boundary between virtual and physical world and illustrates it in a practical use case. The I4SSM tool set presented is implemented and running following the ideas described in the article. It implements the basic OPC UA server aggregation needed to get data, it adds on top of this a set of objects to makes it possible to link acquired data with models, in a generic server, without need to rebuild it when the model evolves. It also offers several possible ways to load models, using AML to Nodeset2 converter, custom converter, or loading libraries built from models and tools made available on GitHub by the OPC Foundation [16]. Additionally to the classical acquisition server use case, I4SSM can be used in the virtual world, or as a boundary between virtual and physical environment, as proven with the automotive cell use case. However, I4SSM can still be improved and completed. Next steps would be to use it OPC UA Historical Access (HA) functionalities to replay some failure cases in collaboration with a simulation tool, performances could be improved again to be able to deal with even bigger amount of data, and the implementation of the PubSub OPC UA specification would open new possibilities, such as edge to cloud communication and aggregation.

Acknowledgements

This work was conducted and funded with the support of the French tax system, thanks to the Research Tax Credit (CIR in French).

We would like to thank INTEL, and especially Jan VAN OFFEREN for the support, the test platform and help on performance improvement. We also thank Jacques MEZHRAHID, Nicolas NGUYEN, Eric OURSEL and Sylvain PLAZANET for their support and expertise. Finally, we would like to thank Rafael BALDERAS HILL, the Virtual Commisionning team, and N. LASSABE for their expertise and their support on the redaction of this paper.

References

[1]. P. Bondeson, S. Liss, Roadmap of virtual commissioning inertia, Thesis, *KTH Royal Institute of Technology*, Stockholm, Sweden, 2018.

[2]. I. Gonzalez, A. J. Calderon, A. Mejias, J. M. Andujar, Novel Networked Remote Laboratory Architecture for Open Connectivity based on PLC-OPC-LabVIEW-EJS Integration. Application in Remote Fuzzy Control and Sensors Data Acquisition, *Sensors,* 16, 11, 2016, p. 1822.

[3]. Lüder A., Hundt L., Keibel A., Description of manufacturing processes using AutomationML, in *Proceedings of the 15th IEEE Conference on Emerging Technologies and Factory Automation (ETFA),* 2010, pp. 1-8.

[4]. OPC UA - Part 1: Overview and Concepts Release 1.04 Specification*,* 2017.

[5]. Federal Office for Information Security (BSI), OPC UA Security Analysis, 2017.

[6]. OPC UA - Part 7: Profiles Release 1.04 Specification*,* 2017.

[7]. Platform Industry 4.0, RAMI4.0 – a reference framework for digitalisation, 2018.

[8]. Johansson Markus, Vyatkin Valeriy, Aggregating OPC UA Server for Generic Information Integration, 2017.

[9]. OPC UA - Part 3: Address Space Model Release 1.04 Specification*,* 2017.

[10]. OPC Foundation website (https://opcfoundation.org)

[11]. Rainer Drath, et al., AutomationML in a Nutshell, in Handbuch Industrie 4.0, Vogel-Heuser B., Bauernhansl T., ten Hompel M. (Eds.), *Springer Reference Technik. Springer Vieweg*, Berlin, Heidelberg, Bd. 2, pp. 213-258.

[12]. DIN SPEC 16592, Combining OPC Unified Architecture and Automation Markup Language, December 2016.

[13]. Robert Henßen, Miriam Schleipen, Interoperability between OPC UA and AutomationML, *Procedia CIRP,* Vol. 25, 2014, pp. 297-304.

[14]. Stefan Profanter, Benedict Simlinger, From modelling to execution – OPC UA Information Model Tutorial (https://opcua.rocks/from-modelling-to-execution-opc-ua-information-model-tutorial/), 2020.

[15]. R. Balderas-Hill, J. Delbos, S. Trebosc, J. Tsague, G. Feroldi, J. Martin, T. Master, N. Lassabe, Improving interoperability of Virtual Commissioning toolchains by using OPC-UA-based technologies, in *Proceedings of the 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* September 2021, Vasteras, Sweden, pp. 01-07.

[16]. OPC foundation GitHub, https://github.com/OPCFoundation

(013)

Development of an AI Maturity Model for Small and Medium-sized Enterprises

B. Schmidgal, M. Kujath, S. Kolomiichuk, M. Rentzsch and S. Häberer
Fraunhofer Institute for Factory Operation and Automation IFF,
Sandtorstr. 22, 39106 Magdeburg, Germany
Tel.: + 49391 4090140
E-mail: boris.schmidgal@iff.fraunhofer.de

Summary: The relevance of Artificial Intelligence (AI) for manufacturing and service operations in the industrial sector is growing rapidly. Various economic actors, including small and medium-sized enterprises (SME) that are the backbone of the European economy, are already well aware of specific use cases of AI in the industrial context and demand effective AI solutions to their challenges. However, existing approaches and models for identifying the appropriate AI use case for efficiency improvement potentials can cover the growing market demand to a limited extent only. Uncertainties about the actual capabilities, a lack of a strategic focus, missing success metrics, poor data quality, or false expectations of AI systems can all lead to failure of AI projects. To counter this, an AI maturity assessment can assist companies to overcome challenges on the way to a sustainable planning, deployment and adaptation of AI in industrial applications. This paper examines the need for a holistic method and presents the Fraunhofer IFF AI Maturity Model – a method based on the needs of SME in the manufacturing industry.

Keywords: Artificial intelligence, AI-supported business model innovation, Use case-specific assessment, Maturity level, Feasibility analysis, Decision support.

1. Introduction

Artificial intelligence (AI) is regarded as an important technology that is indispensable for preserving Germany's economic performance. The German government has recognized the relevance of this key technology as well as its potential for additional economic growth, and has developed a framework for action to promote and fully exploit this potential. Essential framework conditions have been merged into a national "AI Strategy", which is designed as a learning strategy that needs to be continuously readjusted jointly by politics, science, industry and civil society (The German Federal Government 2020). In particular, the course is to be set for the expansion of competencies and expertise, the promotion of young researchers and knowledge transfer, as well as for modern research infrastructures and international networking. Thus, AI is expected to not only change the way people work, but also the individuals' everyday life and society in general.

Ministries, economic actors and major German associations agree: A successful digital transformation is of existential importance for prosperity, growth and innovative strength. By leveraging large data sets (Big Data), businesses are increasingly striving to create new value for their clientele and develop new, innovative business models (German Engineering Association 2020). AI-driven business model innovation already extends to various industrial application fields, such as warehousing and sorting by autonomous vehicles, analytics for product development, or quality control by means of AI-driven image recognition in the field of machine and plant manufacturing. Learning algorithms are already used in large and multi-national companies and can infer the wear and tear of machines from certain data such as pressure, temperature, acoustic noise, power consumption or vibrations. Business-relevant patterns and key figures can be derived from this and used as decision support for technicians.

2. Background and Motivation

While certain AI applications, such as interaction with customers (chatbots) and automated data collection and evaluation have become almost routine in the industrial context, German SMEs are reticent yet. A survey on the use of AI conducted by Bitkom e.V. in 2020 – the industry association of the German information and telecommunications sector – shows that 9 percent of German industrial companies with 100 to 199 employees use AI in their company in the context of Industry 4.0, whereas 11 percent of German industrial companies with 200 to 499 employees use AI (Bitkom 2020). Consequently, one can estimate an adaptation rate of AI in the industrial context in German SMEs of around 10 percent.

However, the same survey with same respondents revealed that the benefits of AI in the context of Industrie 4.0 are widely known and companies recognize economic potentials. According to the study, 43 percent cited that AI can provide advantages for predictive maintenance as the most important benefit of AI in the context of Industry 4.0, followed by productivity enhancement, and optimization of production and manufacturing processes with 41 and 39 percent accordingly. Hence, even though the advantages and use cases are mostly known, AI-based solutions are still barely used.

As one of the objectives of application-oriented research, it is necessary to identify the gap between the state of knowledge and state of implementation of AI and its economic possibilities within the SME landscape. This gap is due to existing barriers to use AI. A survey of AI experts by The Scientific Institute for Infrastructure and Communication Services revealed that a lack of expertise and skilled workers is a strong (36 %) or very strong (64 %) obstacle to the implementation of AI solutions in SME (Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste 2019). This reflects the extent of the general shortage of skilled workers in SME: IT specialists in particular are rare on the labor market. SME are often unable to keep up with the salaries paid by large companies.

Inadequate availability of data in sufficient quantity and quality is another major obstacle. Due to the smaller size of the company, the possibilities for data collection and data utilization are more limited than in large companies. Many of the SME are not sufficiently digitized to ensure an adequate collection of production data and machine data, among other things, by means of software and sensor technology. Since the presence of suitable datasets along the value chain is a key prerequisite for training AI systems, successful deployment of AI is directly dependent on the company's level of digitization. Data security concerns, lack of acceptance among employees, and limited financial resources are further obstacles that were also identified and analyzed by AI experts and trainers.

3. The AI Maturity Model

Based on these findings, concrete needs were defined, which are to be covered by a methodical approach. The goal is to derive a standardized, holistic procedure for identifying relevant use cases for AI deployment and providing a decision template.

Existing AI applications mostly focus on increasing efficiency. However, AI offers SME greater potential through service or product innovations, in which existing or easily captured customer data or process data combined with AI methods form the basis for the development of new value propositions. If these are integrated into corresponding revenue mechanics and value chains, innovative, AI-based business models can be developed.

However, to realize such business models, in addition to domain knowledge – the knowledge of how the industry works, what customers want, and the established processes – methodological knowledge in the context of AI is required above all. Therefore, it is necessary to identify the status quo of the resource endowment and requirements typical for the company size class. Therefore, the Fraunhofer IFF developed a third-party assessment of the current maturity level in order to derive necessary capacities and resources for empowering employees and exploiting potential by means of AI. The assessment was designed with the claim to be objective, holistic and realistic. The maturity model as a component of a use case analysis is predominantly aimed at companies in the manufacturing sector.

First of all, it is important to understand why the maturity model approach was preferred for making companies capable of acting in the field of AI. On the one hand, maturity models can be used to describe the change in the object under consideration (assessment) in order to derive recommendations for action based on this, which are aimed at achieving the next higher level. In addition to performance evaluation and improvement, maturity models are also used for internal and external comparisons. They provide a benchmark for best-in-class companies or the entire industry (Bruin et al. 2005). In the corporate context, maturity models can be helpful tools, especially for the responsible managers, to be able to classify the current development status of the company in relation to the past and the future. Maturity models focus the assessment on a specific area, but can be applied to a broad field (Häberer et al. 2017). Figure 1 depicts the Fraunhofer IFF AI Maturity Model that represents an incremental development trajectory required to achieve a networked and fully automated production and logistics environment with the help of AI use.

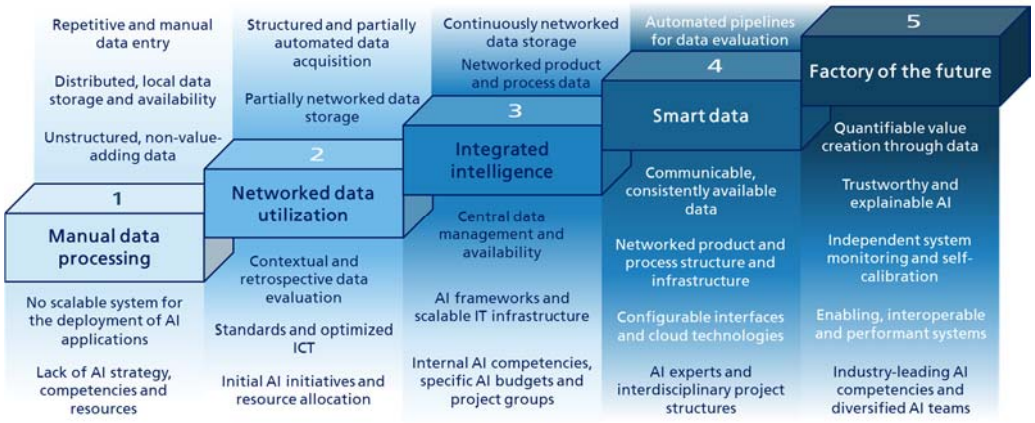


Fig. 1. The AI Maturity Model – five stages to AI maturity and their respective characteristics.

The model consists of five stages of AI integration:

- Stage 1: Manual data processing
- Stage 2: Networked data utilization
- Stage 3: Integrated intelligence
- Stage 4: Smart data
- Stage 5: Factory of the future

The following stage characteristics were established to make a clear differentiation between the individual maturity levels so that the differences to the next higher or next lower stage are distinct:

1. **Manual data processing:** Recurring and manual data entry; Distributed, local data storage and availability; Unstructured, non-value-adding data; Lack of a scalable system for deploying AI applications; Lack of AI strategy, skills, and resources in-house.

Example: A paper-based, non-automated data capture and paper-based process control system in screw manufacturing is present. In production planning and control, data evaluation is dominated by humans and data availability is only given in individual processes.

2. **Networked data utilization:** Structured and partially automated data acquisition; partially networked data storage; context-related and retrospective data evaluation; standards and optimized ICT; embryonic AI initiatives and resource allocation.

Example: A building materials manufacturer considers its data to be relevant to the business because data is evaluated on a rules-based approach and used to support employees in their decision-making. The team leader in the construction department can evaluate his shift planning in a partially automated and algorithm-supported manner and store it in a knowledge database, which the management can access.

3. **Integrated Intelligence:** Continuously networked data storage; Networked product and process data; Central data management and availability; AI frameworks and scalable IT infrastructure; Internal AI competencies, specific AI budgets and project groups.

Example: Data scientists at a turbine manufacturing company use data mining analysis techniques to evaluate and interpret extensive and complex data sets for error detection in production series. As part of a dedicated AI strategy, a data warehouse has already been built as an interdepartmental data storage and analytics platform. Analysts in development, production, quality assurance and warranty processing have access to this data warehouse to perform ad hoc analyses and create standard reports supported by machine learning algorithms. After particularly complex calculations, the company can now forecast for each part and each product variant the frequency with which certain defects occur as a rule.

4. **Smart data:** Automated pipelines for data evaluation; Communicable, consistently available data; Networked product and process structure and infrastructure; Configurable interfaces and cloud technologies; AI experts and interdisciplinary project structures.

Example: Commercial vehicle manufacturer collects process data at every step of the manufacturing process. Data pre-processing and initial analysis takes place on the edge. Through standardised data exchange interfaces, information is exchanged between the shop floor and central PPS system. The collected data is then processed in the cloud. AI is used to forecast completion times, delays and possible disruptive events in production. The algorithms suggest the optimal production planning and exchange the data with upstream and downstream processes (supply chain management, transport of finished products). Thus, the control of these processes is also data-based on the actual state of production and existing production state forecasts for defined forecast horizons.

5. **Factory of the future:** Quantifiable value creation through data; Trustworthy and explainable AI; Independent system monitoring and self-calibration; Enabling, interoperable and performant systems; Industry-leading AI competencies and diversified AI teams.

Example: High-tech electronic chip manufacturer has fully automated production lines. Production planning is carried out by AI algorithms depending on order situation, deadline chain, material availability. Quality control is carried out inline using current production data, evaluation of images of the products and semi-finished products by AI-based systems. Quality deviations are automatically compensated for by the production control system. Human intervention is only required in cases that are not clearly defined and in exceptional circumstances. The supporting processes (logistics, infrastructure) are also data-driven and depend directly on the main process (production). Through the external data exchange, production capacities as well as external service providers and suppliers are controlled across all locations.

Based on the assumption that one level of AI integration is characteristic for each stage of development, the individual AI levels of a use case, with increasing maturity, merge into one another, but are not mutually exclusive. Therefore, the distinction between AI levels for the respective use case levels should be regarded as intersecting.

The main purpose is that an organization should not only discover at which level its areas of application are, but in particular what kind of solutions are goal-oriented in order to get to the next higher level. This gives the organization an indication of what it needs to achieve if it wants to reach the next higher level of AI maturity. Thus, it should not be the goal of an SME to achieve stage 5 but rather the next stage of development.

The maturity model does not fulfil the claim of evaluating the entire company in terms of a uniform AI maturity, substantiated by a single numeral degree. Instead, it aims to evaluate individual use cases within

the company, representing a business process or function that describes a sequence of individual activities that are carried out successively to achieve a business or operational result with a discernible value (IBM 2017). As process-related components of an organization, use cases are seen as complex workflows that are intended to contribute to business success. For this reason, the objective is to identify the economically most prospective business use cases for ensuring a successful and structured introduction of AI within a company.

4. The Fraunhofer AI CheckUp

The maturity model is one component of an overall service that aims to make manufacturing companies ready for the use of artificial intelligence – the so-called AI CheckUp. The AI CheckUp is a method developed for the analysis of processes to identify suitable use cases for AI-based solutions within SME. The AI CheckUp arose from the development need of a modular structure to cover different scientific and economic requirements for AI in an industrial context:

- Evaluation of the existing database (quality, availability, etc.);
- Identification of improvements in the existing data management and utilization (problem monitoring);
- A credible evaluation system of use cases and realistic references of AI solutions implemented to date;
- Manufacturer-independent orientation and consulting;
- Third-party assessment of the AI and data management competencies available in the company through a holistic multi-method approach (interviews, workshops, analysis of the database, etc.);
- Estimation of quantitative benefits;
- Checklist with minimum requirements for a project;
- Identify new research needs.

The approach is to compare the as-is situation with the requirements that an AI-based solution requires, identify existing data sources, and design the to-be data workflows for the AI-based solution. Eventually, an identification of required resources (professional, technological, methodological) and well defined success criteria is to be established.

The first step is to raise awareness and educate the employees of a company around the topic of AI and clarify the potentials and restrictions of AI implementation. There is also a focus on the business goals and strategy in the sense of concretizing what a company wants to achieve independently of AI. Experts in the field impart fundamental knowledge about AI and provide answers to questions such as:

- Why should I get involved with AI?
- How do I find out whether AI is the right technology for me?
- What areas of application are there for AI in my company and where can AI specifically be helpful?
- Which AI tools exist and which of them can fit my business use cases?
- What data do I need at all and how do I need it to train my AI?

Secondly, operational use cases are to be identified and prioritized. On-site inspections (plant visits) and personal interviews form the basis for a structured approach and solution development. The goal is to successfully carry out a company-specific use case determination for various alternative actions. For this purpose, a questionnaire based on the fields of action of AI in the industrial context was developed to enable a systematic evaluation of AI use cases. Since the maturity model is adapted for the requirements, goals and challenges of each individual use case to be assessed, an assessment is always use case-specific and benchmarking of use cases does not appear to be purposeful. Therefore, use case-specific and general fields of action were developed that are relevant to cover a company's needs for actions in a holistic way. The AI CheckUp focuses on the following seven fields of action of AI in the industrial context:

1. Quality of the collected data
2. Handling / Processing of data
3. Infrastructure and access to technologies and tools
4. AI strategy and leadership
5. Employees and AI acceptance
6. AI evaluation and impact assessment
7. Data sovereignty and collaboration

Figure 2 depicts the seven fields of action that are to be inquired and assessed by the experts. Each field of action has one or several central questions, intended to guide the experts' data collection, but do not limit the field of action to these key questions alone. Based on the corresponding answers from the interviews with management, skilled workers and specialists, as well as the information gathered through observations and process analyses during the plant tours, each of the seven fields of action can be evaluated and classified according to the level characteristics of the maturity model at level 1 to 5 for the respective use case.

Hence, the AI maturity level is determined. The application areas are classified objectively in terms of the requirements and prerequisites for AI. The aim is to realistically assess the potential of AI to achieve these goals. An effort-benefit assessment of the measures is carried out so that problem definitions can be derived and aligned with the corporate strategy. The claim is a realistic assessment of where the identified and prioritized use cases of the company stand in relation to AI and which AI-supported solutions make economic sense. Based on the maturity degree of the respective use case, necessary development steps can be derived and prioritized, taking into account an integrated risk assessment (such as dependencies on partners and suppliers).

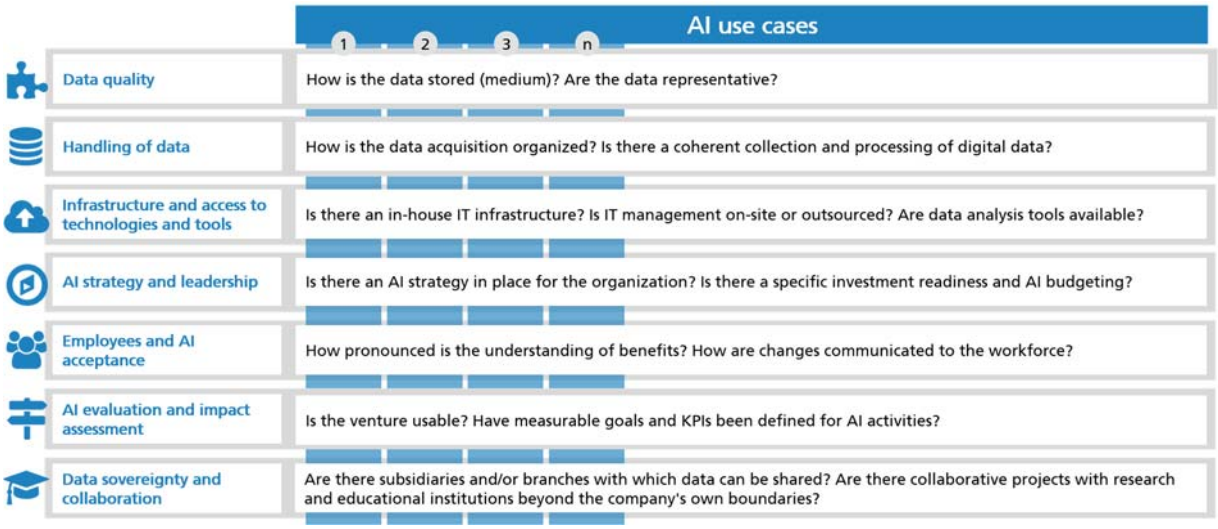


Fig. 2. Seven fields of action of AI and their key questions for information acquisition.

What follows is an AI use case development. Based on the information intake and the determined maturity degree, efficiency improvement potentials can be evaluated and individual measures for increasing the AI maturity of the respective use case can be derived. The measures can be of technical, organizational or methodical nature. Together with the employees, the experts derive operational use cases with potential for increasing efficiency and reducing costs, and match the problems with possible solutions (problem-solution-fit).

Ultimately, a decision template is developed and provided. For the purpose of decision support of the SME in possible investment projects, the Fraunhofer IFF developed a management tool for systematically developing identified use cases further, visualizing its elements, and serving the decision makers as a decision support template – the so-called AI Use Case Canvas. Its development is based on the Business Model Canvas by Alexander Osterwalder which is an established "strategic management tool to describe how an organization creates, delivers and captures value" (Osterwalder 2021). Comparable approaches with similar structuring principles for business process modeling of AI applications are available on the market, such as OwnML on the website www.ownml.co (OwnML 2022). The Business Model Canvas arranges the business model elements in nine basic building blocks and establishes a uniform structure, with the aim to show the logic of how a company intends to make money. Similarly, the AI Use Case Canvas was developed to describe business use cases in a structured way. Decision makers can discuss, evaluate and develop the building blocks, and discover correlations between them. This can serve as a template for decision support for the further development of existing and future use cases. The tool can be integrated into the company's strategic planning and provides an overview of a company's business use case within its external environment (market forces, industry forces, key trends, and macroeconomic forces) (Osterwalder 2010). Approaches for optimizing the economic mechanics of a specific business use case can thus be determined and developed. As a result, new product and service ideas as well as efforts towards process innovation can be initiated. The AI Use Case Canvas consists of 14 elements and aims to holistically visualize a specific use case in the company:

1. **Value proposition:** It describes the essence of the use value that the use case provides to the user.
2. **Internal customer segments:** Involved, affected or interested internal parties and organizations, which exert influence on the use case and influence each other (in particular own employees, shareholders and service providers).
3. **External customer segments:** Target groups outside of the company to whom the use case is intended to add value (in particular customers, the general public, authorities and lawmakers).
4. **Business model:** The as-is logic of how a company generates added value and secures earnings. The revenue mechanism are in the foreground here.
5. **Data situation / data quality:** The data situation / data quality describes how well data are suited to fulfil their purpose for the use case.
6. **Technology and infrastructure:** The set of tools, technologies (software and hardware), infrastructure, and methods used to provide the use case with the required resources and to achieve (create) the value proposition.
7. **Domain and method knowledge:** Existing domain knowledge of how the industry works, customer needs, and established processes, as well as methodological knowledge in the context of AI, data management and business models.
8. **Key partners:** The partnerships that are created through networking with individuals and organizations that can contribute to the business success of the use case.
9. **Key resources:** The key assets for the successful implementation of the use case, or the

resources available for the use case with which operations can be executed effectively.

10. **Key activities:** Processes that are essential for the success of the business model or use case. Key activities describe which activities are of particular importance for the success of the use case.
11. **Risks**: Risks associated with the AI use case project, possible negative outcomes, possibility of loss and failure. The risk-based assessment of AI describes the extent and quality of addressing system criticality as an expression of the damage potential of an algorithmic system.
12. **Cost structure:** The main expenses and costs associated with this use case (in particular fixed costs such as salaries, wages, rent, and infrastructure expenses, variable costs (proportional to the volume of goods or services produced), economies of scale, and economies of scope).
13. **Revenue streams:** Revenues that can be generated from each customer segment, such as sales of goods or services, user fees, membership fees, leasing, renting, lending, royalties, and advertising fees.
14. **Key performance indicators for success of failure:** Business metrics that reflect the success of the use case in various areas.

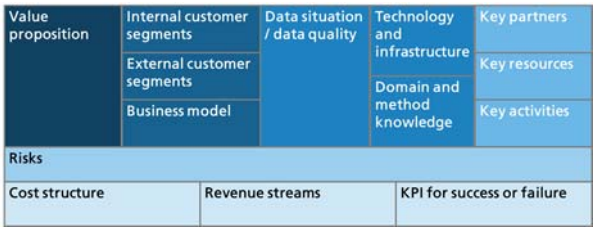


Fig. 3. The Fraunhofer IFF AI Use Case Canvas

The AI Use Case Canvas is developed and illustrated together with the managers and specialists of the company, so that decision makers always have the possibility to continue working with this structure and use it as a strategic decision template. The objective is to accompany the SME in its transformation into a future- and AI-capable company and to make it capable of acting in the field of AI in a number of steps.

5. Conclusions

In order for companies to profitably exploit the potential of new AI technologies, it is essential to determine the current state of its business use cases. This is where the maturity model as a component of the AI CheckUp can provide support for SMEs. The aim of the AI Maturity Model is to support companies in taking advantage of the economic opportunities offered by AI and machine learning in particular. In order to make the most of the potential, it is important to pursue a human-centric AI approach and to develop measures and solutions that employees can trust and accept. By means of the identification of business use cases for AI, the analysis of fields of action and a use case-specific classification, an evaluation of efficiency improvement potentials can be made so that individual measures for increasing the AI maturity can be derived. In the end, an SME can benefit from decision support including an effort-benefit analysis as tangible added value. Thus, in the context of digitization, SMEs can undertake an incremental development of their business model, with manufacturers evolving from a pure product sale to an as-a-service business model, for example, and be able to keep up with multinational companies.

Acknowledgements

The development of the Fraunhofer IFF AI Maturity Model was supported by Christian Löwke, Andreas Herzog and Lina Lau. We would like to thank you for your professional contribution.

References

[1]. Bitkom. Industrie 4.0 – so digital sind Deutschlands Fabriken, May 19, 2020, p. 8. https://www.bitkom.org/sites/default/files/2020-05/200519_bitkomprasentation_industrie40_2020_final.pdf, Accessed: 14.11.2021, in German.
[2]. De Bruin, T., Rosemann, M., Freeze, R., Kulkarni, U., Understanding the Main Phases of Developing a Maturity Assessment Model, in *Proceedings of the 16th Australasian Conference on Information Systems (ACIS),* 2005 p. 3.
[3]. German Engineering Association (VDMA), Leitfaden Künstliche Intelligenz – Potenziale und Umsetzungen im Mittelstand, *VDMA Bayern*, Munich, Bavaria, 2020. https://www.vdma.org/documents/34570/1052572/Leitfaden+K%C3%BCnstliche+Intelligenz-Potenziale+und+Umsetzungen+im+Mittelstand.pdf/ce38a591-68cb-9775-101e-d7cad064b149?t=1615364023575). Accessed: 07.01.2021 (in German).
[4]. Engineering Lifecycle Management / 6.0.4. Anwendungsfälle definieren, *IBM Corporation,* 2017. https://www.ibm.com/docs/de/elm/6.0.4?topic=requirements-defining-use-cases. Accessed: 07.01.2022 (in German).
[5]. Häberer S., Kujath M., Flechtner E., Industrie 4.0-CheckUp, wt Werkstattstechnik online, wt Jahrgang 107, 2017, Heft 03 , pp. 141-142.
[6]. OwnML, The Machine Learning Canvas, https://www.ownml.co/machine-learning-canvas Accessed: 13.01.2022.
[7]. Osterwalder, A., Strategyzer AG, Löwenstrasse 2 CH, 8001 Zürich, Switzerland 2021, https://www.alexosterwalder.com/, Accessed: 21.12.2021.
[8]. Osterwalder, A., Pigneur, Y. Business Model Generation. A Handbook for Visionaries, Game Changers, and Challengers, *John Wiley & Sons*, 2010, pp. 200-210.

[9]. The German Federal Government. Strategie Künstliche Intelligenz der Bundesregierung: Fortschreibung 2020, 2020. https://www.ki-strategie-deutschland.de/files/downloads/201201_Fortschreibung_KI-Strategie.pdf. Accessed: 07.09.2021 (in German).

[10]. Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste. Künstliche Intelligenz im Mittelstand – Relevanz, Anwendungen, Transfer. WIK GmbH, Rhöndorfer Straße, 6853604 Bad Honnef. 2019. https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5, Accessed: 14.11.2021 (in German).

(014)

Management and Path Planning Solution for Parking Facilities using Dynamic Load Balancing

F. D. Sandru, V. I. Ungureanu and I. Silea
Automation and Applied Informatics Department, University Politehnica Timisoara, Timisoara, Romania
E-mail: florin-d.sandru@upt.ro, vlad.ungureanu @upt.ro, ioan.silea@upt.ro

Summary: The growth of personal vehicle ownership has had a direct impact on the need to accommodate vehicles into structures that haven't been conceived to handle such requirements of usage. One impacted area by this change is the optimal usage of existing parking facilities, considering an increase in parking space needs, existing facilities will be filled faster than before. While navigating empty or near-empty facilities results in a reduced effort in finding available spaces, the moment a driver must navigate the existing infrastructure almost completely or multiple times has negative effects both economically and for the environment. This paper presents a method of managing the position of vehicles in parking facilities based on a multivariable approach with a load balancing component. The presented solution provides minimal functionality even if data sources are unavailable thus allowing permanent service availability.

Keywords: Parking, Route optimization, Reduced travel time, Load balancing, distance, Dijkstra, SUMO.

1. Introduction

One of the aspects that impact both vehicle owners and infrastructure providers both private and public is the storage of the vehicle while it is not in use. Infrastructure providers need to plan for the current and future demand of the offered facilities while needing to provide appropriate maintenance and management for existing solutions, guidelines for this are available in works like [1] where already proven strategies are laid out. Vehicle owners need to determine the optimal cost (travel time, monetary) while choosing a parking solution. While the individual decision regarding the monetary factor has only an indirect effect on infrastructure providers (loss of income, this can be mitigated by designing the facility to the right capacity), offering vehicle drivers the choice of a specific parking space can impact the infrastructure by generating traffic congestions on more desired parts of the facility. A significant number of research articles have as a scope management solution for parking facilities, a distinction needs to be made in between the type of solutions they offer, while papers like [2] provide specific solutions for part of the management process (access control) other papers like [3], make a holistic approach and provide an all-in-one solution for the given infrastructure. Considering advances in mobile computing solutions (smartphones, dedicated navigation systems), the driver is presented the option of getting real-time traffic information and adjusting driving behavior accordingly. Some of the problems associated with this solutions are presented in [4], according to the article users of such systems might end up in more difficult driving scenarios or traffic congestions will be artificially created because the route chosen might not be able to handle the increase in traffic, part of the problem has its cause due to competition in-between solution providers, in case of an enclosed and unique management system, this is not the case.

The navigation component (guidance of the vehicle from start to destination) is most associated with map-based systems to a minima-based search in a graph also called the shortest path problem. A review study of various algorithms is made available in [5], for this paper the Dijkstra algorithm was chosen as the basis for implementation due to its good performance output regarding computational power and previous proof of concepts being available. The Dijkstra algorithm can be improved by reinterpreting the cost for a specific use case, one example of such extension is made in [6] where the cost is associated with the fuel consumption of vehicles traveling on the road infrastructure, a similar approach is made also in this paper but considering as a cost the total travel time and the congestion generated by vehicles.

Congestion issues are typically seen in computer networks and computing tasks where the non-optimal use of resources can cause serious problems, one mitigation method is the use of load balancing. Load balancing consists of the distribution of tasks among the available system resources. A characterization of load balancing types centered is made in [7], even though cloud computing centered the methods described can be applied to the optimal vehicle routing process. Few papers have explored this methodology in a parking management context, the examples include [8] and [9] where the distribution was made on the scale of multiple parking facilities.

2. Solution Description

The goal of the solution is to provide the optimal travel time from a facility and not an individual's point of view. This is achieved by using a multi-parameter approach based on the information that can be made available inside the parking facility.

2.1. Definition of Input Data

The basic data inputs were considered the ones with the least technological implication: the road infrastructure consisting of entry and exit points, the connecting roads with characteristics regarding length, and the total amount of parking spaces and allowed maneuvering.

More complex types of data are considered for the advanced feature of the algorithm, those data types include access and entry times for dedicated infrastructure nodes, the number of vehicles targeting the parking facility, the destination when exiting the parking facility, and the congestion of each traffic element. One important factor in regards to this type of dynamic data is that their acquisition will increase both the initial cost of the facility due to the sensors needed or infrastructure needed to acquire them and the general computing requirements.

2.2. Algorithm Description

As previously mentioned, the developed algorithm can be seen as an extended version of the Dijkstra algorithm, in the following ways: the cost function of the starting point is not zero as with the standard approach as it reflects the entry time to the facility, it considers subdivisions of the graph without generating new point, it features the balancing component by increasing the cost for paths that already have traffic assigned to them.

In a common navigation system, the minima is computed based on the physical distance between the start and the trip destination, the proposed algorithm deviates from this definition and defines the primary minima condition as the smallest amount of time spent in traffic. This change is supported as map providers include besides the road length also information about various speed limits, thus the first component used for evaluation becomes the one described in eq. 1

$$t_{seg} = l_{seg}/s_{seg}, \quad (1)$$

where:
t_{seg} = necessary time to traverse a map segment [s];
l_{seg} = legnth of the segment [m];
s_{seg} = the maximun allowed speed to traverse the segment [m/s].

The information obtained by this step is purely static and doesn't take into account the real-world environment like infrastructure use resulting in a degradation of the projected arrival cost.

Adding the dynamic characteristic data about the other traffic participants is needed, this can be either sourced from measurement systems or is known in case the path planning is done by a centralized mechanism. This component is presented in eq 2.

$$d_{seg} = l_f \times l_{seg}/s_{seg}, \quad (2)$$

where:
l_f = load factor for the segment;
l_{seg} and s_{seg} having the same meaning as before.

The load factor is expressed using eq. 3:

$$l_f = \begin{cases} \frac{\sum l_{veh}}{l_{seg}}, res > 1 \\ 0, \ res < 1 \end{cases}, \quad (3)$$

where:
l_{veh} = length of the vehicles traveling on the segment;
res = usage condition, the computed values is used only if the vehicle demand exceeds the available infrastructure.

This approach allows to inflate the cost in case the road structure usage is above the road capacity, the factor is only applied once the strain is above the capacity.

Additionally to this, to better capture the parking specific maneuver downtime on the road network, a time penalty is added for vehicles that would be parking or exiting the parking space Δt_m [s].

Thus the actual cost function becomes the one described in eq. 4:

$$c_t = \sum_{n=0}^{seg\_max} t_{seg} + d_{seg} + \Delta t_m \quad (4)$$

Another specificity added to the algorithm is the fact that the check for capacity and infrastructure entry time is done before allowing vehicle access to the given infrastructure.

The complete flow can be seen in Fig. 1.

3. Simulation and Results

A medium-sized parking lot was designed to be used as a demonstrator for the deployment, in the initial conception phase a dedicated computer-aided design software (LibreCAD) was used, the output of the projection phase is visible in Fig. 2.

The parking facility features the following elements: a single one-way street connecting the two entry and exit points. The parking lot is divided into parking areas consisting of the parking spaces connected to a road segment. The parking space dimensions were chosen to be 2.5 m for with and 5m for length. A general speed limit of 15Km/h is considered for the traffic inside the parking facility while the access road has a speed limit of 50 Km/h.

The simulation necessities, network definition, static and dynamic information extraction, route generation, and vehicle control were done using the SUMO (Simulation of Urban MObility) package features, developed by the German Aerospace Center, this is "an open-source, highly portable, microscopic and continuous traffic simulation package designed to handle large networks" [10], while its primary scope is not the simulation of parking facilities the features it offers were of great help in implementing the simulation scenarios and solution by extending and relying upon it in the developed code.

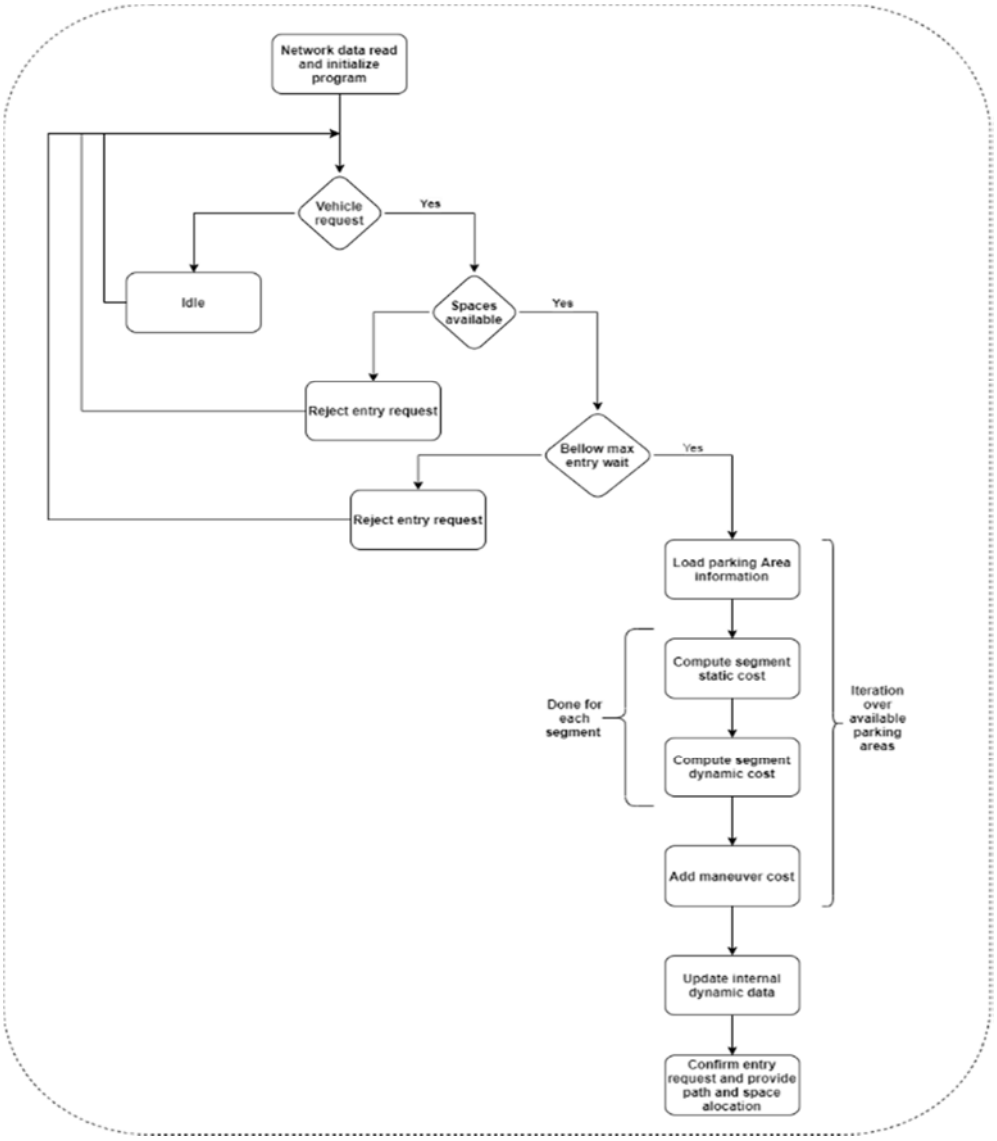


Fig. 1. Flow chart representing the proposed algorithm.

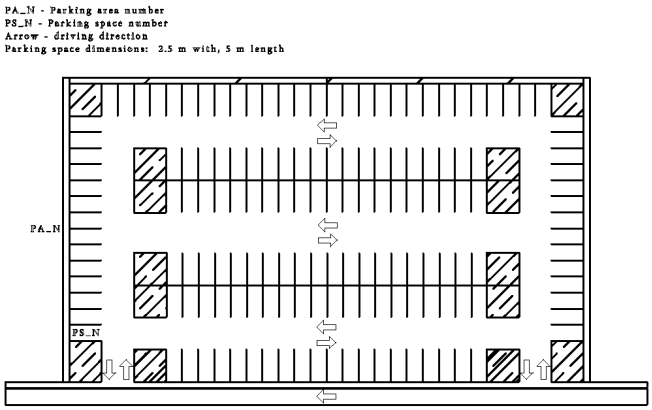


Fig. 2. Parking facility designed for validation.

As the SUMO tooling requires a specific input format for being able to simulate the environment, the initial schematic was recreated in the netedit utility denoting the network nodes and the associated restrains. The network provides direct information used by the solution scripting for parameters like junctions, direction change restrictions, segment length, max segment speed, and the association between the parking area and the access road. Specific to the simulation is the computation of a parking space

offset, due to the association being made on parking area level with each parking space the with of the individual space needs to be accounted for when computing the distance travelled. Another specificity of the simulation is the slit of the parking area due the presence of a junction, this is visible in the lateral parking areas thus increasing the number of areas that need to be monitored. The parking facility netedit specific design is visible in Fig. 3.

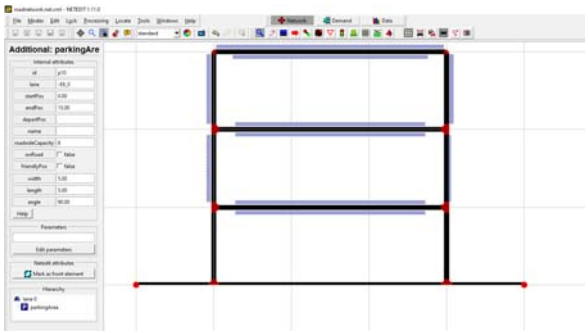


Fig. 3. Parking facility design translation for simulation.

Currently, this step has been done manually, the available scripting allows the direct import only of OSM (Open Street Map) format, the authors are considering the development of a script that would make it possible to do type conversions DXF file format to the XML file format used by SUMO.

The control of the simulation is done from a python-based implementation of the parking manager using the TraCI API, Traci allows this by creating a TCP-based connection to the SUMO application and using that connection to pass data. The high-level overview of the SW components used is visible in Fig. 4. Application internals where specifically developed for this study while existing elements are marked separately.

The simulation phase considered 3 different scenarios for choosing a parking space for a vehicle next in queue:

a) Guidance based solely on proximity to the target;
b) A randomized allocation of parking spaces with guidance instructions;
c) The proposed extension to the algorithm to balance the network usage.

The evaluation criteria considered was the least time spent idling, this value is taken from the simulation and as per TraCI API [11] is defined as "The waiting time of a vehicle is defined as the time (in seconds) spent with a speed below 0.1 m/s since the last time it was faster than 0.1 m/s."

As a general path target, vehicles always originate in the right outmost extremity of the access road and their destination is a parking space within the facility.

In a direct comparison, the 3 scenarios have clear differences from a performance point of view, the time-based criteria show the ability of the developed solution to reduce the general time spent in traffic, by having the same demand this is visible in Fig. 5.

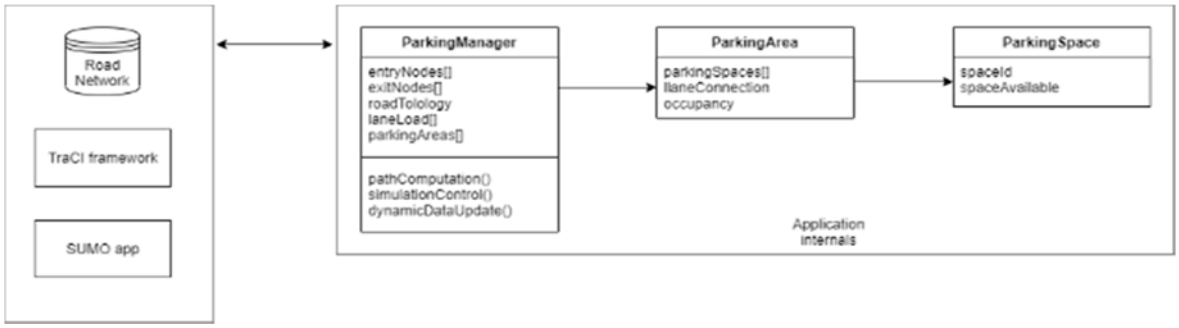


Fig. 4. Software overview.

Fig. 5. Solution performance in loading the complete parking facility.

The individual max idle time per solutin are presented in Fig. 6, Fig. 7 and Fig. 8.

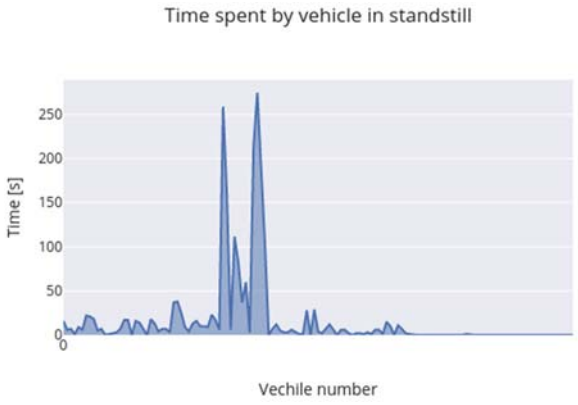


Fig. 6. Idle time for strategy a).

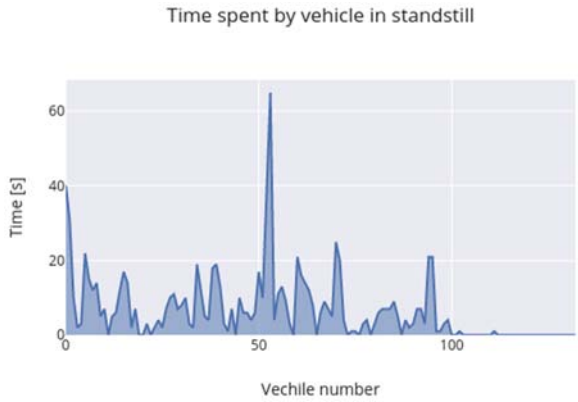


Fig. 7. Idle time for strategy b).

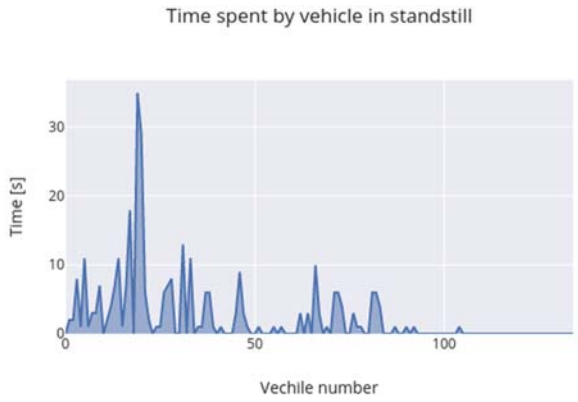


Fig. 8. Idle time for strategy c).

Allocation based on the minimal driving distance, this approach generates excessive strain on the first connection points to the parking area and results in congestion. Waiting times observed during these simulations were the highest while choosing a random parking area among the available structures yields better results than the closest available area.
Higher waiting times are still visible this is due to the bottlenecking of the entryway.

4. Proposals for the Physical Realization of the System

Human-machine interface: Depending on the level of automation of the vehicle human interaction might or might not be needed, in the case of a non-autonomous vehicle, the information about parking space availability, the assigned position, and the navigation can either be done using an application running on a smartphone or directly on a vehicle integrated navigation system.

Access control: While the enforcement of the entry and exit rights is usually done using physical barriers this approach would only lead to delays, considering this a solution where the access rights are checked by the management system is the desired one. As the parking space is assigned by the system any occupancy of a non-assigned parking space could be considered a violation in the terms of usage in case an unauthorized vehicle is occupying a parking space blocking that vehicle on the used parking space till the service payment is made is considered the preferred approach. The monitoring is proposed to be done using camera systems capable of detecting the presence of vehicles and associating their position to the model of the facility.

Vehicle navigation: The maneuvering of the vehicle inside the parking structure can be one of the most challenging aspects of operation, the map data of the parking structure would either need to be made available beforehand and associated to a general navigation map or merged upon the proximity to the entry position and deleted upon exit to reduce the storage requirements, another approach could consist in using a dedicated reduced map consisting only of the parking facility inside a specific application on a smartphone. Positioning is another key factor, while a combination of a GNSS system with a map matcher component would provide a certain degree of precision, this might not meet the constraints to be within a few meters. In this case, a solution based on using inertial data like described in [12] proves itself useful but would not be able to compensate for a complete loss of GNSS signal throughout a complete trip. A possible solution could consist in the usage of a camera-based system that would be able to recognize specific signage and associate it to the corresponding infrastructure elements.

5. Conclusions

The solution can determine based on a given infrastructure the necessary parking lot information and provide access control features in case the capacity is reached while outperforming classic navigation solutions in terms of time spent in traffic.

Future extensions should consider more simulation scenarios and the inclusion of more complex driving behavior (like take over).

The authors are planning after the extended validation phase to also realize a physical implementation of the solution.

References

[1]. Todd Litman, Parking management: strategies, evaluation and planning. Victoria, *BC: Victoria Transport Policy Institute*, 2016.

[2]. E. Hassania Rouan and A. Boumezzough, RFID Based Security and Automatic Parking Access Control System, *Business Intelligence*, 2021, pp. 434–443.

[3]. M. Owayjan, B. Sleem, E. Saad and A. Maroun, Parking management system using mobile application, in *Proceedings of the Sensors Networks Smart and Emerging Technologies Conference (SENSET' 2017)*, 2017, pp. 1-4.

[4]. IEEE Spectrum: Technology, Engineering, and Science News, 2021. https://spectrum.ieee.org/computing/ hardware/your-navigation-app-is-making-traffic-unmanageable (accessed Jun. 12, 2021).

[5]. Madkour, Amgad, et al., A survey of shortest-path algorithms, *International Journal of Applied Engineering Research*, Vol. 13, No. 9, 2018, pp. 6817-6820.

[6]. Zhang, Jin-dong, Yu-jie Feng, Fei-fei Shi, Gang Wang, Bin Ma, Rui-sheng Li, and Xiao-yan Jia, Vehicle routing in urban areas based on the oil consumption weight-Dijkstra algorithm, *IET Intelligent Transport Systems*, 10, No. 7, 2016, pp. 495-502.

[7]. Katyal, Mayanka, and Atul Mishra, A comparative study of load balancing algorithms in cloud computing environment, *International Journal of Computer Applications*, 117, 24, 2014, pp. 33-37.

[8]. Kim, Oanh Tran Thi, Nguyen H. Tran, Chuan Pham, Tuan LeAnh, My T. Thai, and Choong Seon Hong, Parking assignment: Minimizing parking expenses and balancing parking demand among multiple parking lots, *IEEE Transactions on Automation Science and Engineering*, 17, no. 3, 2019, pp. 1320-1331.

[9]. A. Souza, Z. Wen, N. Cacho, A. Romanovsky, P. James and R. Ranjan, Using Osmotic Services Composition for Dynamic Load Balancing of Smart City Applications, in *Proceedings of the IEEE 11th Conference on Service-Oriented Computing and Applications (SOCA' 2018)*, 2018, pp. 145-152.

[10]. P. A. Lopez et al., Microscopic Traffic Simulation using SUMO, in *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC' 18)*, 2018, pp. 2575-2582.

[11]. Python: module traci._vehicle, Sumo.dlr.de, 2021. https://sumo.dlr.de/daily/pydoc/traci._vehicle.html (accessed Jan. 2, 2022).

[12]. Sandru F. D., Nanu S., Silea I. and Miclea R. C., Kalman and Butterworth filtering for GNSS/INS data, in *Proceedings of the 12th IEEE International Symposium on Electronics and Telecommunications (ISETC'2016)*, 2016, pp. 257-260.

(015)

Switching Propulsion Mechanisms of Tubular Catalytic Micromotors

P. Wrede [1,2], M. M. Sánchez [1], V. M. Fomin [1,3,4] and O. G. Schmidt [1,5,6]
[1] Institute for Integrative Nanosciences (IIN), Leibniz Institute for Solid State and Materials Research (IFW) Dresden, Dresden, Germany
[2] Max Planck Institute for Intelligent Systems and Max Planck ETH Center for Learning Systems, Stuttgart, Germany
[3] Institute of Engineering Physics for Biomedicine, National Research Nuclear University "MEPhI", Moscow, Russia
[4] Laboratory of Physics and Engineering of Nanomaterials, Department of Theoretical Physics, Moldova State University, Chişinău, Republic of Moldova
[5] Material Systems for Nanoelectronics, TU Chemnitz, Chemnitz, Germany
[6] Center for Materials, Architectures and Integration of Nanomembranes (MAIN), TU Chemnitz, Chemnitz, Germany
Tel.: +49 711-689-3424, fax: + 49 711-689-3412
E-mail: wrede@is.mpg.de

Summary: Different propulsion mechanisms have been suggested for a variety of chemical micromotors. Their high efficiency and thrust force enable several applications in the fields of environmental remediation and biomedicine. In particular, bubble-recoil based motion has been modeled by three different phenomena: capillary forces, bubble growth, and bubble expulsion. However, these models have been suggested independently based on a single influencing factor (i.e., viscosity), limiting the understanding of the overall micromotor performance. Here the combined effect of medium viscosity, surface tension, and fuel concentration on the micromotor swimming performance is analysed. Hence the dominant propulsion mechanisms describing its motion are accurately identified. Using statistically relevant experimental data, a holistic theoretical model is proposed for bubble-propelled tubular catalytic micromotors that includes all three above-mentioned phenomena. The proposed model provides deeper insights into their propulsion physics toward optimized geometries and experimental conditions. These findings pave the way for highly optimized catalytic micromotors for various applications.

Keywords: Microrobots, Chemical propulsion, Viscosity sensing, Micromachines, Micro-/nanotechnology.

1. Introduction

For more than a decade researcher are developing new chemical micro-devices, including micromotors. Owing to their small size these highly controllable man-made machines enable precise applications for targeted medical treatments [1, 2] and environmental remediation [3, 4]. Moreover, micromotors are interesting model systems for studying motion mechanisms as well as other physical phenomena at low Reynolds number regimes. A particularly interesting type of chemically-driven propulsion is the bubble-recoil mechanism. Micromotors utilising this specific mechanism show excellent motion performance in terms of speed and thrust force with simple designs. They are mainly made of asymmetrical tubes coated with an inner surface of a catalytic material (e.g., Pt, Ag, Pd, enzyme) and an outside functional layer (e.g., Au, Fe, SiO_2, TiO_2) for their guidance, in situ reactions or biofunctionalization purposes [5]. A chemical reaction between the fuel liquid and a catalyst coating the inner surface of the micromotor converts chemical energy into mechanical motion. The most commonly used reaction is the decomposition of H_2O_2 into O_2 and H_2O using Pt as catalyst. Bubbles are formed by O_2, which is produced inside the tube, leading to a uni- or bidirectional movement, depending on the micromotor geometry [6].

To reach a fundamental understanding of the propulsion physics of such micromotors, we fabricate micromotors with different lengths and semi-cone angles using two-photon lithography. Based on these structures, the influence of different concentrations of fuel, surfactant and the viscosity of the environment as well as geometric parameters on the propulsion mode is analysed [7]. We show that different propulsion mechanisms can be observed individually or simultaneously in the same experiment. This provides new insights into the propulsion mechanisms and the possibility to further optimize and simulate the behaviour of bubble propelled micromotors for operating in more complex media.

2. Simulation and Switching of Propulsion Mechanis

The motion of conical bubble propelled micromotors is described by three established mechanisms, the capillary force [8], bubble growth [9] and jet-like propulsion [10]. Our experiments imply

that, depending on the composition of the fuel fluid, a switching of propulsion mechanisms occurs [7]. The use of a surfactant, here SDS, reduces the surface tension of the medium hence reducing the drag force acting on the bubble during its motion through the micromotor. This leads to an increased micromotor velocity for SDS concentration rising from 1.25 % till 10 %. This effect reverses for SDS concentrations larger than 10 %. The amount of surfactant present absorbable by the micromotor scales with the available surface area. After reaching the threshold of 10 % SDS, when the used micromotors reveal their maximum SDS absorbance, the positive effect of the surfactant on the micromotors speed diminishes. A dominant influence of the bubble growth mechanism is observed for the concentrations higher than 20 % SDS (at 2.5 % H_2O_2) as well as for the concentrations of 0.4 % MC and higher (with viscosity above 0.004 Pa·s) at 2.5 % H_2O_2 and 5 % SDS (Fig. 1). In those cases, insufficient force is exerted on the bubble to detach it from the tube. This causes the bubble to keep growing much bigger than the larger opening of the micromotor. Due to the motion of the bubble in the micromotor, a relatively small contribution of the capillary force is also present. Anyway, it can be neglected for a sufficiently long time because no second bubble is formed inside the micromotor. With increasing MC concentration, the viscosity of the fluid, and therefore the drag force counteracting the propulsion force of the micromotor, increase. This leads to an increase in propulsion speed. Additionally, the MC forms a passivating layer around the Pt inside the micromotor. This results in a reduction of the chemical activity of platinum. Hence, less H_2O_2 is decomposed and thus less O_2 is produced. As a consequence, the bubble release frequency decreases, allowing the bubbles to collect more O_2, which results in an increased bubble diameter. Our experimental data suggest that starting from the concentration of about 20 % H_2O_2 (the specific value depends on the system under study), the bubble expulsion mechanism dominates over the other two mechanisms (Fig. 1) by two reasons. Firstly, the bubbles move along the tube so quickly, that their radii remain smaller than that of the tube. Accordingly, there is no contact between the bubble and the inner wall to induce a capillary force. Secondly, due to their small size, the bubbles leave the micromotor without adhering, so that no growth force emerges. The main reason for these observations is the increased amount of the available H_2O_2 when increasing concentration. This, in turn, leads to an increased production of O_2, and hence, to a quicker formation of bubbles. Accordingly, the bubble release frequency increases, and so does the micromotor speed. Due to this increased frequency and the associated shorter time, during which the bubbles stay in the tube, they can also collect less oxygen leading to a decreased bubble radius and a smaller rise in speed for H_2O_2 concentrations over 10 %. For all remaining fluid parameters, all three mechanisms need to be considered to describe the micromotors motion (Fig. 1).

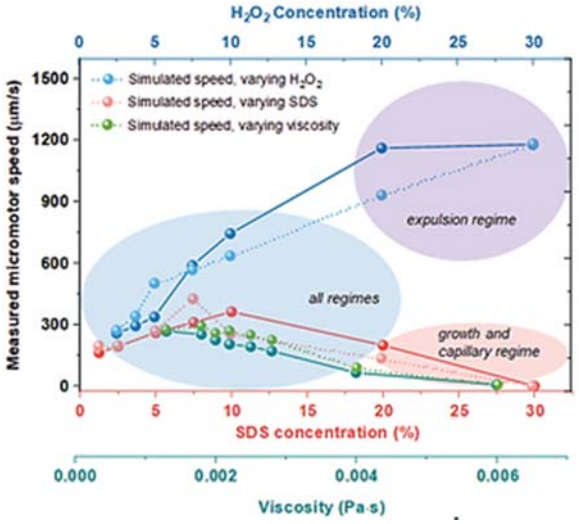


Fig. 1. Measured and simulated micromotor speeds for various fluidparameters. (After Wrede et al. [7]).

Hence, for our simulations, all three mechanisms are taken into account, in particular, for the H_2O_2 concentrations lower than 20 %, while as for the H_2O_2 concentration above this value, only the bubble expulsion mechanism functions. Considering all mechanisms in our model, the simulated speed as a function of the H_2O_2 concentration (ranging from 2.5 % to 30 %) is in good agreement with the experimental data. All three propulsion mechanisms are also considered for the whole range of the MC concentrations (from 0.05 % to 0.25 %, corresponding to viscosities from 1.3 to 6 mPa·s), as well as for the whole range of the SDS concentrations (from 1.25 % to 10 %). The growth mechanism is dominant for the MC concentrations of 0.4 % and 0.6 % as well as for the SDS concentrations of 20 % and 30 %.

Additionally, we evaluate the effect of different geometric parameters on the performance of the here presented micromotors, scaling the micromotor semi-cone angle and length. In particular, 50 µm long micromotors with the varying semi-cone angle (2.5°, 5°, and 10°), as well as micromotors with the semi-cone angle of 5° and varying lengths of 25, 50, and 100 µm, are analysed. An increase of the micromotor length from 25 to 100 µm results in an increase of the micromotor speed, while an increase in the semi-cone angle slows the micromotor down. To better understand this behaviour, in our simulations, we assume certain dependencies: when increasing the semi-cone angle of the micromotor, we keep the smaller radius constant, therefore the larger radius increases. The same applies for an increase in length. Such an increase in the larger opening of the tube leads to the production of high-volume bubbles. Using our experimental data for a micromotor with the length of 50 µm and the semi-cone angle of 5°, the bubble radius is twice as large as the opening radius for different H_2O_2 concentrations (up to 10 %). For larger H_2O_2 concentrations, the bubble radius is similar to the opening radius. This effect is caused by a higher bubble expulsion frequency. The latter depends on the

active catalytic area of the micromotor, which causes an increase in the oxygen production rate. An increase of both the semi-cone angle and the length of the micromotors leads to an increase in the catalytic surface area. Particularly, an increase of the semi-cone angle from 2.5° to 5° for a conical micromotor with a length of 50 μm results in a surface area decrease of 20 %. Additionally, larger semi-cone angles increase the drag force acting on the micromotor. In summary, an increase in bubble expulsion frequency and bubble radius is predicted for increasing micromotor length and semi-cone angle. At the same time, an increased speed is predicted for smaller semi-cone angles and longer micromotors.

3. Conclusion

A physical explanation of the bubble-propelled micromotor motion is performed using the established capillary force, bubble growth, and jet-like propulsion mechanisms. Switching of propulsion mechanisms is for the first time unveiled at certain values of the H_2O_2, MC, and SDS concentrations. Besides, it is observed that for certain experimental parameters, a concerted action of all three propulsion mechanisms is observed. A theoretical model is proposed to calculate the speed of the bubble-propelled micromotors, including the contributions of all three known propulsion mechanisms, resulting in good agreement with the obtained experimental data. This offers new possibilities to optimize and predict the performance of bubble-propelled micromotors for different applications in various working environments, including new applications, such as, in-situ sensing of the medium viscosity, as well as for fundamental understanding of the impact of various geometrical parameters and experimental conditions (length, semi-cone angle, surface properties) on the propulsion of conical catalytic micromotors.

References

[1]. Medina-Sánchez, M., Xu, H. and Schmidt, O. G. Micro- and nano-motors: The new generation of drug carriers, *Therapeutic Delivery,* 9, 4, 2018, pp. 303–316.

[2]. Wang, B., Kostarelos, K., Nelson, B. J. and Zhang, L., Trends in micro-/nanorobotics: materials development, actuation, localization, and system integration for biomedical applications, *Advanced Materials,* 33, 2021, pp. 1–44.

[3]. Srivastava, S. K., Guix, M. & Schmidt, O. G., Wastewater mediated activation of micromotors for efficient water cleaning, *Nano Letters,* 16, 2016, pp. 817–821.

[4]. Jurado-Sánchez, B. & Wang, J., Micromotors for environmental applications: A review. *Environmental Science: Nano,* 7, 2018, pp. 1530–1544.

[5]. Zhang, Y., Yuan, K. & Zhang, L., Micro/ Nanomachines: From functionalization to sensing and removal, *Advanced Materials Technologies,* 4, 2019, pp. 1–22.

[6]. Mei, Y. F., Naeem, F., Bolaños Quiñones, V. A., Huang, G. S., Liao, F., Li, Y., Manjare, M., Naeem, S., Solovev, A. A. and Zhang, J., Tubular catalytic micromotors in transition from unidirectional bubble sequences to more complex bidirectional motion, *Applied Physics Letters* 114, 2019, pp. 1–16.

[7]. Wrede, P., Medina-Sánchez, M., Fomin, V. M. and Schmidt, O. G., Switching propulsion mechanisms of tubular catalytic micromotors, *Small,* 17, 2021, pp. 1–11.

[8]. Klingner, A., Khalil, I. S. M., Magdanz, V., Fomin, V. M., Schmidt, O. G. and Misra, S., Modeling of unidirectional-overloaded transition in catalytic tubular microjets, *Journal of Physical Chemistry C,* 121, 2017, pp. 14854–14863.

[9]. Manjare, M., Yang, B. and Zhao, Y. P. Bubble-propelled microjets: Model and experiment. *Journal of Physical Chemistry C,* 117, 2013, pp. 4657–4665.

[10]. Li, L., Wang, J., Li, T., Song, W. and Zhang, G., Hydrodynamics and propulsion mechanism of self-propelled catalytic micromotors: Model and experiment, *Soft Matter*, 10, 2014, pp. 7511–7518.

(019)

Stability Margins for Linear Periodically Time-Varying Systems

Xiaojing Yang
School of ASE, Beihang University, 37 Xueyuan Rd., 100191 Beijing, China
Tel.: + 86 13240488262, fax: + 86 13240488262
E-mail: XYang@buaa.edu.cn; Pandonglei@126.com

Summary: Singular Perturbation Margin (SPM) and Generalized Gain Margin (GGM) are Phase Margin (PM) and Gain Margin (GM) like stability metrics. In this paper, the problem of SPM and GGM assessment for Linear Periodically Time-Varying (LPTV) systems is formulated. Chang transformation makes it possible to reduce the SPM analysis for Hill equations, which is essentially a stability problem of higher order LPTV systems due to the SPM gauge introduced dynamics, to the second order LPTV systems. Based upon Floquet Theory, the SPM and the GGM assessment methods for the second order and the general order LPTV systems are established, respectively.

Keywords: Stability margin, Singular perturbation, Regular perturbation, Linear system, Periodically time-varying systems, Hill equation.

1. Introduction

Research on periodically time-varying systems can be traced back to M. Faraday [1] in 1830s; the first detailed theory study for Linear Periodically Time-Varying (LPTV) systems is given by E. Mathieu [2]; G. Floquet established Floquet theory [3] in 1883, which forms the basis of a great many of the descriptions of parametric behaviors [4]; one of the most significant earlier papers on periodically time-varying systems was G. W. Hill [5], whose name has been given to the general class of the second order periodic differential equations, and his research is the first investigation of a practical problem in such field. Due to a growing number of applications of LPTV systems in different fields, since the first half of the 20th century, more study effort was put on Meissner equation [6], Mathieu equation [7], the solutions [8] and the stability behavior of the above equations [9], etc. Through the continuous hard work of scholars for the recent 60-70 years, the research methodology of LPTV systems has been formed a relatively mature body of knowledge, and can be roughly divided into three categories: 1, find the time invariant equivalent system for the periodically time-varying system and further discuss the problems of stability and pole assignment [10-13]; 2, analyze the controllability and oberservability of LPTV systems based on the state transition matrix of one period [14, 15]; 3, Lyapunov methods [16-19].

Stability is a key problem in the design of automatic control systems, and in engineering applications, not only the question whether the system is stable is must known information, but also the quantitative criteria of stability is pivotally demanded. Recently, Singular Perturbation Margin (SPM) and Generalized Gain Margin (GGM) [20] have been proposed as the classical Phase Margin (PM) and Gain Margin (GM) like stability metrics for Single Input Single Output (SISO) Linear Time Invariant (LTI) and Nonlinear Time Invariant (NLTI) systems [21] from the view of the singular perturbation and the regular perturbation. Due to the distinguishing feature of the exponential stability of the equilibrium point, i. e. the robustness to the nonvanishing regular perturbations, the concepts of SPM and GGM have been investigated towards to the establishment of the quantitative exponential stability metrics for general Multi Input Multi Output (MIMO) Nonlinear Time Varying (NLTV) systems, defined by the maximum singular perturbation parameter, denoted by $\varepsilon_{\max}$, and the maximum and minimum regular perturbation parameters, denoted by $k_{\min}$ and $k_{\max}$ that render the perturbed closed-loop system at onset of instability to gauge the capability in accommodating singular perturbations (parasitic dynamics) and regular perturbations (parametric dispersions). Furthermore, the long term effort of such research is the development of the theoretically based and practically efficient SPM and GGM analysis methodology to facilitate the nonlinear and time-varying control system design.

In this paper, the problem of SPM and GGM analysis for LPTV systems is formulated in Section Objectives; the main results: decoupling technique of the singularly perturbed LPTV systems; SPM gauge design; SPM and the GGM assessment methods, examples etc., are listed in Section 3; Section 4 concludes the whole paper's with some insightful remarks.

2. Objective

The SPM and GGM analysis for a closed-loop and well-designed LPTV control system (also called Nominal System)

$$\dot{x} = A_{nom}(t)x\,, \tag{1}$$

where $x \in \mathbb{R}^n$ is the state vector, $A_{nom}(t)$ is bounded and continuously differentiable, satisfying

$$A_{nom}(t) = A_{nom}(t+T) \tag{2}$$

with T the system period, is referred as to the quantitative assessment of the capability of the exponentially stable null equilibrium (origin) in accommodating singular perturbations and regular perturbations:

(1) Design the SPM gauge and GGM gauge for the specific perturbations;
(2) Determine the stability of the perturbed system using stability criterion when different singular perturbation parameters and gain parameters are given, respectively;
(3) Obtain the SPM and GGM by the marginally stable situation of the perturbed system;
(4) Reveal the relation between the characteristics of the nominal system (1) and its exact SPM and GGM.

3. Perturbed Hill Equation

3.1. Stability Criterion and SPM/GGM Gauges

We first introduce a necessary and sufficient exponential stability condition for LPTV systems that will be used in the sequel.

Lemma 1 [4] *The natural response of an LPTV system is stable, if and only if no eigenvalues of the Discrete Transition Matrix (DTM) have magnitudes greater than one and that an eigenvalue with unity magnitude is not degenerate.*

The eigenvalues of the DTM is also called Floquet characteristic multipliers of the periodic system. The transition matrix of a general order LPTV system is analytically intractable in Lemma 1, but for the second-order cases, the general form

$$\ddot{x} + \alpha_1(t)\dot{x} + \alpha_0(t) = 0$$

can be converted into an expression without a first derivative, and the classical form is

$$\ddot{y} + (a - 2q\psi(t))y = 0, \quad \psi(t) = \psi(t+\pi), \tag{3}$$

where function $\varphi(t)$ has a period of π and a, q are constant parameters, is identified as Hill equation. Hill equation is not only the model in many applications, but also has significant meaning for general second-order equations. When the singular and regular perturbations are considered in the closed-loop, the block diagram of Hill equation with corresponding perturbation sites is shown by Fig. 1.

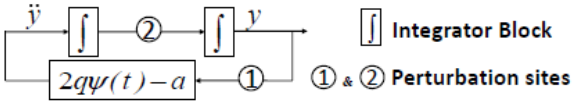


Fig. 1. Block Diagram of Hill Equation with Perturbations.

Design the SISO LTI SPM gauges at the perturbation sites, which might have the equation forms as [20] minimum-phase gauge, zero-error gauge or approximately-linear phase gauge, etc., and the singularly perturbed system is given by

$$\begin{aligned} \dot{x} &= A_{11}(t)x + A_{12}(t)z \\ \varepsilon\dot{z} &= A_{21}(t)x + A_{22}z \end{aligned}, \tag{4}$$

where the singular perturbation parameter ε describes the time-scale separation, $A_{11}(t), A_{12}(t), A_{21}(t), A_{22}$ are function matrices and parameter matrix that are dependent on the perturbation site and the SPM gauge design, and $z \in \mathbb{R}^m$ is state vector of the fast dynamics.

3.2. Decoupled Slow and Fast Systems

The singularly perturbed system (4) can be totally decoupled to slow and fast system by Chang transformation.

Lemma 2 *With the SISO LTI SPM Gauge, the decoupled slow system of the singularly perturbed Hill equation (4) is a Hill equation with the same period to the nominal system.*

Proof. Consider the two perturbation positions, Site 1 and Site 2, in the loop of Fig. 1.

Site 1: With the SISO LTI SPM Gauge, the system matrices of the singularly perturbed system in form of (4) are given by

$$A_{11} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, A_{12}(t) = \begin{bmatrix} 0 & \cdots & 0 \\ & \left(-a + 2q\psi(t)\right)C_f & \end{bmatrix},$$
$$A_{21} = B_f \begin{bmatrix} 1 & 0 \end{bmatrix}, A_{22} = A_f,$$

where A_f, B_f, C_f are the parameter matrices of the SISO SPM gauge. Chang transformation [22] is given by

$$\begin{bmatrix} \xi \\ \zeta \end{bmatrix} = \begin{bmatrix} I - \varepsilon H(t)L(t) & -\varepsilon H(t) \\ L(t) & I \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = S(t) \begin{bmatrix} x \\ z \end{bmatrix}$$

where function matrices *L(t)* and *H(t)* satisfy Differential Riccati Equation (DRE) and Differential Sylvester Equation (DSE)

$$-\dot{L}(t) = L(t)A_{11} + \frac{1}{\varepsilon}A_{21} - L(t)A_{12}(t)L(t) - \frac{1}{\varepsilon}A_{22}L(t)$$
$$-\dot{H}(t) = A_{11}H(t) + A_{12}(t) - H(t)L(t)A_{12}(t)$$
$$-\frac{1}{\varepsilon}H(t)A_{22} - A_{12}(t)L(t)H(t)$$

For Chang transformation, the iterative solutions of the DRE and DSE in form of

$$L(t)=\sum_{j=0}^{\infty}\varepsilon^j\bar{L}_j(t),\quad H(t)=\sum_{j=0}^{\infty}\varepsilon^j\bar{H}_j(t)$$

where

$$\begin{aligned}&\bar{L}_0=A_{22}^{-1}A_{21},\quad \bar{H}_0=A_{12}^{-1}A_{22}\\&\bar{L}_{k+1}=A_{22}^{-1}\left[\dot{\bar{L}}_k+\bar{L}_k\left(A_{11}-A_{12}\bar{L}_k\right)\right],k=0,1,2,...\\&\bar{H}_{k+1}=\left[-\dot{\bar{H}}_k+\bar{H}_k\bar{L}_kA_{12}+\left(A_{11}-A_{12}\bar{L}_k\right)\bar{H}_k\right]A_{22}^{-1}\\&\qquad\qquad k=0,1,2,...\end{aligned}$$

Then, the slow system is given by

$$\begin{aligned}A_{slow}&=A_{11}-A_{12}(t)L(t)\\&=\begin{pmatrix}0&1\\(a-2q\psi(t))C_fL_1(t)&(a-2q\psi(t))C_fL_2(t)\end{pmatrix}\end{aligned}$$

where $L_1(t),L_2(t)$ are the column vectors of $L(t)$, and the second order equation becomes

$$\ddot{x}+(2q\psi(t)-a)C_fL_2(t)\dot{x}+(2q\psi(t)-a)C_fL_1(t)x=0 \quad (5)$$

By transformation

$$x(t)=\exp\left\{-\frac{1}{2}\int_0^t(2q\psi(t)-a)C_fL_2dt\right\}y(t) \quad (6)$$

Eq. (5) is transformed to

$$\begin{aligned}\ddot{y}=-\Big\{&(2q\psi(t)-a)C_fL_1(t)-q\dot{\psi}(t)C_fL_2(t)\\&-\frac{1}{4}(2q\psi(t)-a)^2\left(C_fL_1(t)\right)^2\Big\}y\end{aligned}$$

which is a Hill equation with the same period to function $\psi(t)$.

Site 2: With the SISO LTI SPM Gauge, the system matrices of the singularly perturbed system become

$$A_{11}=\begin{bmatrix}0&0\\-a+2q\psi(t)&0\end{bmatrix},A_{12}(t)=\begin{bmatrix}0&C_f&0\\0&\cdots&0\end{bmatrix},$$
$$A_{21}=B_f\begin{bmatrix}0&1\end{bmatrix},A_{22}=A_f,$$

By Chang transformation, the slow system is given by

$$\begin{aligned}A_{slow}(t)&=A_{11}(t)-A_{12}L(t)\\&=\begin{pmatrix}C_fL_1(t)&C_fL_2(t)\\-(a-2q\psi(t))&0\end{pmatrix}\end{aligned}$$

and the second order slow system becomes

$$\begin{aligned}&\ddot{x}-\left(C_fL_1+\tfrac{C_f\dot{L}_2}{C_fL_2}\right)\dot{x}+\left(C_f\dot{L}_1-\tfrac{C_f\dot{L}_2C_fL_1}{C_fL_2}\right.\\&\left.+(2q\psi(t)-a)C_fL_2\right)x=0\end{aligned} \quad (7)$$

which can be denoted as $\ddot{x}+\alpha_2(t)\dot{x}+\alpha_1(t)x=0$. Then, by transformation

$$x(t)=\exp\left\{-\frac{1}{2}\int_0^t\alpha_2(t)dt\right\}y(t)$$

Eq. (7) is transformed to be

$$\ddot{y}(t)=-\left\{\alpha_1(t)-\frac{1}{2}\dot{\alpha}_2(t)-\frac{1}{4}\alpha_2^2(t)\right\}y(t)$$

which is a Hill equation. □

3.3. SPM Assessment of Hill Equations

By Lemma 2, it is known that with an SISO LTI SPM Gauge, the singularly perturbed Hill equation can be decoupled to a fast system and a slow system. The following Theorem 1 is a direct result of the definition for SPM and Lemma 1.

Theorem 1 *The SPM of the LPTV nominal system is equal to the maximum singular perturbation parameter of the slow system satisfying*

$$\varepsilon_{\max}=\sup\{\bar{\varepsilon}:\varepsilon\in(0,\bar{\varepsilon})\}$$

when $\varepsilon\in(0,\bar{\varepsilon})$ *no eigenvalues of the DTM of the slow system have magnitudes greater than one.*

When the perturbation is added in Site 1, design the LTI SPM gauge as a second order minimum-phase fast system, which has a transfer function

$$L_{fast}(s)=\frac{\omega_0^2}{(\varepsilon s)^2+2\zeta\omega_0(\varepsilon s)+\omega_0^2}$$

Consider a Mathieu equation, which is a special case of Hill equation.

$$\ddot{y}+\left(a-2q\cos(2t)\right)y=0$$

Numerically calculate the DTM for Mathieu equation and check the eigenvalues of the DTM for the stability situation (Lemma 1). The stability diagram is shown by Fig. 2.

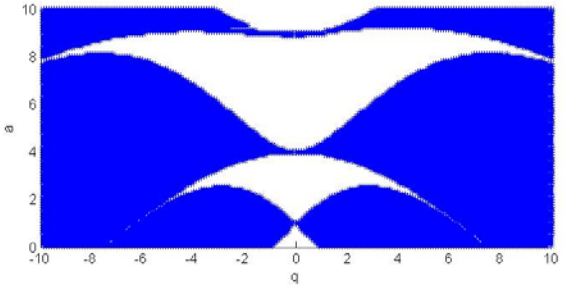


Fig. 2. Stability diagram for the Mathieu equation. Blank regions correspond to stable solutions and shaded regions to unstable solutions. The diagram is symmetrical about *a* axis.

In the expression of the slow system, matrix *L(t)* can be obtained by the iterative expression and the slow system is given by

$$A_{slow} = A_{11} - A_{12}(t)L(t) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ (2q\cos(2t) - a)\omega_0^2 & 0 \end{pmatrix} L(t)$$

$$L(t) = \sum_{j=0}^{\infty} \varepsilon^j \bar{L}_j(t)$$

$$\bar{L}_0 = A_{22}^{-1}A_{21},\ \bar{L}_{k+1} = A_{22}^{-1}\left[\dot{\bar{L}}_k + \bar{L}_k\left(A_{11} - A_{12}\bar{L}_k\right)\right]$$
$$k = 0,1,2,...$$

The parameters are: $a = 3, q = 1, \omega_0 = 2, \zeta = 0.2$.

Numerically calculate the DTM for the slow system and check the eigenvalues of the DTM for the stability situation (Lemma 1). The stability situation is different with different singular perturbation parameter ε, and the results are shown by Table. 1.

Table 1. Stability of the slow system with different singular perturbation parameters (Site 1)

No.	ε	Max magnitude of the eigenvalues of the DTM	Stability
1.	0.01	1.0	yes
2.	0.03	1.0	yes
3.	0.04	1.0	yes
4.	0.05	1.05	no
5.	0.06	1.06	no

According to Table. 1, the definition of SPM and Theorem 1, it is obtained that $\varepsilon_{\max} = 0.04$.

3.4. SPM Assessment of General Order LPTV Equations

The DTM, denoted by $\phi(\theta,0)$, is a key point for the stability margin assessment, and according to Floquet theory

$$x(t + m\theta) = \phi(\theta,0)^m x(t)$$

let $t = 0$ and $m = 0,1,2,...,n$ in the above equation, through simulation to know the value of $x(0), x(\theta),..., x(n\theta)$, and then the elements of $\phi(\theta,0)$ becomes an n^2 algebraic equation.

Due to the numerical methods for DTM, the algorithms of section 3.3 can be generalized to any order LPTV system to obtain the singular perturbation margin.

3.5. GGM Assessment of Hill Equations

Meissner Equation is a well-known Hill equation, when the function $\psi(t)$ is unit rectangular waveform in (3), which is shown by Fig. 3, and can be viewed as the pair of constant coefficient equations

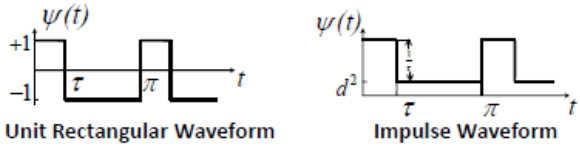


Fig. 3. Meissner Equation and Impulse Coefficient Hill Equation.

$$\ddot{x} + (a - 2q)x = 0, \quad 0 \le t < \tau$$
$$\ddot{x} + (a + 2q)x = 0, \quad \tau \le t < \pi$$

The DTM of Meissner equation is given by

$$\phi(\pi,0) = \begin{bmatrix} \cos d(\pi - \tau) & \frac{1}{d}\sin d(\pi - \tau) \\ -d\sin d(\pi - \tau) & \cos d(\pi - \tau) \end{bmatrix} \begin{bmatrix} \cos c\tau & \frac{1}{c}\sin c\tau \\ -c\sin c\tau & \cos c\tau \end{bmatrix},$$

where $d = \sqrt{(a + 2q)}, c = \sqrt{(a - 2q)}$.

Case 1. With constant regular perturbation parameter *k* in the closed loop, either Site 1 or Site 2 in Fig. 1, the DTM of the regularly perturbed system is given by the above $\phi(\pi,0)$, albeit the parameters *k* dependent. Denote

$$\bar{d} = \sqrt{(a + 2q)k},\ \bar{c} = \sqrt{(a - 2q)k}$$

The characteristic multiplier can be written as

$$\lambda_{1,2} = r/2 \pm \sqrt{(r/2)^2 - 1}$$

and

$$r = \text{trace}\left\{\phi_{pert,k}(\pi,0)\right\} = 2\cos\bar{d}(\pi - \tau)\cos\bar{c}\tau - \left(\frac{\bar{d}}{\bar{c}} + \frac{\bar{c}}{\bar{d}}\right)\sin\bar{d}(\pi - \tau)\sin\bar{c}\tau$$

By Lemma 1, the stability diagram for Meissner equation with GGM parameter *k* is shown by Fig. 4.

In Fig. 4, the blank area represents the stable solutions of Meissner equation with the parameter couple (q,a). When the nominal system with parameter couple (q_0,a_0), is perturbed by regular perturbation parameter k, then (q,a) will move along the line $a/q=3$, until the couple (q_f,a_f) first touches the boundary of the blank area, i. e. the regularly perturbed system becomes marginally stable, which is shown by the legend square. Then, the GGM $k_{\max}$ and $k_{\min}$ are obtained.

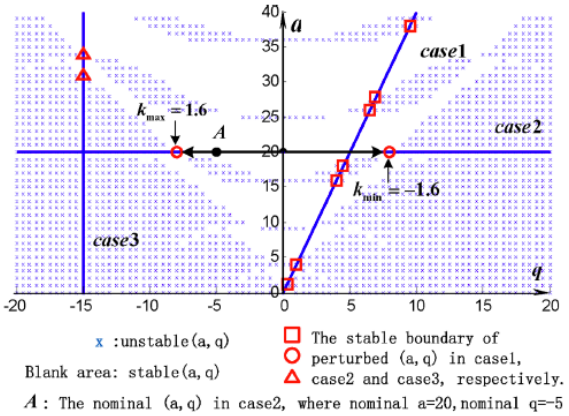


Fig. 4. Stability Diagram for Meissner Equation and GGM Assessment.

Case 2. When Meissner equation is regularly perturbed on parameter q. The parameters $\bar{d}$ and $\bar{c}$ in the DTM of the regularly perturbed system become $\bar{d}=\sqrt{a+2qk}$, $\bar{c}=\sqrt{a-2qk}$. In Fig. 4, when $a=20$, the legend circles show boundaries of stability solutions, and with $q=-1$ as the nominal system parameter (point A), the GGM of Meissner equation is $k_{\max}=1.6$, $k_{\min}=-1.6$.

Case 3. When Meissner equation is regularly perturbed on parameter a, the parameters $\bar{d}$ and $\bar{c}$ in the DTM become $\bar{d}=\sqrt{ak+2q}$, $\bar{c}=\sqrt{ak-2q}$. In Fig. 3, the legendtriangle represent the boundaries of stable solutions.

3.6. GGM Assessment of General Order Equations

Due to the numerical methods for DTM, the algorithems of section 3.5 can be generalized to any order LPTV system to obtain the GGM.

4. Conclusions

The problem of SPM and GGM assessment for LPTV systems is formulated. Chang transformation makes it possible to reduce the SPM analysis for Hill equations, which is essentially a stability problem of higher order LPTV systems due to the SPM gauge introduced dynamics, to second order LPTV systems. Based upon Floquet Theory, the SPM and GGM assessment methods for the second order and general order LPTV systems are established, respectively.

Acknowledgements

The author is grateful of the financial support from National Natural Science Foundation of China No. 61803010 and Aeronautical Science Foundation of China No.2018ZA51002.

References

[1]. M. Faraday, On a peculiar class of acoustical figures; and on certain forms assumed by groups of particles upon vibrating elastic surfaces, *Proceedings of the Royal Society of London,* 3, 1830, pp. 229-318.
[2]. E. Mathieu, Memoire sur le movement vibratoire d'une membrane de forme elliptique, *J. Math. Pure Appl*., 13, 1868, p. 137.
[3]. G. Floquet, Sur les equations differenctiales lineaires. *Ann. L'Ecole Normale Super*., 12, 1883, p. 47.
[4]. J. A. Richards, Analysis of Periodically Time- Varying Systems. *Springer*, 1983.
[5]. G. W. Hill, On the part of motion of the lunar perigee which is function of the mean motions of the sun and moon, *Acta Mathematica,* 8, 1886, pp. 1-36.
[6]. E. Meissner, Uber Schuttelerscheinungen im System mit periodisch veranderlicher Elastizitat, *Schweiz. Bauztg*., 72, 1918, pp. 95-98.
[7]. J. R. Carson, Notes on theory of modulation, *Proc. IRE*, 10, 1922, pp. 57-64.
[8]. H. Jeffreys, Approximate solutions of linear differential equations of second order, *Proc. Lon. Math. Soc*., 23, 1924, p. 428.
[9]. B. van der Pol, M. J. O. Strutt, On the stability of solutions of Mathieu's equation, *Philos. Mag.*, 5, 1928, pp. 18-39.
[10]. W. Y. Yan, R. R. Bitmead, Decentralized control techniques in periodically time-varying discrete-time control systems, *IEEE Transactions on Automatic Control*, 37, 10, 1992, pp. 1644-1648.
[11]. M. S. Chen, Y. Z. Chen, Static output feedback control for periodically time-varying systems, *IEEE Transactions on Automatic Control,* 1999, 44, 1, pp. 218-222.
[12]. S. Kalender, H. Flashner, A new approach for control of periodically time varying systems with application to spacecraft momentum unloading, in *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit,* 2006.
[13]. J. Zhang, C. Zhang, Robustness of discrete periodically time-varying control under LTI unstructured perturbations. *IEEE Transactions on Automatic Control,* 45, 7, 2000, pp. 1370-1374.
[14]. P. Brunovski, Controllability and linear closed-loop controls in linear periodic systems, *Differential Equations*, 6, 4, 1969, pp. 296-313.
[15]. Y. T. Hagiwara, Periodically time-varying controller synthesis for multiobjective control of discrete-time systems and analysis of achievable performance." *Systems & Control Letters,* 6, 2, 2011, pp. 296-313.

[16]. S. Bittanti, 30 years of periodic control-form analysis to design, in *Proceedings of the 3rd Asian Control Conference, Shanghai*, 2000, pp. 1253-1258.

[17]. X. Su, M. Lv, H. Shi, Q. Zhang, Stability analysis for linear time-varying periodically descriptor systems: a generalized Lyapunov approach, in *Proceedings of the IEEE World Congress on Intelligent Control & Automation*, 2006.

[18]. X. Xie, J. Lam, P. Li, A novel h∞ tracking control scheme for periodic piecewise time-varying systems, *Information Sciences*, 484, 2019, pp. 71-83.

[19]. R. C. L. F. Oliveira, M. C. De Oliveira, P. L. D. Peres, Special time-varying Lyapunov function for robust stability analysis of linear parameter varying systems with bounded parameter variation, *Iet Control Theory & Applications* 3, 10, 2009, pp. 1448-1461.

[20]. X. Yang, J. J. Zhu, A. S. Hodel, Singular perturbation margin and generalized gain margin for linear time invariant systems, *International Journal of Control*, 88, 1, 2015, pp.11-29.

[21]. X. Yang, J. J. Zhu, Singular perturbation margin and generalized gain margin for nonlinear time invariant systems, *International Journal of Control*, 89, 3, 2016, pp. 1-18.

[22]. P. V. Kokotovic, H. K. Khalil, J. O'Reilly, Singular perturbations methods in control: analysis and design, *Academic Press,* New York*,* 1986.

(020)

Obstacle Segmentation for Autonomous Guided Vehicles through Point Cloud Clustering with an RGB-D Camera

M. Pires [1], P. Couto [1,2], A. Santos [3] and V. Filipe [1,4]
[1] University of Trás-os-Montes e Alto Douro (UTAD), 5000-801 Vila Real, Portugal
[2] CITAB, Quinta de Prados, 5000-801 Vila Real, Portugal
[3] Active Space Technologies, Parque Industrial de Taveiro, Lote 12, 3045-508 Coimbra, Portugal
[4] INESC Technology and Science (INESC TEC), 4200-465 Porto, Portugal
Tel.: + 351259350356
E-mail: vfilipe@utad.pt

Summary: The movement of materials in industrial assembly lines can be done efficiently using Automated Guided Vehicles (AGVs). However, visual perception of industrial environments is complex due to the existence of many obstacles in pre-defined routes. With the INDTECH 4.0 project, we aim to develop an autonomous navigation system, allowing the AGV to detect and avoid obstacles based on the processing of depth data acquired with a frontal depth camera mounted on the AGV. Applying RANSAC algorithm and Euclidean Clustering to the 3D point clouds captured by the camera, we are able to isolate obstacles from the ground plane and separate them into clusters. The clusters give information about the location of obstacles with respect to the AGV position. In the experiments conducted in outdoor and indoor environments the results revealed that the method is effective, returning high percentages of detection for most tests.

Keywords: Autonomous navigation, Obstacle avoidance, AGV (Automated Guided Vehicle), ROS (Robot Operating System).

1. Introduction

Automation is one of the fastest developing fields of robotics. With the ever-growing interest in automating processes, both for efficiency or safety, the ability to provide robots with autonomous navigation capabilities is of paramount significance [1, 2].

Several sensor technologies and types of navigation systems are available for the autonomous navigation of mobile robots. A key part of mobile autonomous robots is the ability to detect objects in their path and to be able to navigate freely with minimum risk of collision. Because of this, in industrial environments, AGVs are usually provided with laser scanners to prevent them from running into people or things. However, the effectiveness of these systems can be affected by adverse weather conditions such as direct sunlight, rain, snow, or fog. Even considering the use of laser scanners prepared to operate outdoors, other vehicles (e.g., forklifts) may not be detected, thus leading to collisions. As an alternative, 3D point clouds captured with a depth camera could be used in these complex conditions.

There are different ways to perform object detection using 3D point clouds. Most solutions tackle this problem by using neural networks adapted to this kind of data. An important paper in this type of methods presents the VoxelNet algorithm [3] that works with sparse 3D point clouds, without the need of manual feature engineering. Further studies lead to the development of algorithms like the one presented in [4] that uses RGB images to generate detection proposals for 3D detection and PointRCNN [5] that uses a bottom-up 3D proposal generation method.

Ground plane and obstacle detection are essential tasks for collision free navigation of autonomous mobile robots [6]. One of the most common methods to detect the ground plane in a scene is based on processing depth information. After ground plane extraction, objects can be detected in tridimensional data by using point cloud segmentation, which is the method explored in this work. An example of this kind of methodology is presented in [7] where the authors present three different point cloud segmentation methods and compare their performance with the k-means segmentation algorithm. In [8], the authors explored the idea of combining different segmentation methods and post-processing algorithms for point cloud segmentation of aerial images of urban areas. A 3D point cloud segmentation survey is presented in [9] where a comparative analysis on various state of the art point cloud segmentation methods is made.

On the project "INDTECH 4.0", developed by a consortium of companies and universities, one of the work packages aims to study and develop autonomous navigation solutions, based on vectors of flexibility, collaboration, adaptability and modularity, to be implemented in the car assembly industry. In this context, an Intel RealSense D435i camera was mounted on the front of an AGV to collect depth data. The 3D point clouds acquired are continuously processed to ensure obstacle avoidance, which is an essential task for real-time outdoor autonomous navigation of AGVs.

In this paper, we present a method to detect potential obstacles and to measure their relative position to the robot using point cloud data provided by the depth camera. This way, the AGV is guided through vision technology, without any previous knowledge of the

environment that can be static or dynamic, indoor or outdoor, in a process called natural navigation.

The remainder of this paper is structured as follows. Section 2 is a description of our methodology, detailing the various processing phases of our method. Section 3 presents and analyses the results according to various metrics relevant for this kind of algorithm. Section 4 presents the conclusions of this paper and suggestions of future work.

2. Proposed Methodology

A method to detect obstacles and their locations by means of 3D point clouds provided by the Intel RealSense D435i depth camera is proposed. The method processing sequence is shown in Fig. 1.

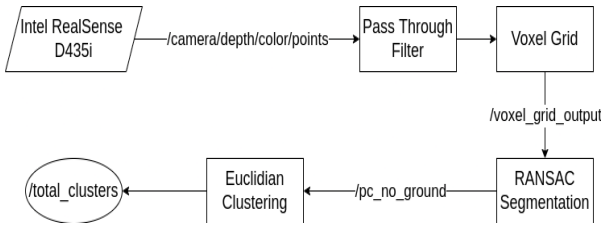


Fig. 1. Processing Flowchart.

The method consists in a ROS (Robot Operating System) node that subscribes to the point cloud provided directly by the depth camera in the *"/camera/depth/color/points*" ROS topic.

This point cloud is subjected to three different stages of processing, resulting in the separation of obstacles into point clusters. Each of these clusters represents a different obstacle and, through the clusters, we can obtain the position of each obstacle relative to the AGV. After processing, all detected clusters are merged in a point cloud that is published in the "*total_clusters*" ROS topic.

In addition to ROS, functionalities provided by the Point Cloud Library (PCL) [10] have been used in all processing steps.

2.1 Pass-through Filter and Voxel Grid

The first processing stage consists in applying a pass-through filter. This filter is used to restrict the point cloud limits to a relevant distance window.

The ultimate use of this algorithm is to apply it in a navigation system of a low speed AGV for load transportation. Thus, it is possible to limit the computational effort while conserving a high relevance of the results since, due to the low speed, we do not need to detect obstacles that are far away from the robot.

Another processing tool used to save computational time is the voxel grid to down sample the input point cloud. This filter divides the point cloud into voxels with user-defined size. The points present in each voxel are approximated by their centroid. This process results in a point cloud with less points, but that still describes the original surface accurately.

In our tests, we defined the limits of the distance window in the interval [1.5, 5] m along the z-axis for the pass-through filter and a voxel edge length of 0.07 m for the voxel grid.

2.2 RANSAC Segmentation

To successfully identify obstacles in the camera field of view, it is necessary to segment the volume of the point cloud that corresponds to the ground plane. This is done to deliver a higher degree of separation between objects, facilitating the process of grouping points into clusters. For this purpose, the RANSAC (Random Sample Consensus) algorithm is used [11].

RANSAC is an iterative algorithm for fitting mathematical models to experimental data. This method takes a number of arbitrary points from the experimental set and instantiates a model using those points. Having a candidate model defined, the number of points that are within a defined tolerance to the plane is counted (inliers). When the number of inliers exceeds a threshold, the candidate model is chosen. If the number of inliers is too low, the process is repeated until a plane is identified.

The PCL implementation of RANSAC further includes the option of constraining the plane fit to a specified axis within an angular tolerance. This helps in identifying the ground plane by restricting the detection to horizontal planes.

Once the ground plane is identified, inliers are removed, and the resulting point cloud solely contains objects that may represent obstacles.

2.3. Euclidean Clustering

After removing the ground plane, Euclidean Clustering is carried out to identify the various clusters of points present in the point cloud.

Euclidean Clustering consists in separating sets of points according to the Euclidean distance between those points. If two points are within a predefined distance threshold, they are considered to be in the same cluster. However, if that distance is higher than the threshold, the points are segregated into different clusters.

The PCL implementation of Euclidean Clustering also contains the ability to define the minimum and maximum cluster sizes to mitigate the detection of undesirable sets of points.

Once all the clusters are defined, we can visualise them and measure the position to each obstacle. For the interest of avoiding AGV collisions, our method returns the coordinates of the closest point of each cluster.

3. Results

Throughout this work, various tests have been carried out to optimise this algorithm. These tests were

performed in the premises of Active Space Technologies and consisted in recording ROS bags using the D435i depth camera with various types of obstacles located at different distances. We tested the algorithm for both static and mobile obstacles and in indoor and outdoor environments.

To represent static obstacles, chairs placed at 1.8 m, 2.5 m and 4.5 m were used. The distance to the chairs was measured taking the closest part of the chair into account. This set up allows us to evaluate the algorithm performance for different distances across the defined depth window ([1.5, 5] m). For each of the three distances, a ROS bag is recorded with two chairs side-by-side, indoors and outdoors.

Another test was made with three chairs, one at each distance mentioned above, to assess simultaneous detection of various obstacles at different distances. The final test with chairs consisted in moving the camera perpendicularly to the three chairs, in the direction of the chair furthest away. This test simulates object detection of static obstacles with a circulating AGV.

For mobile obstacles, we performed tests in which people moved in front of the camera in various directions. A final test was made in which the camera was also moving among wandering people, simulating a more complex scenario that would be common for an industrial AGV.

Every test mentioned above has been done indoors and repeated outdoors.

This batch of tests enables us to get a sense of the algorithm's performance in different conditions considered relevant for the project. Figure 2 represents the typical outputs of this algorithm concerning the three major stages of processing.

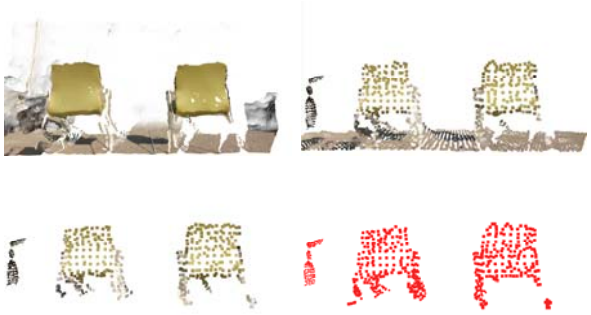


Fig. 2. Results for the four stages of processing. From top-left to bottom-right: original point cloud, down sampled and range limited point cloud, point cloud after ground plane removing and point cloud with detected clusters (two chairs).

3.1 Object Detection Effectiveness

To measure the object detection effectiveness, we counted the number of times that each object is detected and compared it with the number of frames of each test. This was done using the ROS bags corresponding to the placement of chairs side-by-side at various distances and using indoor and outdoor measurements.

For each ROS bag, the point from each cluster closest to the camera is obtained and published in a ROS topic (*cluster_closest_point*) for analysis. This topic contains the xyz coordinates of the closest point of each object. The total number of point clouds analysed by the algorithm is also counted.

A python script is used to subscribe to the (*cluster_closest_point*) topic and the information is assembled and plotted to a 3D graph. This graph gives us a sense of the position of obstacles in the corresponding test.

In addition to visualising the data, a clustering algorithm is applied to this set of points, originating clusters of points that correspond to detections of the same object. Using the clusters of closest points, we count the number of points that represent each obstacle detected by the camera. That value determines how many times each obstacle is detected.

By comparing these values with the number of point clouds processed by our object detection algorithm we are able to calculate the effectiveness of the object detection.

Each ROS bag has a duration of approximately 15 seconds, which at □30 fps represents a total of □450 frames for each test run. The process described above is repeated 10 times for each ROS bag and the results are presented in Tables 1 and 2.

Table 1. Indoor detection effectiveness.

Distance (m)	Indoor Detection Effectiveness [%]
1.8	92.72 ± 5.04
2.5	70.59 ± 6.98
4.5	51.02 ± 0.70

Table 2. Outdoor detection effectiveness.

Distance (m)	Outdoor Detection Effectiveness [%]
1.8	99.05 ± 0.25
2.5	99.12 ± 0.86
4.5	89.22 ± 1.51

From these results we conclude that the method is very effective, returning high percentages of detection for most tests. Also, the percentages of detection of both objects are very similar for all distances.

It is seen that the detection rates drop more significantly for indoor obstacles that sit farther away from the camera. This is due to the poorer luminosity and the fact that for farther objects, the point cloud includes parts of the ceiling, extending the dimensions of the point cloud and making the object detection process more difficult. In this case, the performance bottleneck is mostly due to the increased size of the point cloud, hence explaining why, for outdoor conditions, an efficacy drop is not as drastic.

We can also conclude that, albeit detection rates vary according to environmental conditions, there is not a variance in detection efficacy between the two obstacles.

Thus, detection success decreases with distance to the obstacle, but does not appear to change for obstacles at the same distance. The percentage of obstacle detection is also significantly more stable in outdoor conditions, as can be concluded by the lower standard deviation values.

3.2 Distance Measurements

A major requirement of this work is not only to detect obstacles in front of the robot but also to measure their position relative to the camera.

Using the same ROS bags, with two chairs side-by-side at various distances, we analysed the distance measurements obtained with our algorithm.

The results in Tables 3 and 4 show the mean and standard deviation values for the distances obtained in each test, for indoor and outdoor conditions.

Table 3. Indoor distance measurements analysis.

Indoor Measurements		
Distance (m)	**Average Distance (m)**	**Standard Deviation (m)**
1.8	1.85	0.03
2.5	2.60	0.09
4.5	5.01	0.11

Table 4. Outdoor distance measurements analysis.

Outdoor Measurements		
Distance (m)	**Average Distance (m)**	**Standard Deviation (m)**
1.8	1.91	0.03
2.5	2.56	0.13
4.5	4.95	0.20

For distance measurements, we are able to obtain reasonable values, with an expected drop in accuracy for farther objects. These measurements are entirely dependent on the accuracy of the Intel RealSense D435i depth camera, and, as demonstrated, we can rely on the good accuracy provided by the depth sensor for our application.

3.3 Moving Object Detection

For moving obstacles, the performance of the algorithm shows favourable results, although the analysis of these tests has been made merely by visual inspection.

The proposed method is able to segment and follow moving obstacles smoothly with high degree of accuracy and with high frame rate without apparent fail in detection.

Figure 3 shows two frames from the mobile AGV/wandering person test where it is clear that the person is segmented accurately and there are no false positives. The detection remains accurate during the whole test regardless of the movement of the AGV and of the person.

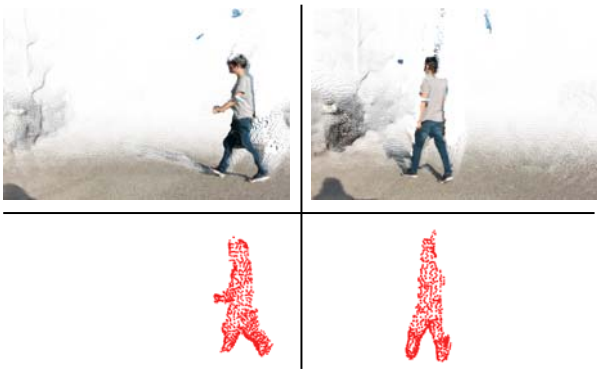


Fig. 3. Moving obstacle detection (wandering pedestrian). In this test, both the person and the AGV are moving. Top: original point clouds; bottom: detected clusters

3.4 Frame Rates

Since this algorithm is intended for real-time obstacle avoidance, it is also important to analyse the frame rate.

Using a computer with an Intel(R) Core(TM) i7-6700HQ CPU @ 2.60 GHz, we do not register significant frame rate drops relatively to the native frame rate of the point cloud (□30 fps), with the maximum frame rate drop being of □58 % in one of the tests, but that still results in a rate of □12 fps for obstacle detection. That being said, most tests performed had a frame drop inferior to 10 % relative to the native frame rate of the camera, with the average frame drop in all tests being of 18 %. These values mean that our method is applicable in a slow moving AGV for real-time object detection.

4. Conclusions

In this work, an object detection algorithm using point cloud data from an Intel RealSense D435i is proposed. This algorithm is meant to be applied in an autonomous navigation system with obstacle avoidance capabilities for AGVs in dynamic industrial environments.

The presented results show this method can detect obstacles and measure their position relative to the AGV in various environmental conditions. This is very important in a SLAM (Simultaneous Localization and Mapping) context to enable robust autonomous navigation of robots. This method is also able to function in real-time, which is crucial for autonomous navigation.

The algorithm detects objects with high efficacy and returns results with inconsequential frame drops from the camera's native frame rate. The resulting frame rates of detection are adequate for a slow moving AGV, as intended for our application.

The obstacle position measurements returned by our method also present high degrees of precision that improve with illumination and as obstacles get closer to

the camera, characteristics that are innate to the Intel RealSense D435i depth camera.

Further developments of this algorithm may provide the capability to create a system that generates odometry measurements using depth camera data, such as implementing a visual odometry system for SLAM. Such algorithms are useful to perform sensor fusion with an inertial measurement unit or can even replace this kind of sensors. Finally, the methodology can be extended to process RGB images, acquired simultaneously by the Intel camera, for object detection and classification.

Acknowledgements

This work was funded by Project "INDTECH 4.0 – new technologies for smart manufacturing", POCI- 01-0247-FEDER-026653, financed by the European Regional Development Fund (ERDF), through the COMPETE 2020 - Competitiveness and Internationalization Operational Program (POCI). The work is also funded by national funds from FCT - Portuguese Foundation for Science and Technology, under the project UIDB/04033/2020.

References

[1]. Pires M. A., Natural navigation solutions for AMRs and AGVs using depth cameras, MSc diss., *Universidade de Coimbra*, 2021.

[2]. Lynch L., Newe T., Clifford J., Coleman J., Walsh J. and Toal D. Automated ground vehicle (AGV) and sensor technologies - a review, in *Proceedings of the 12th IEEE International Conference on Sensing Technology (ICST'18)*, 2018, pp. 347-352.

[3]. Zhou Y. and Tuzel O. Voxelnet: End-to-end learning for point cloud based 3D object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2018, pp. 4490-4499.

[4]. Qi C. R., Liu W., Wu C., Su H. and Guibas L. J., Frustum pointnets for 3D object detection from RGB-D data, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2018, pp. 918-927.

[5]. Shi S., Wang X. and Li H. PointRCNN: 3D object proposal generation and detection from point cloud, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2019, pp. 770-779.

[6]. Kircali D. and Tek F. B. Ground plane detection using an RGB-D sensor, in *Proceedings of the 29th International Symposium on Computer and Information Sciences*, 2014, pp. 69-77.

[7]. Habermann D., Hata A., Wolf D. and Osório F. S. 3D point clouds segmentation for autonomous ground vehicle, in *Proceedings of the III IEEE Brazilian Symposium on Computing Systems Engineering*, 2013, pp. 143-148.

[8]. Vosselman G. Point cloud segmentation for urban scene classification, *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, Vol. 1, No. 257-262, 2013, p. 1.

[9]. Nguyen A. and Le B. 3D point cloud segmentation: A survey, in *Proceedings of the 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM),* 2013, pp. 225-230.

[10]. Rusu R. B. and Cousins, S. 3D is here: Point cloud library (PCL), in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2011, pp. 1–4.

[11]. Fischler M. A. and Bolles R. C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, Vol. 24, No. 6, 1981, pp. 381–395.

(023)

Competency-based Education of the Mechatronics Engineer in the Transition from Manufacturing 3.0 to Industry 4.0

Eusebio Jiménez López [1], Francisco Javier Ochoa Estrella [2], Gabriel Luna-Sandoval [3], Flavio Muñoz Beltrán [2], Francisco Cuenca Jiménez [4] and Marco Antonio Maciel Monteón [3]
[1] Universidad Tecnológica del Sur de Sonora-ULSA Noroeste, CIAAM,
Dr. Norman E. Borlaug Km. 14. Cd Obregón, Sonora, Mexico
[2] TecNM/ Instituto Tecnológico Superior de Cajeme,
Carretera Internacional a Nogales Km. 2 s/n, Cd Obregón, Sonora, Mexico
[3] Universidad Estatal de Sonora, Carretera,
San Luis Río Colorado - Sonoyta km. 6.5, Parque Industrial, San Luis Río Colorado, México
[4] Universidad Nacional Autónoma de México, Av. Universidad 3004, Col, Copilco Universidad,
Coyoacán, 04510, Ciudad de México, CDMX, Mexico
Tel.: + 52 (644) 414-8687, fax: + 52 (644) 414-8687
E-mail: ejimenezl@msn.com

Summary: The changes in industrial processes that are being introduced by the new industrial revolution called "manufacturing 4.0" have broad repercussions on engineering education. Engineering education depends on the competencies required by the industrial sector; however, the technological transition is not constant and homogeneous in companies or in productive regions, especially in developing countries such as Mexico, so there are still many productive systems that operate under the philosophy of manufacturing 3.0. It is necessary to design an adequate profile and determine an appropriate educational model for the mechatronic engineer in the context of the transition between manufacturing 3.0 and industry 4.0, which allows him/her to have the necessary competences to face the current and future challenges of the companies. This paper proposes Competency Based Education (CBE) as the basis for training mechatronic engineers.

Keywords: Mechatronics, Industry 4.0, Education, Manufacturing 3.0, Competences, Engineering.

1. Introduction

Industry 4.0 not only presents a radical challenge to companies but also extends to education, especially to engineering education. The incursion of disruptive technologies (Big Data, the Internet of Things, Artificial Intelligence, Cloud Computing, Machine Learning, Augmented Reality, etc.) in industrial processes motivates the training of highly qualified and intelligent professionals who are able to be trained with multiple skills, such as the ability to work in different business units, strategic thinking, computer skills, centralization capacity, general aptitude, leadership, culture protection and information security, among others [1].

Similarly, the engineer must acquire (core or transversal) process and content competencies, social, technical, resource management and systems competencies. These include the ability to interact with machines, data analysis, planning and programming; negotiation, collaboration, team building, flexibility, rapid learning ability, problem detection and resolution, and autonomous initiative among others [2].

However, although numerous authors propose changes in engineering education due to the accelerated incursion of Industry 4. 0 the skills and competencies suggested in engineering education are difficult and complex to acquire, mainly in developing countries due to numerous factors, such as: the lack of an effective relationship between companies and universities, the existence of obsolete educational technology, teachers who are not updated and do not have a good command of computer tools, unconsolidated educational models, the weakness of national value chains and the lack of an effective public technological policy, among others.

Although those countries that are at the forefront of the implementation of Industry 4.0 can afford a better evolution in education, in developing countries such as Mexico, this transition is not continuous or homogeneous, so it is necessary to study and propose appropriate educational methods or approaches to train the engineer according to the local reality (mainly focused on Manufacturing 3.0) and according to the new challenges posed by Industry 4.0.

The competency-based approach can be a way to train new engineers, especially mechatronic engineers, since it fosters comprehensive training, flexibility and self-management, promotes active learning, develops technical and social skills and encourages engineers to solve problems in complex situations, among other relevant characteristics.

2. Technological Transition: Technical Aspects

The current technological transition implies a new challenge for mechatronic engineering as machinery, processes and factories are evolving from traditional automation to intelligent industry. Today, concepts such

as manufacturing cell or CIM (Computer Integrated Manufacturing) have evolved to cyber-physical systems and digital twins, integrated with artificial intelligence, advanced simulation and autonomous and cooperative robots. Similarly, the design and manufacture of various parts and components are now performed by advanced software and simulation, as well as by additive manufacturing systems. Information is stored and processed by Big Data and cloud computing techniques, and protected by cybersecurity methods. Training is aided by virtual and augmented reality techniques. To these challenges must be added the great improvements in interconnection and communication provided by the Internet of Things and the improvement of IT integration systems in companies. All these technological improvements bring great challenges only on the technical side for mechatronics engineers.

With regard to the various companies, especially in Mexico, the technological transition (from manufacturing 3.0 to industry 4.0) has become a serious problem, as they face the challenge of deciding whether a technological upgrade (almost total change of production systems) that implies a considerable economic investment or a technology reconversion that involves a substantial improvement to the production systems they already have and, therefore, less economic investment, is relevant. This type of analysis and decision making undoubtedly involves mechatronics engineers, so it will be necessary to use reverse engineering techniques for the evaluation of the technologies available to the companies if a technological reconversion criterion is to be applied. Likewise, it will be necessary to be involved in the technology transfer process, which implies knowledge of the purchase, installation and start-up of the systems acquired and knowledge of intellectual property for the study of patents and licenses in the event that a technological update is chosen.

The transition between two industrial revolutions generates an important series of scientific and technological challenges, not only for companies, but also for governments, the education sector, trade and strategic competition between regions or countries.

3. Mechatronics Education in Industry 4.0

The synergic integration of Mechanics, Control, Electronics and Computing in the development of products, machines, processes and production systems has always had great challenges for the mechatronics engineer, so now we must consider that to the traditional challenges of mechatronics we must add various requirements and technical requirements demanded by industry 4.0, especially the various techniques of artificial intelligence, digitization and the internet of things.

The fourth industrial revolution requires multidisciplinary knowledge (as in the case of Mechatronics) and expertise. Multidisciplinary skills are mandatory and real-time collaboration will be more important than ever for success in the manufacturing and service industries [3]. One of the crucial infrastructures that must be valued in the training of the mechatronics engineer is computational as it is constantly changing and has a wide influence on the technologies supporting Industry 4.0, so constant updating of curricula can help the computational infrastructure not to be overtaken by changes in processing speed or become obsolete in a short period of time. In fact, mechatronics has gone from being purely associated with essentially autonomous systems, such as robots, to providing the intelligent objects and systems that are the building blocks of cyber-physical systems and, therefore, of systems based on the Internet of Things and the cloud, and all this has been made possible due to the evolution of computer technologies, so the importance of these technologies in the education of the mechatronics engineer is transcendental.

Mechatronics was born at the height of the third industrial revolution and, consequently, together with computing, informatics and robotics, it brought a significant technological impact to the industrial world over a period of at least 50 years. Today, the education of the mechatronics engineer faces the major challenges of transforming CIM systems (core of manufacturing 3.0) to modern cyber-physical systems (CPS) that integrate various Digital Twins (DT) and are the basis of Industry 4.0.

Cyber-physical entities are described as any entity composed of physical and cyber-physical elements that interact autonomously with each other, with or without human supervision [4]. The main characteristic of a Cyber-Physical System is its ability to merge the physical and virtual worlds, in particular, through its embedded software.

One of the technologies on which cyber-physical systems are supported is simulation, described as the imitation of the behavior of the properties of a system. One of the most important concepts within CPS are Digital Twins that use simulation as part of their processes.

A digital twin is defined as "a formal digital representation of some asset, process, or system that captures attributes and behaviors of that entity suitable for communication, storage, interpretation, or processing within a given context" by the Industrial Internet Consortium [5].

In general CPS are understood from the point of view of integration that combines systems or processes that occur in reality with computational processes, while Digital Twins or digital copy are used in real-time improvement and optimization processes. This implies that the Digital Twin relates to its physical counterpart in two ways: 1) It receives from it information generally from sensors and 2) The valuable or processed information from the digital copy is sent to the physical part to achieve specific objectives. Cyber-physical systems and Digital Twins are two of the central concepts that must necessarily be taken into account for the design or redesign of mechatronics career curricula and for the design of multidisciplinary competencies.

On the other hand, the most significant transformation related to the way products are

manufactured is digitalization. The fourth industrial revolution aims to optimize the computerized third industrial revolution (Industry 3.0). This requires the development of intelligent equipment with access to more data, thus becoming more efficient and productive by making decisions in real time [6]. In this sense, the different educational models around mechatronics should consider as a priority two fundamental aspects for the training of engineers: 1) The formal study of the various algorithms and tools of Artificial Intelligence and 2) The conformation of intelligent teaching and learning environments. There is a range of tools that support Artificial Intelligence, such as Deep Learning, Machine Learning, Neural Networks, Fuzzy Logic, Genetic Algorithms, Natural Language Processing, Knowledge Engineering, Expert Systems, Data Mining and others. These tools will be commonly used in industries in the coming years.

Faced with the constant challenges of manufacturing 4.0, mechatronics and mechatronics education must reinvent themselves. To this end, designers of mechatronic systems of all varieties and types must be aware of a set of fundamental principles if they are to ensure success. Some of these principles are as follows: technology; complexity; operational software; hardware; reliable and rechargeable power supplies; innocent human error; connectivity; privacy, dependency, ubiquity, and hybrid society, among others [7].

4. Competency-Based Approach

The training of the new engineer not only demands an update of the curriculum but also requires universities to have a strong connection with companies and the adoption of new educational approaches, such as Competency-Based Education (CBE) and active learning methodologies, as well as a rational contextualization of the conditions in which the technological transition is taking place in each country and in each productive region. Competencies are understood as a combination of skills, abilities, knowledge, aptitudes, attitudes and values to perform well in complex and authentic contexts.

Competency-Based Education is both goal-oriented and outcome-oriented. It is an approach embedded in adult learning theory, which posits that adult learners are more likely to engage in learning that is focused on a specific goal or outcome. This means that both teaching and learning in CBE are oriented toward the development of an explicitly stated and described skill, and measured by observing how the learner performs this skill [8]. A Competency-Based Education approach focuses on the development and demonstration of competencies by a learner. Thus, a CBE perspective puts the learner at the center of the model and provides a comprehensive education oriented to the competencies demanded in the world. CBE has strong implications in the teaching-learning process. This approach alters the curricular organization, the way content is taught, the learning environments, the assessment instruments and the evidence of the level of development of the defined set of competencies [9].

On the other hand, with the purpose of making students access active and meaningful learning, active methodologies were introduced. Prince and Felder compiled six most innovative teaching methods, namely, inquiry learning, problem-based learning, project-based learning, case-based learning, discovery learning, and just-in-time teaching. These teaching methods are learner-centered and motivate students with real-world problems, cases, and observations, which triggers students desire to learn knowledge [10].

There are several proposed taxonomies for engineering education. One of them consists of seven categories [11]: 1) Engineering specific work includes general engineering work, specialized engineering work, and engineering combined with other work, 2) Non-engineering work specific work includes general, supervision and leadership, and management and administration, 3) Communication 4), Interpersonal interactions include personal and teamwork, 5) Personal dispositions include general and discipline 6), Adaptive dispositions include personal development, lifelong learning and change management, and 7) Advanced dispositions include achievement orientation, impact and influence, conceptual thinking, analytical thinking, initiative, self-confidence, interpersonal understanding, concern for order, information seeking, teamwork and cooperation, experience and service orientation.

The competency model may be ideal for the comprehensive training of the mechatronics engineer whose professional activity will be to solve problems in the transition from manufacturing 3.0 to industry 4.0, as this model offers students the possibility to hone their ability to recognize, build and manage their skills. It allows students to evaluate and improve their performance, solve difficulties, plan innovative strategies and interpret diverse situations in an ethical and responsible manner. The competencies integrate a response to society, the culture of quality, globalization and business competitiveness.

Although in Mexico, as in other Latin American countries, CBE has been implemented for two decades, this educational approach has not been homogenized nor has it been effective in engineering education. Some universities in other countries have been applying the competency-based approach and have even adopted active methodologies such as Project Based Learning [12] and Learning Factories [13] to train engineers in the current industrial context. It is necessary that CBE be established in an integral way in engineering careers in Mexico and in other Latin American countries, since in this way it will be possible to design effective curricula and learning strategies, as is the case of Mechatronics Engineering, in accordance with the demands of the new industrial revolution.

5. Conclusions

The current profile of the mechatronic engineer must be designed taking into account the context of each

country and the implications generated in the transition from manufacturing 3.0 to industry 4.0. The objective of their training should be the acquisition of competencies and technical and social skills that allow them to solve problems in both technological approaches.

The CBE is a good option for the formation of the mechatronic engineer in the context of the current technological transition, because it is possible to design and evaluate the necessary competencies in engineering students for each requirement needed by the industry. It is necessary to incorporate in the curricula and study plans of the mechatronics career special topics of cyber-physical systems, digital twins and artificial intelligence with the aim of making inroads into the base technologies of industry 4.0, and promote the development of competencies and skills in programming and computing in students, since this topic is of great relevance and importance in the current context of technological applications.

Mechatronics as an area of technological application and education of the same, face the difficulties and challenges involved in transforming CIM systems that represent the technologies of manufacturing 3.0 to cyber-physical systems that make up the core of industry 4.0.

References

[1]. Kaur R., Awasthi A., Grzybowska K. Evaluation of Key Skills Supporting Industry 4.0—A Review of Literature and Practice, in Grzybowska K., Awasthi A., Sawhney R. (Eds.), Sustainable Logistics and Production in Industry 4.0. EcoProduction (Environmental Issues in Logistics and Manufacturing). *Springer, Cham*. 2020.

[2]. BRICS skill development working group: Skills development for Industry 4.0. 2018.

[3]. Stankovski S., Ostojić G., Zhang X., Baranovski I., Tegeltija S., and Sabolč Horvat. Mechatronics, Identification Tehnology, Industry 4.0 and Education. in *Proceedings of the 18th IEEE International Symposium INFOTEH-JAHORINA*, 20-22 March, 2019.

[4]. DeSmit Z., Elhabashy A. E., Lee J. Wells L. and Camelio J. A., Cyber-Physical Vulnerability Assessment in Manufacturing Systems, *Procedia Manufacturing*, 5, 2016, pp. 1060-1074.

[5]. Jian D., Tian M., Qing Z., Zhen L. y Ji‐Y., Design and application of digital twin system for the blade‐rotor test rig, *Journal of Intelligent Manufacturing*, 2021.

[6]. Souza, M. L. H., C. Andre da Costa, G. de Oliveira Ramos, Rodrigo da Rosa Righi, A survey on decision-making based on system reliability in the context of Industry 4.0, *Journal of Manufacturing Systems*, 56, 2020, pp. 133–156.

[7]. Russell D., Reinventing Mechatronics - Final Thoughts, in Reinventing Mechatronics, Xiu-Tian Yan, David Bradley, David Russell and Philip Moore (Eds.), *Springer*, 2020, pp.179-186.

[8]. Mace K. L., Welch C. E., The future of health professions education: considerations for competency-based education in athletic training, *Athletic Training Education Journal,* 14, 03, 2019, pp. 215–222.

[9]. Félix L. C., Rendon A. E. and Nieto J. M., Challenge-based learning: an I-semester for experiential learning in Mechatronics Engineering, *International Journal on Interactive Design and Manufacturing,* 13, 2019, pp. 1367–1383.

[10]. Prince M. J. and Felder R. M., Inductive teaching and learning methods: Definitions, comparisons, and research bases*, Journal of Engineering Education*, Vol. 95, 2006, pp. 123-138.

[11]. Woollacott L. C., Validating the CDIO syllabus for engineering education using the taxonomy of engineering competencies. *European Journal of Engineering Education*, 34, 6, 2009, pp. 545–559.

[12]. Atmojo U. D., Project-based Learning at Aalto Factory of the Future - A Flexible Production-based Industrie 4.0 Learning Factory, in *Proceedings of the 11th Conference on Learning Factories (CLF'2021)*, 1–2 July, 2021, Graz, Austria, 2021.

[13]. Montoya A., Guarín A. and Mora J., Advantages of Learning Factories for Production Planning Based on Shop Floor Simulation: A Step towards Smart Factories in Industry 4.0, in Industry 4.0 - Impact on Intelligent Logistics and Manufacturing, Tamás Bányai (Ed.), Chapter 2, *IntechOpen*, 2020.

2nd IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 (ARCI' 2022), 2-3 February 2022, Andorra la Vella, Andorra

(024)

Simulation of a Collision and Obstacle Avoidance Algorithm for Cooperative Industrial Autonomous Vehicles

J. Grosset [1,2], A.-J. Fougères [2], M. Djoko-Kouam [2], C. Couturier [1] and J.-M. Bonnin [1]
[1] IMT Atlantique, IRISA, Rennes, France
[2] ECAM Rennes, Louis de Broglie, Campus de Ker Lann, Bruz, Rennes 35091, France
E-mail: juliette.grosset@ecam-rennes.fr

Summary: Industry 4.0 leads to a strong digitalization of industrial processes, but also a significant increase in communication and cooperation between the machines that make it up. This is the case with intelligent autonomous vehicles (IAVs) and other cooperative mobile robots which are multiplying in factories, often in the form of fleets of vehicles, and whose intelligence and autonomy are increasing. The implantation and deployment of IAVs fleets raises several challenges: acceptability by employees, vehicle location, traffic fluidity, vehicle perception of changing environments. Simulation offers a good framework for studying solutions for these different challenges. Thus, we propose in this paper the extension of a collision detection algorithm to deal with the obstacle avoidance issue. The conclusive simulation will allow us to experiment in real conditions.

Keywords: Simulation, Autonomous vehicles, Cooperative collision avoidance.

1. Introduction

One of the challenges of Industry 4.0, is to determine and optimize the flow of data, products and materials in manufacturing companies. To realize these challenges, many solutions have been defined [1] such as the utilization of automated guided vehicles (AGVs), intelligent autonomous vehicles (IAVs) and other cooperative autonomous robots. The implantation and deployment of intelligent autonomous vehicle fleets raises several challenges: acceptability by employees, vehicle location, traffic fluidity, vehicle perception of changing environments (dynamic). In this context, autonomy is reduced to predetermined trajectories.

To improve the autonomy of a fleet, one way it to develop a collective intelligence to make the behaviors of vehicles adaptive. We will focus on a class of problems faced by IAVs related to collision and obstacle avoidance. This occurs when two vehicles need to cross an intersection at the same time, known as a *deadlock situation*. But also, when obstacles are present in the aisles and need to be avoided by the vehicles safely.

In this paper, we will propose an enhancement to the collision avoidance algorithm experimented in the study [2], in order to handle the problem of obstacle avoidance.

2. State of the Art

For the consumer sector, autonomy of vehicles is well determined with 6 levels of autonomy [3]. However, no such scale exists in the industrial context. Our objective is to improve the IAV autonomy integrated in a fleet based on collective intelligent strategies. Among the problems to be solved to make IAVs more autonomous, we can particularly note the location and positioning of vehicles [4], as well as the avoidance of other vehicles or obstacles [5]. This last problem can be solved by the cooperation between IAVs [2, 6].

The capacity to exchange information between the different IAVs of a fleet should improve this autonomy [7, 8]. The study [2] proposed a cooperation strategy based on the exchange of messages to determine the priority to pass an intersection between IAVs. The solution requires the vehicle to know its own position, and to be able to communicate with the other vehicles.

The collision avoidance algorithm presented in [2] allows IAVs to communicate and cooperate using different types of messages.

The communication between IAVs is done with 3 different types of messages:

- Hello_msg: message to indicate its presence with its position
- Coop_msg: message before an intersection area to determine priority.
- Ack_msg: message to confirm receipt of a Coop_msg.

The European Institute of Telecommunications Standards (ETSI) has published a standard for this kind of cooperative awareness messages (CAM) (ETSI EN 302 637-2 standard [9, 10]) and decentralized environmental notification message (DENM) (ETSI EN 302 637-3 standard [11]). These specifications and messages are approved and constitute building blocks for the safety of future intelligent transport systems (ITS) [12]. The purpose of CAM messages is similar to Hello_msg in [2]. They make it possible to locate vehicles in real time relative to each other. DENM messages are alert messages. They are issued at the time of an unexpected event in order to cooperate, warn and disseminate information in the geographical are

concerned. They are important messages that would complete the range of possible messages to be exchanged to cooperate and avoid collisions for the IAVs in the Bahnes et al. algorithm [2].

3. Algorithm Improvement

The collision avoidance algorithm of [2] makes it possible to deal with the priority of different vehicles when approaching an intersection. However, it does not deal with the problems of detection, communication and avoidance of fixed or moving obstacles (e.g. human operators).

We extend the algorithm of Bahnes et al. to handle the presence of fixed or moving obstacles in Fig. 1.

Then, we simulate the algorithm staying within the framework of the three scenarios proposed by [2]. These simulations rely on an agent-based model where IAVs are identified [13, 14]. Indeed, agent-based simulation for IAVs is the most common in the same way as simulations based on discrete events or robotics software [15].

IAV agents have the ability to exchange messages and are equipped with radar. This allows them to detect vehicles in front of them. For instance, given an IAV agent a_i, if another IAV agent a_j in front of it is stopped or travelling at a slower speed, the IAV agent a_i can detect it with its radar and stop accordingly to avoid hitting it, as in Fig. 2.

To improve the collective autonomy of the IAVs it is essential that they have a good capacity for individual autonomy. The individual autonomy of the IAVs strengthens their collective autonomy.

The messages exchanged by the different IAVs remain consistent with the Bahnes's algorithm. We just propose two new message types for the collaborative perception (added to the three messages defined by Bahnes et al.: Hello_Msg, Coop_Msg and Ack_Msg):

- Obstacle_msg: a message sent by an IAV agent to the other IAV agents circulating in the warehouse to indicate the presence of a perceived obstacle.
- Alert_msg: a message sent by an IAV agent to the other IAV agents circulating in the warehouse to indicate an unavoidable obstacle.

4. Simulation Results

The traffic plan presented in Fig. 3. It shows the different scenarios that we consider as a benchmark plan to compare results. Ten IAVs are distributed over 3 circuits: the red IAVs on the first circuit, the blue ones on the second and the yellow ones on the third.

It involves different intersections, where vehicles can arrive from different sides like in a warehouse (4 intersections are shown in Fig. 3.). Thus, it provides the different characteristics of an industrial environment and allows us to realize simulated experimental tests in line with realistic scenarios.

We notice in the simulation that the avoidance is well respected and the obstacles are perceived by the IAVs. Therefore, the simulation validates the extended Bahnes's algorithm with collision avoidance and fixed or dynamic obstacle detection processing.

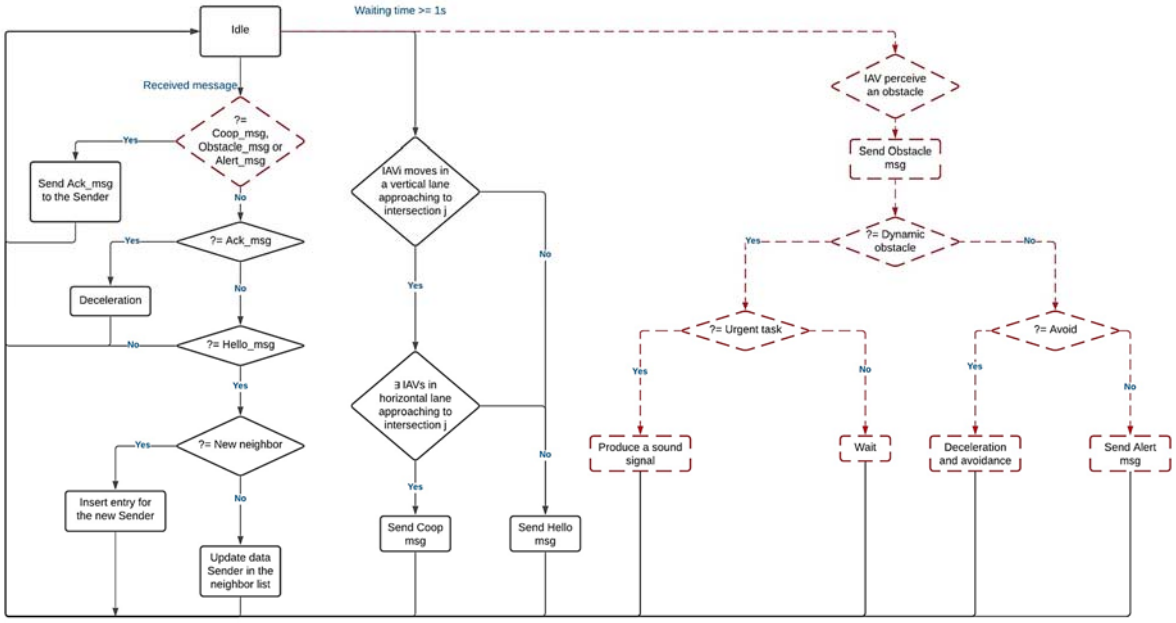


Fig. 1. Improvement of Bahnes's algorithm to treat the problem of collision and obstacles

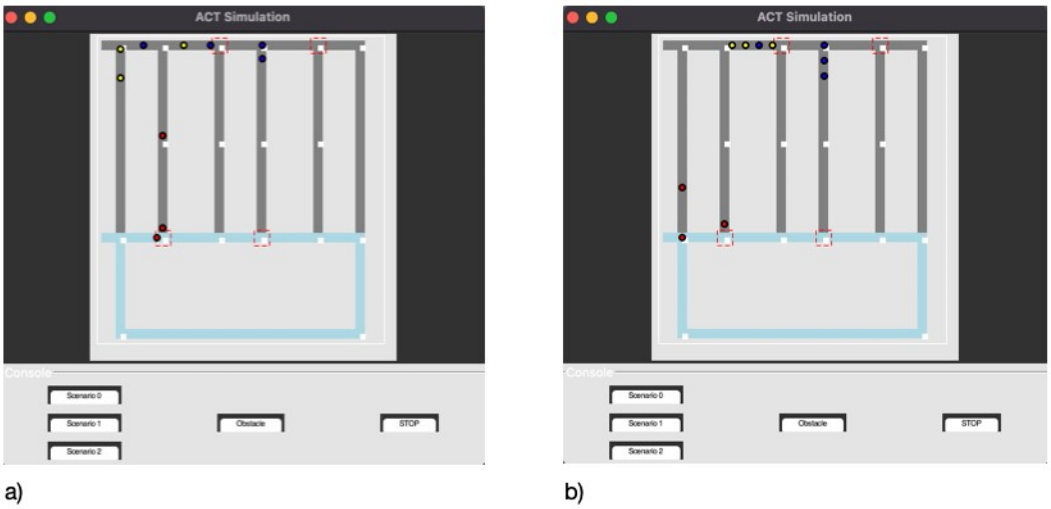


Fig. 2. Simulation of radar use: (a) at the top of the picture: one blue and three yellow IAVs arrive near the intersection, (b) while waiting for the yellow IAV to pass the intersection, the radar of the blue IAV and the two other yellow IAV allow them to stop and keep their distance to avoid colliding

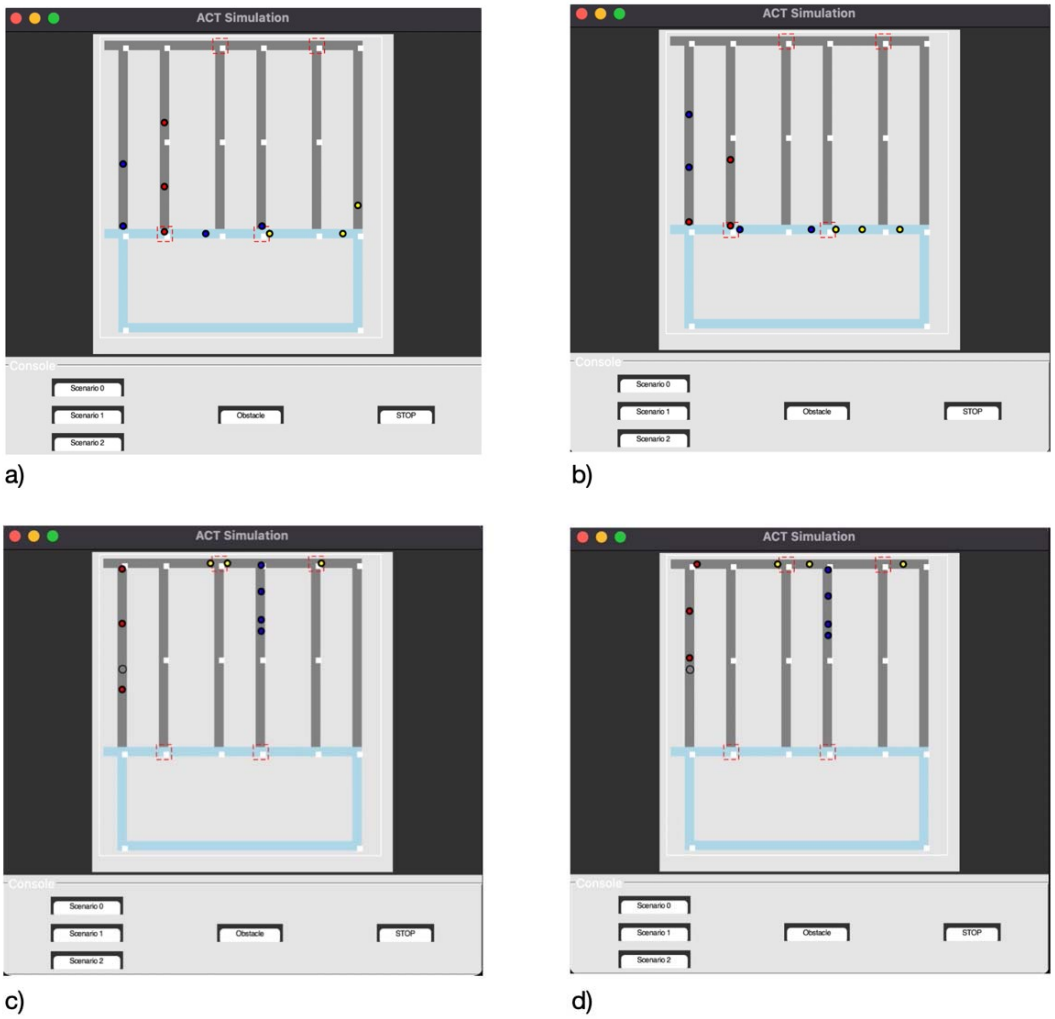


Fig. 3. Simulation of the scenarios: (a) in the center of the picture: a blue and yellow IAV arrive at an intersection, (b) the yellow IAV passed the intersection after communicating with other IAVs, (c) on the left side of the picture: a red IAV perceives a fixed obstacle in front of him, (d) a red IAV avoided the obstacle.

5. Conclusions and Perspectives

In an Industry 4.0 context, many actors cross paths in different areas of a warehouse: vehicles, operators, obstacles (objects that fall or are left in aisles may appear).

The algorithm in [2] proposed a message-based communication protocol between vehicles to prioritise passage through an intersection. We extended it in order to have the possibility to handle the detection of these fixed and mobile obstacles (Fig. 1).

We have validated the algorithm by a simulation approach with the traffic plan presented in [2].

As an extension of the work and the aim to propose a cooperation protocol for a collaborative collision and obstacle detection, the messages will respect the standards defined by ETSI [9, 11].

In future work, we will work to simulate others levels of autonomy using collective strategies – cooperation between IAVs but also cooperation integrating the infrastructure and the environment of the IAVs. We also plan to do real experimentations with robots.

References

[1]. H. Andreasson, A. Bouguerra, M. Cirillo, D.Nikolaev-Dimitrov, D. Driankov, L. Karlsson, A. J. Lilienthal, F. Pecora, J. Pekka Saarinen, A. Sherikov, and T. Stoyanov, Autonomous transport vehicles: Where weare and what is missing. *IEEE Robotics & Automation Magazine*, 22, 1, March 2015, pp. 64–75.

[2]. N. Bahnes, B. Kechar, and H. Haffaf, Cooperation Intelligent Autonomous Vehicles to enhance containerterminal operations, *Journal of Innovation in Digital Ecosystems,* 3, 1, June 2016, pp. 22-29.

[3]. H. Khayyam, B. Javadi, M. Jalili, and Reza N. Jazar, Artificial Intelligence and Internet of Things forautonomous vehicles, in Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems, R. N. Jazar, L. Dai (Eds.), *Springer International Publishing*, 2020, pp. 39-68.

[4]. A. Ndao, M. Djoko-Kouam, and A.-J. Fougères, Matrix Beaconing for the Location of Autonomous Industrial Vehicles on a Simulation Platform, in *Proceedings of the Third Int. Conf. on Advances in Signal Processing Artificial Intelligence (ASPAI 2021)*, Porto, Portugal, November 17-19, 2021, pp. 80-86.

[5]. A. M. Nascimento, L. F. Vismari, C. B. S. T. Molina, P. S. Cugnasca, J. B. Camargo, J. R. de Almeida and A. Y. Hata, A systematic literature review about the impact of artificial intelligence on autonomous vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, 21, 12, 2019, pp. 4928-4946.

[6]. M. Hafner, D. Cunningham, L. Caminiti, and D. Del Vecchio, Cooperative collision avoidance atintersections: Algorithms and experiments. *IEEE Transactions on Intelligent Transportation Systems*, 14, 3, 2013, pp. 1162-1175.

[7]. C. Medrano-Berumen and M. İlhan Akbaş, Validation of decision-making in artificial intelligence-based autonomous vehicles, *Journal of Information and Telecommunication*, 5, 1, 2021, pp. 83–103.

[8]. A. Daniel, K. Subburathinam, B. Muthu, N. Rajkumar and S. Pandian, Procuring Cooperative Intelligence in Autonomous Vehicles for Object Detection through Data Fusion Approach, *IET Intel. Transport Syst*., 14, 2020, 11, pp. 1410–1417.

[9]. Intelligent Transport Systems (ITS); Vehicular Communications; Basic set of Applications; *Part 2: Specification of Cooperative Awareness Basic Service*, ETSI EN 302 637-2 V1.3.2, 2014.

[10]. N. Lyamin, A. Vinel, M. Jonsson and B. Bellalta, Cooperative awareness in VANETs: On ETSI EN 302637-2 performance, *IEEE Transactions on Vehicular Technology*, 67, 1, 2017, pp. 17-28.

[11]. Intelligent Transport Systems (ITS); Vehicular Communications; Basic set of Applications; *Part 3: Specification of Decentralized Environmental Notification Basic Service*, ETSI EN 302 637-3 V1.2.1, 2014.

[12]. J. Santa, F. Pereniguez-Garcia, A. Moragyn and A. Skarmeta, Experimental evaluation of CAM and DENM messaging services in vehicular communications, *Transportation Research Part C: Emerging Technologies,* 46, September 2014, pp. 98–120.

[13]. A.-J. Fougиres, A Modelling Approach Based on Fuzzy Agents, *Int. J. of Computer Science Issues*, 9, 6, 2013, pp. 19-28.

[14]. C. M. Macal, Everything you need to know about agent-based modelling and simulation, *Journal of Simulation*, 10, 2, 2016, pp. 144-156.

[15]. D. Bechtsis, N. Tsolakis, D. Vlachos, and J. S. Srai, Intelligent autonomous vehicles in digital supply chains: a framework for integrating innovations towards sustainable value networks, *Journal of Cleaner Production*, 181, 2018, pp. 60-71.

(025)

Artificial Intelligence and Measurements

R. Taymanov, K. Sapozhnikova, and A. Shutova
D. I. Mendeleyev Institute for Metrology, 19, Moskovsky pr., Saint Petersburg, Russia, 190005
Tel.: + 78122519920, fax: + 78127130114
E-mail: k.v.s@vniim.ru

Summary: The emergence of artificial intelligence (AI) increases human abilities in solving cognitive tasks, expands the scope of measurement as well as forces a rethinking of the basic philosophical question about the material and the spiritual (ideal). The development of AI-based systems will lead to mass production of their various types with relatively weak AI, which, along with systems with strong AI, will change the life of society but create social problems. The paper describes the effectiveness of using measurement systems with AI in manufacturing, medicine, transport, assessment of the product and service quality, determination of human abilities, etc. The directions of activities that can accelerate the positive trends in the new development stage but will delay the emergence and facilitate the solution to related social problems are highlighted.

Keywords: Artificial intelligence, Measurement, Weak and strong artificial intelligence, Development of civilisation, Voice assistant.

1. Introduction

Any creature with a central nervous system regards itself as a part of the environment where it takes action to survive. In this environment, the creature has to separate itself from the other environment components and distinguish components that have special features, e.g., they are potential food, pose a threat, or provide an increase in the population.

The emergence of human beings required the ability to understand the environment much more widely and recognize more specific differences in the environment components. It also turned out to be necessary to identify certain abstractions in this environment which were important for forming the view of the surrounding world. This view allowed planning the life of humans themselves and their tribe for a long time. At first, the view took the form of religion and then that of philosophy. Both forms were changing over time, and the diversity of their versions was growing.

It is now common in philosophy to divide the world into two categories: the material that determines the content of space as well as time and the spiritual (ideal), or in short, matter and consciousness. Advances in artificial intelligence (AI) development, robotics, and additive technology are increasingly leading to discussions on the concept of the spiritual.

It is possible to assume that the spiritual is our perception of the material. Then, the set of design and technological documentation related to some product is a form of the ideal vision of this product based on knowledge allowing to produce the product.

However, this knowledge is incomplete. For example, it does not include information related to some features of materials applied, which can result in changing product characteristics in the course of its operation. Taking into account that to emphasize the unknown in comparing the material and spiritual is desirable, we recommend turning to the opposition "measurable – immeasurable". This transition is reasonable since "when you can measure what you are speaking about and express it in numbers, you know something about it" as Sir W. Thompson, Lord Kelvin, observed far back. (The concept of "immeasurable" and the mentioned Thomson's observation should be referred to the level of knowledge at the time of the statement formulation).

Both philosophy and measurement science should contribute to enriching knowledge of the world and to increasing the effectiveness of activities aimed at its perfecting. Accordingly, the proposed opposition places emphasis on the need to broaden the scope of measurements. Indeed, the possibilities of measurements are increasing with time. One can notice it concerning measuring the properties of natural intelligence and those of artificial intelligence.

2. Vision of Artificial Intelligence, Its Possibilities, and Prospects

The number of papers dealing with AI has increased dramatically in the current century. There exist international standards in this area and dozens of draft standards, both national and international, relating to AI. However, the known definitions of AI are contradictory [1]. Meanwhile, AI is mainly examined in three ways: it can be a set of properties, set of technologies, and an engineering discipline.

Within the frames of this paper, AI is considered to be a set of properties that allow analyzing data that come from input channels, getting knowledge, constructing representations, forming concepts, and reinterpreting them depending on the results of self-learning.

Various sensors and other measuring transducers can be input devices of the channels. Based on the

outcome of data analysis, the AI can identify correlations between the data, recognize images and measure multidimensional quantities, classify them, predict changes in the ratios of various parameters over time, and make decisions or propose them to a human operator.

Nowadays, as a rule, the human operator formulates the tasks. In many cases, the AI finds the algorithm to solve them, sometimes without informing the operator about it. However, in a few years, humans will only set strategic tasks. Tactical tasks, e.g., suppression of interfering influences or troubleshooting, will be handled by the AI independently.

Thus, while in the past technical systems were created to increase the capacity of individual human organs (hands, feet, eyes, etc.) and to speed up the solution to simple mathematical problems, the development of AI opens up the prospect of increasing the human abilities to solve tasks of a cognitive nature.

Since the level of complexity and variety of tasks to be performed by systems based on AI is different, it is reasonable to produce a number of system types that differ in terms of AI level.

It is now common to divide AI into weak (narrow) and strong (general). Systems based on weak AI extract information from a limited set of data and are only able to cope with the specific tasks that they were trained to perform. They are usually oriented towards tasks that, when carried out by a human, do not require high qualifications but oblige him/ her to use regularly updated reference data. Often, communications of similar systems with humans, i.e. conversational skills with a limited vocabulary, are also required. Such tasks include, e.g., working as a waiter/waitress, security guard, tour guide, interpreter, storekeeper, etc. It should be noted that the systems based on weak AI are now relatively widespread, while the number of their types and manufactured product output are growing steadily.

A system with strong AI is much more expensive. It can reason, make judgements in the face of uncertainty, plan, learn, integrate prior knowledge into decision-making, suggest new ideas, etc. It can be assumed that systems with strong AI will also be produced in several types, depending on the specific tasks they solve.

However, it is possible that in the process of the AI-based system operation, such as mobile robots, there will happen a need to coordinate the actions of several robots to solve a new problem. In particular, it will be necessary to perform the actions simultaneously or jointly process information coming from many sources, the number of which is greater than that of channels a single system with AI has.

Apparently, the robot firmware will allow for this kind of cooperation. For example, a robot will "realize" that on its own it cannot solve the task set by an operator and will initiate the cooperation. Essentially, this proactive request for help will mean an emotional response, which, if perceived by another robot, will mark the beginning of an era of robot socialization. A technical revolution of this kind is inevitable, as it opens up new prospects for the development of civilization.

Naturally, there will be a code of ethics for the AI-based systems that should exclude undesirable consequences of their socialization. Significantly, in 2021, the Code of Ethics for Artificial Intelligence was published in Russia [2]. It establishes "general ethical principles and standards of behaviour to guide participants in AI relationships ... as well as of corresponding mechanisms for implementing the provisions of this Code." The Code applies to relations related to ethical aspects of the development (design, engineering, piloting), implementation, and use of AI technologies at all stages of the life cycle..." The Code states that "human beings, their rights, and freedoms shall be regarded as the highest value".

However, it is hardly possible that such a document will be able to fundamentally change the direction of robotics development. As the number of robots capable of emotional responses increases above some limit, more developed socialization is likely, possibly involving interested humans.

Then we can expect the spontaneous selection of leaders. They will be able to set tasks for groups of robots. This does not exclude tacit correction of a robot firmware with their participation. As a result, there have been assumptions about the possibility of conflicts between robots and humans [3]. The history of civilization provides numerous examples of the emergence of similar situations in the relationships between human communities. Usually, they concern violations of ethical laws established by both religions and state laws.

3. AI-based Measurement Systems

The problem of preventing conflicts between robots and humans is not yet a hot topic, although it should not be forgotten. Metrology specialists now face a different challenge: efficiently apply AI for measurement purposes.

The first direction of work is the identification of multidimensional quantities that characterize the occurrence of deviations from nominal values, measurement of the parameters of these quantities, and determination of the deviation dynamics over time. The next stage is the development of methods and means that allow obtaining a desirable result. The essence of the result depends on the task to be solved.

There exist several groups of tasks here.

One of them is to search for irregularities in the materials used in critical products, e.g., to detect a hidden crack in the support of a construction. Conventional methods of ultrasound analysis can detect the presence of irregularities using a panoramic scanning. If the results of the multichannel measurements made with the help of an AI-based system show a correlation of changes, in particular, in many space points, under some impacts on the support,

or over time, then the presence of a crack in the support is diagnosed. This fact makes it possible to prepare a solution aimed at eliminating the hazard.

Another group of tasks is related to medicine, i.e., the detection of abnormalities in human physiological systems. In [4], on the basis of the electrocardiogram processing, Uspenskiy proposes to use AI for analyzing the ratios between blood pressure pulse parameters got at several points on the patient's body surface. Statistics of clinical studies give the basis for the automatic diagnosing of more than forty diseases by characteristic deviations of measured ratios from the nominal values fixed in healthy people. The diagnosis, in its turn, determines the course of treatment to restore the health of the patient.

This method also opens up additional opportunities. As we all know, certain music influences the physiological processes taking place in the body [5]. By analyzing changes in the above ratios, while a patient is listening to selected music excerpts, it is possible to refine the diagnosis and treatment recommendations.

Diagnosing in this way is also technically relatively easy with exposure to force, vibration, heat, or otherwise. This feature makes it possible to help patients having a significant likelihood of abnormal physiological processes.

It is sufficient to arrange measurements of a specially selected group of indicators, which form, using an AI-based system, an estimate of the value of the quantity indicating the approach of an attack. The corresponding signal should be transmitted to the doctor if the patient is in hospital or to the patient himself/herself if he/she has a means of preventing a seizure. Such methods can diagnose diseases at early stages too.

The considered direction of AI-based measuring systems can be effective in the automatic translation of emotionally coloured speech [6]. Special speech processing can single out infrasound multidimensional modulation components and identify, e.g., the irony of the said.

AI can also be efficient in traffic management. The relevance of work in this area increases with the appearance of driverless cars. But also for driving conventional cars in regular traffic flows, AI-based car measuring systems will certainly be in demand. They will be able to track the direction and speed of oncoming and passing cars, calculate the probability of collision with them or with other obstacles, and form signals correcting the car movement.

The above examples concern the application of measurements in areas where they were necessary before but could not provide the required efficiency without AI.

However, in some fields, the term "measurement" can so far only be applied conventionally. This statement relates to the assessment of multidimensional quantities in which the role of an expert is essential. Such quantities include many indicators of the quality of goods and services, the level of knowledge and skills of applicants for a particular job, the emotional expressiveness of the musical composition performance, etc.

However, in some cases, an expert can be biased, i.e. interested in an assessment for reasons unrelated to the materials submitted to him.

The trustworthiness improvement of "measurements" being performed with experts' participation up to an acceptable level that excludes crude errors is the second direction of metrological research and development that relies on the capabilities of AI-based measurement systems.

To minimize the role of subjective factors, expert judgements should be reduced to a common scale. Selecting the experts, one should take into account the competence level bringing the expert number to the number that depends on the task to be solved. This requirement can be met using AI. The useful method is, e.g., the correlation analysis of experts' answers to abstract questions combined with a number of indicators characterizing the features of experts' neurophysiological activity (brain biorhythms, voice sound, etc.) while making their assessments.

The Bayesian approach has been successfully applied in this field too. Examples are given in [7].

The third direction is the measurement of abilities and properties of intelligence, including the emotional sphere.

Usually, the level of human intelligence is assessed by the intelligence quotient IQ. It enables the comparison between individual results and average ones for people of the same age. The assessment is made with the help of tests, i.e., a series of progressively more difficult questions chosen by experts. There exist widespread tests, but the authors are not aware of any standard in this field. The reason seems to be the limited interpretation of intelligence in this method and its emphasis on logic.

However, the role of the emotional sphere in a human intellectual activity is significant. Analysis has shown that AI-based systems make it possible to identify abnormalities in intelligence development, e.g., from the sound of babies' cries, i.e. his/her emotional reactions. The abnormalities can be either negative, which indicates, e.g., the early stages of autism, or positive related to the accelerated development of imaginative thinking.

The ineffectiveness of IQ in investigating human abilities does not undermine the merits of the test method. The application of tests including not logical components only but emotional ones may be efficient in evaluating AI-based systems in case communication with humans is one of their primary functions.

Performing the duties of a waiter, salesperson, security guard, or tour guide could be the examples where the "voice assistant" function can be very appropriate. An IQ assessment with such a test makes it possible to compare the quality of this function in systems produced by different companies.

In particular, in 2019, Loup Ventures prepared 800 questions assigned for testing "voice assistants" Amazon Alexa, Apple Siri, and Google Assistant [9]. They checked the understanding of the question and

whether the answer was correct. The questions concerned the address of an institution, an order to buy something, a request to be reminded of something, etc.

The percentage of correct answers given by Google Assistant, Apple Siri, and Amazon Alexa was 93%, 83%, and 80% accordingly. Comparison with the results of 2015 (86%, 79%, and 61% for the same assistants) shows remarkable progress [9].

However, many AI-based systems are aimed at solving tasks where communication with humans is minimal. In such cases, to use technical indicators to assess product quality is appropriate [10].

The fourth direction is the classification of particular multidimensional quantities, e.g., acoustic signals or visual images. Quite often, the parameter ratios or even the parameters themselves, which are used as characteristic features in quantity identification and classification, are specified with a dominant definitional uncertainty.

As a result, an AI-based measuring system has to compare a number of such ratios to make a final decision. This circumstance gives grounds to classify such measurements as soft measurements [11, 12]. The authors have experience in applying such classification to the automatic translation of animal acoustic signals into the language of human communication [13].

It is noteworthy that the dominant definitional uncertainty of multidimensional quantities is the characteristic feature of some other quantities mentioned above. In the fourth direction, soft measurements are highlighted because they are applied here significantly more frequently.

4. Conclusions

The emergence, widespread use in many areas of human activity, and high improvement rate of AI-based systems indicate that human civilization is rising to a new stage in its development. The coming changes in the life of society are substantial. Accordingly, the main challenge will be to maintain human developmental advantages compared with the systems based on AI.

This problem will be solved if as many people as possible are capable of continuous self-learning and active creativity, which requires significant changes in traditional practices of upbringing and education. It is a critical point when not to be out of time and avoid acute social problems is possible.

References

[1]. R. Taymanov, K. Sapozhnikova, The role of artificial intelligence in measuring systems, in *Proceedings of the XXXI International Scientific Symposium on Metrology and Metrology Assurance (MMA)*, Sozopol, Bulgaria, 7-11 September 2021, pp. 1-5.

[2]. Code of Ethics for Artificial Intelligence (in Russian) Web Portal (https://www.profiz.ru/upl/2021/Кодекс_этики_в_сфере_ИИ_финальный.pdf).

[3]. S. Hawking, Brief Answers to the Big Questions, *John Murray*, 2018.

[4]. V. M. Uspenskiy, Artificial intelligence in the diagnosis of diseases of human internal organs, *Informational and Telecommunication Technologies*, Vol. 50, 2021, pp. 15-21 (in Russian).

[5]. AMTA Web Portal (https://www.musictherapy.org/).

[6]. E. Zinovieva, Yu. Kuznetsov, M. Shakhmatova, I. Baksheeva, I. Danilova, K. Sapozhnikova, R. Taymanov, Metrological approach to emotion recognition in sounding speech, *Philosophy and Humanities in Informational Society*, Vol. 3, Issue 17, 2017, pp. 63-82 (in Russian).

[7]. R. Taymanov, K. Sapozhnikova, S. Prokopchina, What is immeasurable make measurable with artificial intelligence, *Measurement: Sensors*, Vol. 18, 2021, pp. 1-4.

[8]. K. Sapozhnikova, R. Taymanov, I. Baksheeva, S. Kostromina, D. Gnedykh, I. Danilova, Metrological approach to measurements of emotions being expected in response to acoustic impacts, in *Proceedings of the 18th International Congress of Metrology*, Paris, France, 19-21 September 2017, pp. 1-7.

[9]. Artificial Intelligence News Web Portal (https://artificialintelligence-news.com/).

[10]. Draft standard. Artificial intelligence systems. Quality assurance. General (in Russian).

[11]. K. Sapozhnikova, R. Taymanov, S. Kostromina, Widening borders of measurable is a problem of the XXIst century, *Soft Measurements and Computing*, Vol. 37, Issue 12, 2020, pp. 31-44 (in Russian).

[12]. K. Sapozhnikova, A. Pronin, R. Taymanov, Increasing measurement trustworthiness as a necessary part of technology development, *Sensors & Transducers*, Vol. 251, Issue 4, 2021, pp. 61-69.

[13]. K. Sapozhnikova, S. Hussein, R. Taymanov, Iu. Baksheyeva, Music and growl of a lion: anything in common? Measurement model optimized with the help of artificial intelligence will answer, *Journal of Physics: Conference Series*, Vol. 1379, 2019, pp. 1-6.

(026)

Intelligent Sensors Networks for Monitoring and Controlling Complex Systems under Conditions of Uncertainty

S. V. Prokopchina
Financial University, Russian Federation
E-mail: svprokopchina@mqqil.ru

Summary: Within the framework of the Industry 4.0 concept, intensive development of the processes of intellectualization of sensor systems is envisaged. Among the most important specific properties of real measuring processes in complex systems is, first of all, their implementation under conditions of considerable uncertainty. The uncertainty is caused by a priori incompleteness, inaccuracy, vagueness of information about a complex measuring object and its functioning environment, which does not allow to build an adequate model of the object before the measurement experiment, to identify and formalize the influencing factors of the external environment and to develop effective algorithms for the functioning of information and measurement systems.

The report proposes an approach to the intellectualization of measurement systems in conditions of uncertainty by creating intelligent sensor networks based on Bayesian intelligent technologies (BIT) and means of their implementation. Typical modules of such networks are considered, which are integrated sets of various sensors and intelligent measurement information processing systems. The results of the networks are comprehensive assessments of the state of complex objects and recommendations for providing of their sustainable functioning. An important part of such systems is the built-in means of a complete metrological justification of all received solutions. The systems have a hierarchical architecture, according to the levels of management of complex objects, which has the possibility of self-development based on newly received information. This is achieved thanks to models and scales with dynamic constraints on which all BIT algorithms are built. The report provides examples of the use of intelligent sensor networks for monitoring and control of power generation and water supply systems.

Keywords: Intelligent measurements, Bayesian approach, Sensors network.

1. Introduction

Within the framework of the Industry 4.0 concept, intensive development of the processes of intellectualization of sensor systems is envisaged.

Among the most important specific properties of real measuring processes in complex systems is, first of all, their implementation under conditions of considerable uncertainty. The uncertainty is caused by a priori incompleteness, inaccuracy, vagueness of information about a complex measuring object and its functioning environment, which does not allow to construct an adequate model of the object before the measurement experiment, to identify and formalize the influencing factors of the external environment and to develop effective algorithms for the functioning of information and measurement systems.

Other specific properties of complex systems are their dynamism and active relationship with the environment, which determines special requirements for methods of creating models, including measuring ones. The tasks of monitoring the state of complex systems and their effective management, especially in real-time modes of system operation, determine the need to obtain not individual measurement results, but assessments, conclusions, and management recommendations. Such opportunities can be provided by involving artificial intelligence methods in measuring processes. The Bayesian intelligent technologies developed by the author on the basis of the regularizing Bayesian approach make it possible to implement monitoring and management of complex systems in conditions of significant uncertainty and active dynamic influence of the external environment. Methodology, information technologies and applied examples of solving applied problems based on them are given in the author's works, for example, in [1-3].

During the implementation of a number of applied projects related to the creation of intelligent measuring systems for distributed man-made, natural and socio-economic systems, the concept of intelligent sensor networks was developed.

Such networks allow for a comprehensive assessment of the state of distributed systems at any time, to determine the main risks and potentials of both individual sites and the system as a whole.

The intelligent sensor network consists of a central module for intelligent information processing and peripheral modules that collect information on all parts of the network.

As the central module, an intelligent environment built on the Infoanalytic platform is used, which receives information flows from peripheral intelligent modules, applied neural networks for image and document processing, statistical databases, expert assessments and other information. The structure of such a network is shown in Fig. 1.

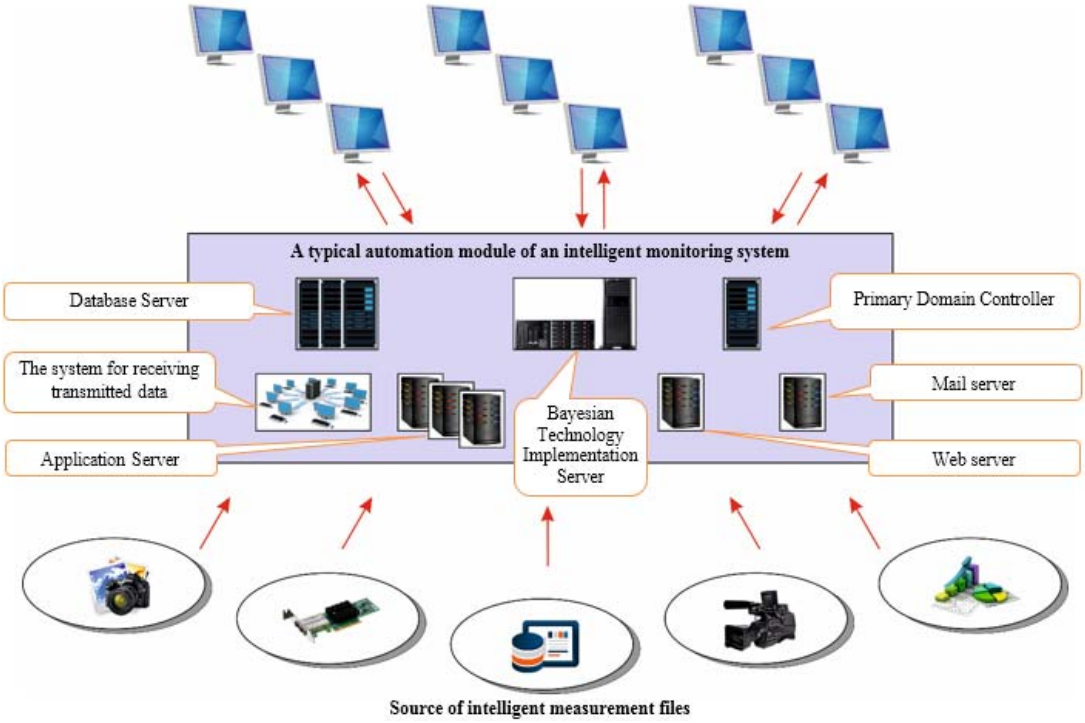


Fig. 1. Block diagram of a typical module of an intelligent sensor network section.

The experience of such developments made it possible to form a typical peripheral sensor network module that implements all the above functions for a separate network section.

It consists of the following subsystems:

- integrated sensor sets that measure system parameters;
- intelligent controllers that implement the functions of intelligent processing of primary measurement information, consisting in the integration of information from individual sensors in order to obtain estimates, and the coordination of protocols for transmitting solutions received on this part of the network to the central module;
- subsystems of interfaces and web services in monitoring mode;
- databases of primary measurement information.

The type of a typical sensor network module is illustrated in Fig. 1. A conceptual model for obtaining a solution in a peripheral sensor network module based on Bayesian intelligent technologies (BITS).

The conceptual model of creating a solution based on Bayesian intelligent technologies (BITS) is based on the methodology and principles of creating models with dynamic constraints based on the regularizing Bayesian approach [1-3]. The stages of the complex solution of the problem of creating an intelligent big data processing system can be represented by the following model:

$$Q=Q\_1*Q\_2*Q\_3*Q\_4*Q\_5*Q\_6*Q\_7*Q\_8*Q\_9*Q\_10*Q\_11 \quad (1),$$

where:

Q is the generalized algorithm for creating a typical sensor network module;
Q1 is the creation of a conceptual model of an object with dynamic constraints (MDO) for this section of the sensor network;
Q2 is the Sensor system selection;
Q3 is the construction of scales with dynamic constraints for each sensor in the intelligent controller;
Q5 is the implementation of measurements, obtaining primary measurement information from sensors;
Q6 integration of individual measurements of various sensors according to the modified Bayes formula;
Q7 interpretation of integrated information and its presentation in the form of an assessment of the state of a network section;
Q8 is the determination of dynamics and dynamic models, trends in individual parameters and in general according to the state of the system in this section of the network and trends in the development of situations;
Q_9 is the assessment of risks and potentials of the situation and display by methods of cognitive graphics;
Q10 is the interpretation of the current situation;
Q_11 is the generating recommendations to improve the situation.

Convolution of information (in formulas (1) and (3) this action is indicated by the symbol *) about the value of the element x_ij, i=1,...I is performed according to the modified Bayes formula for discrete laws of hypothesis distribution in the form of membership functions (linguistic scale of BII for processing non-numeric, qualitative information) or probability distributions (numerical scale of BII for processing non-numeric, qualitative information). The

probability of a solution (hypotheses about the state of an object, the value of a parameter, or the like) is calculated by the formula:

$$\mu\left(h_{kj}^{0\lambda}\middle|h_{kj}^{0}|\ X_I^0|Q_I^0\right) = \frac{\mu\left(h_{kij}^{\lambda}|h_{kij}|x_{ij}|x_i|Q_i\right)\circ\mu\left(h_{k,i+1,j}^{\lambda}\middle|x_{i+1,j}|\ X_{i+1}|Q_{i+1}\right)}{\sum_{i=1}^{k}\mu\left(h_{l,i+1,j}^{\lambda}|h_{l,i+1,j}|x_{i+1}|\ X_{i+1}|\ Q_{i+1}\right)} \quad (2),$$

where:
$h_{kj}^{0\lambda}$ is the integral (for a set of information flows X_I^0)) regularizing Bayesian estimation (RBO);
h_{kij} is the RBO for X_i data stream;
$\mu\left(h_{kj}^{0\lambda}\middle|h_{kj}^{0}|\ X_I^0|Q_I^0\right)$ is the posteriori probability of evaluation $h_{kj}^{0\lambda}$.

The general scheme of the system being developed consists of: a measurement storage server, sensors; a controller; measurement processing and transmission to the server; a workstation; a neural network for generating scripts; a web service; clients; IIS Server Windows Infoanalytics; MS SQL; a mathematical apparatus for data analysis and decision-making; system management.

A typical peripheral module of an intelligent sensor network, as shown in Fig. 1, includes basic measuring instruments and an intelligent controller, as well as the necessary modules for organizing web services. Measuring instruments may include both sensors for measuring physical parameters and virtual measuring instruments for measuring integral, non-physical and non-quantitative parameters. This architecture of a typical module allows you to combine, integrate and process measurement and observation data in monitoring mode. The use of special scales with dynamic constraints [1, 2] allows measurements to be made under conditions of uncertainty, significant dynamism of the measured parameters and the active influence of the external environment.

2. Conceptual Model for Monitoring the State of the Water Supply Network

The conceptual model for monitoring the state of the I-th section of the water supply network based on Bayesian intelligent technologies (BIT) has the following form:

$$Qi=Q\_1*Q\_2*Q\_3*Q\_4*Q\_5 \quad (3),$$

where:
Qi is the integral assessment of the condition of the water supply network section;
Q_1 is the integral assessment of the condition of the well on the section of the water supply network;
Q_2 is the integral assessment of the condition of a linear section of the water supply network ;
Q_3 is the integral assessment of the condition of the external sections of the water supply network;
Q_4 is the integral assessment of the network status based on information from users;
Q_5 is the integral assessment of the network status based on information from experts.

The conceptual model of monitoring the condition of the water supply network as a whole can be written as follows:

$$Q = * \ i=1,I \ Qi \quad (4)$$

Such methods and technologies are used in the implementation of real projects. In particular, examples of applied solutions of an intelligent sensor network for assessing the state of the water supply network of one of the cities of the Russian Federation are given below.

Fig. 2 shows the interpretation of the situation according to the integral factor "Information from metering devices" on 12.11.2021.

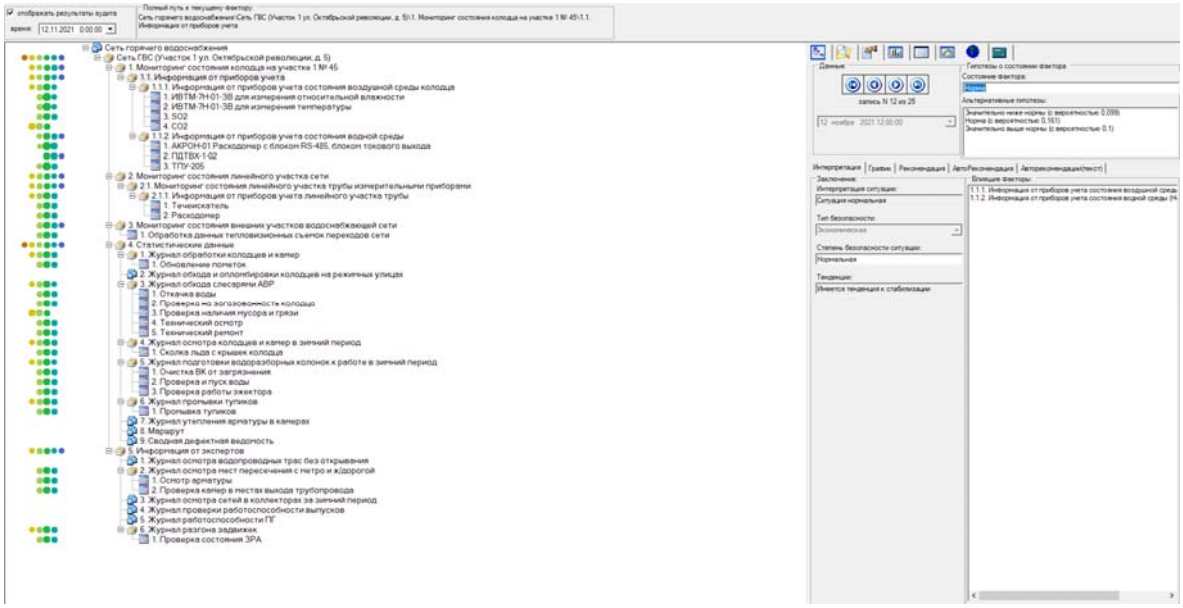


Fig. 2. The interpretation of the measuremental situation.

The results of the work of a typical module in the figure are reflected in the form of cognitive graphics. The color of the circles located to the left of the measured characteristics determines the state of this parameter. In this example, the green color indicates that the parameter is within acceptable limits. The yellow color determines the output of the measured parameter value beyond the range of acceptable values.

3. Conclusion

The article presents methodological aspects and principles of construction of the measuring peripheral module of the system and the version of the Infointegrator with the possibility of remote intelligent processing of various types of information, including measuring data streams from devices, expert assessments of engineers and technicians, statistical data of bypass logs, facts and information from Internet sources with constant connection of production-oriented sites. This architecture is implemented for the first time and allows you to implement a new type of IIoT, namely, intelligent processing of big data of industrial production.

References

[1]. Prokopchina S. V., The concept of Bayesian intellectualization of measurements in the tasks of monitoring complex objects, *News of Artificial Intelligence,* No. 3, 1997, pp. 7-56.

[2]. Prokopchina S.V., Fundamentals of scaling theory in economics, Scientific Library, Moscow, 2021.

[3]. Prokopchina S. V., Cognitive Bayesian measurement networks based on the regularizing Bayesian approach, *Soft Measurements and Calculations*, 2018, No. 2, pp. 56-69.

[4]. Prokopchina S. V., Methodological foundations of scaling in modern Measurement Theory. Classification of measurement scales and their application under uncertainty based on Bayesian Intelligent Technologies, *Journal of Physics: Conference Series,* 1703, 1, 2020, 012003.

[5]. Prokopchina S. V., New Trends in Measurement Science. Bayesian Intelligent Measurements, in *Proceedings of the 5th International Conference on Sensors and Electronic Instrumentation Advances (SEIA' 2019),* Adeje, Spain, 2019, pp. 317-322.

(027)

Methods and Technologies of Bayesian Intelligent Measurements for Human Resources Management in Industry 4.0

S. V. Prokopchina, E. S. Tchernikova
Financial University, Russian Federation
E-mail: svprokopchina@mqqil.ru

Summary: An important place in modern methodologies and technologies of personnel management at the enterprise is occupied by the direction of innovative development, culture of innovation. For this direction, from the position of intellectualization of innovative processes and, in particular, motivation and involvement of personnel, non-quantitative indicators are characteristic, which, following the ISO series standards, should be measurable.

This led to the use of Bayesian intelligent technologies, in particular Bayesian intelligent measurements, for the purpose of measuring indicators of the state of the company's personnel and their management. The aim of the work is to build a systematic methodological basis and means of creating a digital environment for the implementation of the principles of the culture of innovation based on fundamental and best practical psychological work by means of BIT and BII. The implementation of a number of methodological constructions based on individual modules of the Infoanalytic computer system is shown, which contributes to the digitalization of this activity, the dissemination of experience and knowledge for the training and continuous training of specialists, including remotely.

Keywords: Human resources management, Bayesian intelligent measurements.

1. Introduction

The most important part of automation and intellectualization of production enterprises is the personnel of the enterprise. The most advanced production technologies may not ensure its effectiveness in the absence of qualified, organized and involved in innovative processes personnel. A significant number of factors affecting the state of personnel and the impossibility of direct measurement of most of them determine the need to use specialized methods of personnel management in modern conditions of the functioning of enterprises.

An important place in modern methodologies and technologies of personnel management at the enterprise is occupied by the direction of innovative development, culture of innovation. For this direction, from the position of intellectualization of innovative processes and, in particular, motivation and involvement of personnel, non-quantitative indicators are characteristic, which, following the ISO series standards, should be measurable.

This led to the use of Bayesian intelligent technologies, in particular Bayesian intelligent measurements, for the purpose of measuring indicators of the state of the company's personnel and their management.

2. The Main Terms and Indicators of the Culture of Innovation

Cultura of innovation is the sum of codes, norms and habits of employees contributing to the adoption and adaptation of new practices, processes and paradigms in order to increase the value of their activities.

The hypothesis of our research is that such an environment exists and, by identifying its characteristic indicators and measuring them, it is possible to assess the overall state of the innovation sphere at the enterprise and formulate the most positive conditions for creating innovations.

Thus, the aim of the work is to build a systematic methodological basis and means of creating a digital environment for the implementation of the principles of the culture of innovation based on fundamental and best practical psychological work by means of BIT and BII.

The achievement of such a goal will lead to the transformation of the knowledge and practices of innovative companies into an applicable set of tools that allow the company to build a culture of qualitative change, will make available to managers the entire palette of tools for improving the efficiency of the company and unlocking its potential

The object of research of this work is intellectual technologies of digitalization of innovative psychological practices and processes for the continuous development of personality as a member of various social communities, including labor.

The object of the study is also the specific human features of the mental regulation of the activity of individual and group subjects, depending on the natural influence of various factors.

The methodology of the regularizing Bayesian approach and technologies based on it are proposed as a methodology for creating intelligent systems for

conditions of information uncertainty. The effectiveness of the proposed approach is confirmed by its methodological principles focused on information processing in conditions of uncertainty (incompleteness, fuzziness and inaccuracy of psychological information) and the practice of its application to create intelligent systems implemented for a wide range of tasks related to monitoring and management of complex objects, which include the above innovative processes.

In the course of this scientific work, the following results were obtained.

1. The terminological basis of the theoretical provisions of the culture of innovation and motivation is determined.
2. A historical review of the existing concepts of the implementation of psychological innovations and motivation is carried out.
3. A critical review of methodological approaches and best practices for the implementation of innovation processes and motivation of individuals or groups of subjects in accordance with the formulated requirements of methodological and practical relevance.
4. The conditions and factors that favor and hinder the implementation of innovative psychological techniques and motivation of subjects are formulated
5. The regularizing Bayesian approach and Bayesian intellectual technologies based on it are reasonably chosen for the implementation of the principles of psychological development of subjects and communities, including labor.
6. The methods of implementation of psychological innovations and motivation based on the conclusions and methodological decisions made in the work are proposed.

The practical significance of the work is:

1. Developed practical methods and recommendations for the implementation of the principles of the culture of innovation, as well as, as part of it, the motivations of subjects that have been tested on a number of communities. including labor.
2. Conducted psychological projects and experiments, the results of which can be used as effective practices.
3. Implementation of a number of methodological constructions based on individual modules of the Infoanalytic computer system, which contributes to the digitalization of this activity, the dissemination of experience and knowledge for the training and continuous training of specialists, including remotely.

The results obtained make it possible to optimize the personnel management system in the organization, as well as to predict the nature of the impact of certain managerial influences on the level of satisfaction, involvement and effectiveness of employees of the organization.

In this study, we turned to the experience of leading consulting companies, as well as experts, and based on this critical review, we identified four global changes that affect the organization of labor at the present time and should be measured:

1. The sixth technological order;
2. Globalization and decentralization of work;
3. New behavior;
4. The growth of creative professions

All these global trends will only intensify in the future, but it is innovative companies that have already learned how to use them in their work and benefit from these total changes.

Technoclade is a set of generally accepted and applied production technologies and scientific and technological progress

The technological order is not just one change, but a whole series of related and complementary changes. The sixth technoclass is primarily associated with the synthesis of sciences. It is associated with breakthroughs in the field of artificial intelligence and robotics. The company's ability to adapt to new technological conditions will directly affect its viability. The simplest example associated with the emergence and spread of the Internet, we saw in the fifth technoclass.

Our education and the structure of social relations are still in the fourth, industrial technological order. They are focused on working in a hierarchy, consistency and efficiency, having an ideal sample and working according to a template. But the sixth techno-layout allows using robotics, the Internet, things and new analytical abilities of computers to replace not only manufacturing professions, but also some office intellectual tasks. We are already seeing an actively developing market of bots capable of recognizing and solving a client's problem. This means that the company's personnel will have to learn not only to use new technological capabilities, but also to move from purely performing activities to creativity and task setting. Innovative business has such experience. He develops creativity in his employees, supports the model: "Need - training - development - need - training - development".

All this means that companies are turning into universities and colleges. The performing discipline inherent in rigid industrial structures is being replaced by creative and intellectual work. Enterprises should become a platform for continuous development: develop mentoring competencies, implement training tools and various professional collaborations.

It was possible to identify six integral indicators of the state of the enterprise's innovation sphere, which were broadcast unambiguously and were found in the absolute majority of innovative organizations. They include: Transformation, Self-organization, Integrity, Diversity, World-centricity, Creation (entrepreneurship). These indicators together form the basis for the development and implementation of innovations. They interact and complement each other, creating a catalyst effect for breakthrough solutions, discoveries, research and technical creativity.

3. A Measurement Model based on Bayesian Intelligent Technologies for Personnel Management and Innovative Culture of the Enterprise in Conditions of Active Interaction with the External Environment

The main methodological basis of this chapter was the work of Professor S. V. Prokopchina and her scientific school. The following is the concept of S. V. Prokopchina for the formal justification of innovative processes in conditions of uncertainty.

Within the framework of this methodology, it is advisable to present innovative activity as a complex system that actively interacts with the external environment.

The conceptual model of such a system can be represented as a dynamic model of a real system of interrelated innovation processes and, according to the notation introduced above, can be written in the following

$$G_t^{(ine)} = G_t^{(in)} * G_t^{(e)} * G_t^{(y)} \quad (1)$$

where $G_t^{(in)}$ is the model of an innovation system actively interacting with the external environment, $G_t^{(e)}$ is the model of the external environment obtained on the basis of a priori and incoming information; $G_t^{(y)}$ is the model of conditions for the implementation of innovations; * is the symbol denoting the convolution of spaces (compacts) of the specified models in which they are defined.

In order to solve the problem of identifying influencing factors for the development of innovations in Russia, the conceptual model (1) should be detailed to the level of specific influencing factors.

Based on the methodology of the system approach, it is possible to build a model of a complex object as a set of its properties and the relationships between them in interaction with the set of properties of the external environment in the conditions of its study and operation.

On the basis of the RBA methodology, the Infoanalyst computer platform has been developed for the rapid development of digitalization systems of activities. Such a system has the properties of a simple configuration of models - digital images of employees of the organization, the flexibility of measuring, auditing and generating recommendations in conditions of incomplete and inaccurate information. It is designed in the form of an intellectual workplace of a personnel management specialist (IRM-Cadres), a patent for which was obtained by S. V. Prokopchina in 2004.

Within the framework of the task, the main attention should be paid to the composition of environmental factors and conditions of innovation activity. At the same time, the part of the innovation system that determines the processes of digitalization should be allocated to the subsystem of conditions for the rest of the innovation activity.

Then the model of the innovation system (1) can be written as:

$$G_t^{(ine)} = G_t^{(in)} * G_t^{(e)} * G_t^{(d)} \quad (2)$$

where $G_t^{(d)}$ is the digitalization system.

Any model of a complex object can be represented as a composition or a convolution of its properties. We denote the enlarged or integral properties of the innovation system as $Q_{ti}^{(in)}, (i = 1, I)$, and the media $Q_{ti}^{(e)}, (i = 1, J)$ with weights $P_{ti}^{(in)}$

Then the dynamic model of the innovation system can be written as follows:

$$G_t^{(in)} = *_{i=1,I}\left(P_{ti}^{(in)} Q_{ti}^{(in)}\right) \quad (3)$$

Environment model:

$$G_t^{(in)} = *_{i=1,I}\left(P_{ti}^{(e)} Q_{ti}^{(e)}\right) \quad (4)$$

model of digitalization processes:

$$G_t^{(in)} = *_{i=1,I}\left(P_{ti}^{(d)} Q_{ti}^{(d)}\right) \quad (5)$$

Further details of the measured indicators are given in [1-3].

To measure the above integral indicators, the system uses scales with dynamic constraints that allow measuring non-quantitative indicators and making a convolution of individual indicators included in equations (1)- (5) and obtaining an overall assessment of the state of the innovation environment.

Fig. 1 shows an illustration of the implementation of the conceptual model within the innovation sphere of industrial enterprises for Russia. Due to the flexibility and self-learning ability, the system can be used to organize innovative processes of psychological development of employees. At the same time, it is possible to measure all the listed indicators and characteristics of innovative and motivational processes.

4. Conclusion

The paper considers an approach to assessing the state of personnel and the innovative sphere of an enterprise based on Bayesian intellectual measurements. Personnel management is associated with the development of the innovative sphere of the enterprise and the conditions of its development under the active influence of the external environment. An example of the implementation of the concept in the "Infoanalyst" environment is given.

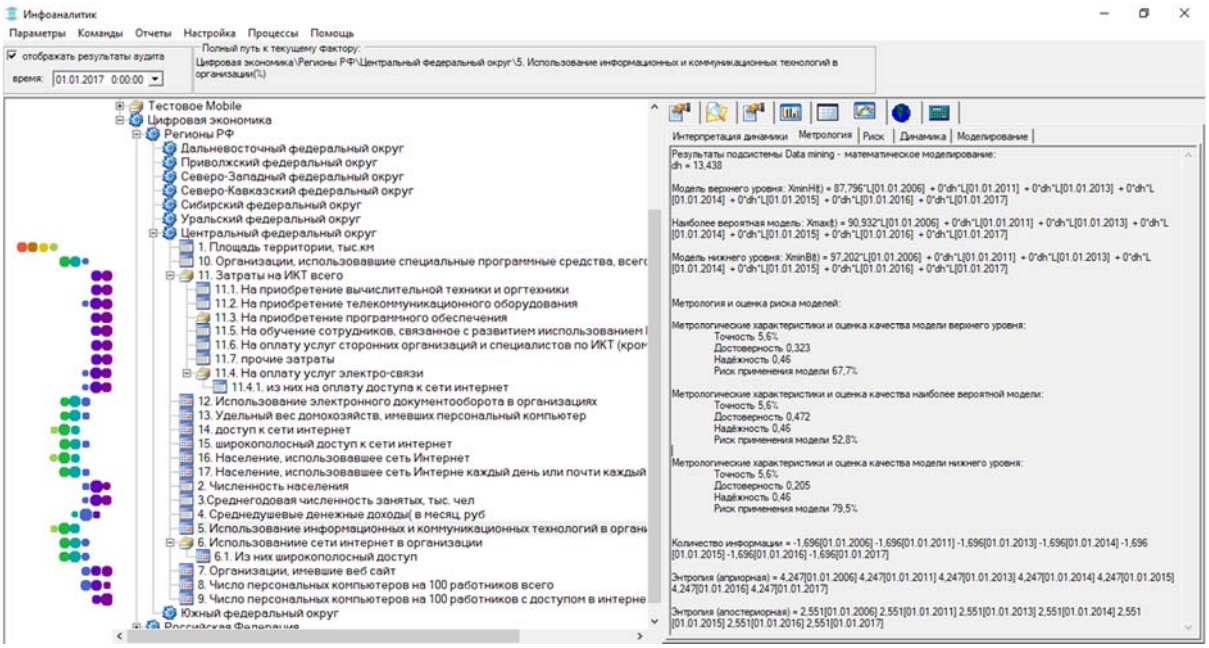


Fig. 1. Implementation of the conceptual model of the innovation sphere in the "Infoanalyst" environment.

References

[1]. Prokopchina S. V., System of mathematical processing of statistical information "Bayesian Mathematical Statistics", in *Proceedings of the International Conference on Soft Computing and Measurement*, St. Petersburg, 2007, June 25-27, pp. 35-45.

[2]. Prokopchina S. V., Soft measurements: methodology and application in scientific, technical and socio-economic problems of the digital economy, *Soft Measurements and Computing,* No. 9, 2018, pp. 4-33.

[3]. Prokopchina S., New trends in measurement theory: Bayesian intelligent measurement and its application in the digital economy, in *Proceedings of the CEUR Workshop*, 2782, 2020, pp. 80–88.

(028)

Intelligent Acoustic Monitoring of Underground Communications

V. P. Koryachko [1] , V. G. Sokolov [2] and S. S. Sergeev [3]
[1] Cad Department of the Supreme Council of the RSRTU, Ryazan, Russia
[2] RSRTU, Ryazan, Russia
[3] TECHNOAS-SK LLC, Kolomna, Russia
E-mail: koryachko.v.p@rsreu.ru, sokolovvg@list.ru, sss@ecnoac.ru

Summary: The article deals with the problem of searching for leaks from pipelines by receiving and processing acoustic emission signals that occur when the liquid flows at the leakage site. The purpose of the work is to review existing solutions for acoustic monitoring, determine the main characteristics of acoustic emission signals and assess the possibility of their recognition in the monitoring process using artificial neural networks.

Keywords: Acoustic emission method, Acoustic leak detector, Correlation leak detector, Acoustic monitoring, Spectral analysis, Correlation analysis, Artificial neural networks.

1. Introduction

A significant part of the city's engineering communications was often built more than a few years ago and is characterized by a significant degree of wear and tear and high losses due to leaks. Even in relatively developed countries, such as Canada, municipal water systems lose, on average, up to 13% of drinking water due to leaks associated with damage to communications. In countries where plumbing systems have not changed for decades, the level of losses is even higher. For example, in the UK this figure is at 23 % [1]. In the Russian Federation, the average leakage in the housing stock in 2004 was estimated at 20-30% of the total water supply to the population [2]. In 2010, the number of hidden leaks on the Moscow water pipeline was estimated at only 5%, but in absolute terms it reached 225 thousand cubic meters per day [3]!

Under these conditions, it is necessary to use remote and non-destructive methods of pipeline inspection in order to identify developing defects as quickly as possible and perform the necessary repairs in time before a serious accident occurs. The following methods are best known:

1. Control of pressure and/or flow of liquid in the pipeline. The necessary devices, as a rule, are installed on pumping stations, on the inputs to the meters, etc. and have the ability to transfer data to remote control centers.

2. Ultrasonic diagnostics of pipelines and welds, based on the propagation of ultrasound in the metal and its reflection from various inhomogeneities. Requires direct access to the pipeline and its preliminary preparation (release from insulation and stripping of sensor installation sites).

3. Magnetic flaw detection based on the registration of changes in the magnetic properties of the metal in the zones of concentration of stresses and deformations. It is possible both in contact (with the installation of equipment on the surface of the pipeline) and in the non-contact version (using portable magnetometer sensors moved by the operator near the pipeline).

4. The method of acoustic emission, based on the reception and listening to sounds that occur when the liquid flows at the leakage site. Can be implemented both with the installation of sensors directly on the pipeline, and with remote listening using a highly sensitive microphone.

The acoustic emission method is considered the most common and requires less effort and less expensive equipment than previous methods. Unfortunately, the method is also characterized by low noise immunity, because both the leakage signal and various kinds of noise (industrial, transport, etc.).) are in the audible range, and it is not always possible to distinguish them from each other. It takes a lot of practical experience and good hearing to determine whether a given sound is a leak and what its volume is.

The functions of the operator can in principle be performed by an intelligent computing system, in which the role of experience is performed by training using a set of reference signals. This approach has proven itself in areas such as image recognition, speech recognition, etc. The most promising direction in intelligent systems at present, of course, are artificial neural networks (ANNs). The average of the advantages that ANNs have often highlight the possibility of solving difficult to formalize problems for which it is difficult to formulate certain criteria and algorithms. This article discusses the main types of equipment for acoustic leak detection and evaluates the main characteristics of acoustic emission signals in order to assess the possibility of using ANNs in this area.

2. Problem Statement and Review of Acoustic Monitoring Equipment

As a result of exposure to adverse factors (corrosion, overpressure, extraneous mechanical influences, manufacturing defects, etc.), holes and cracks appear in the initially continuous wall of the pipeline, through which the leakage of liquid under pressure begins. Leaks are routinely classified by flow rate per unit of time:

- "very weak" - up to 1 l / min;
- "weak" - from 1 to 5 l / min;
- "average" - from 5 to 15 l / min;
- "strong" - more than 15 l / min.

When the liquid jet flows out, acoustic vibrations (acoustic emission) occur, which spread in the soil, liquid and pipeline walls and can be recorded by acoustic sensors. Also in the pipeline and near it there may be other acoustic vibrations - industrial, transport and natural interference, as well as noise created by the liquid when interacting with various inhomogeneities (flange joints, growths, etc.) inside the pipeline (see Fig. 1).

The most common acoustic search for leakage is by the signal of the sensor moving on the surface of the ground near the pipeline and connected to an acoustic receiver in the hands of the operator (sensor D3 and receiver AP in Fig. 1). When the sensor is found directly above the leakage site, the signal recorded by the acoustic receiver will be the greatest. this method is a classic "maximum method", then its error is relatively large and is ~ 1 m at a depth of penetration of the pipe pipeline (1.5 ... 2) m. The appearance of common acoustic leak detectors is shown on the Fig. 2.

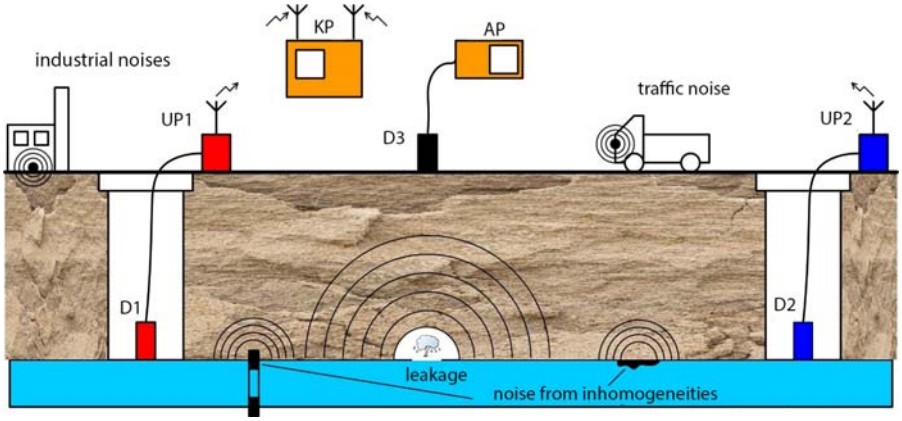


Fig. 1. On the statement of the problem of pipeline acoustic monitoring.

Fig. 2. Appearance of common acoustic leak detectors.

A smaller error in determining the leakage site is provided by the correlation method, in which two sensors are used, which are installed, as a rule, on the pipeline on both sides of the presumed leakage (sensors D1 and D2 in Fig. 1). Signals from the sensors are amplified and transmitted by the UP1 and UP2 transmitters to the KP correlation receiver, which calculates the correlation function of the two signals and determines the time shift between the signals to its maximum. Knowing the distance between the sensors and the speed of the call in the walls of the pipeline, it is possible to determine the coordinates of the leakage site with high accuracy. The typical value of the instrumental error of correlation leak detectors is 0.1 m or less. The appearance of common correlation leak detectors is shown on Fig. 3.

Unfortunately, correlation leak detectors also have drawbacks that can significantly reduce the accuracy of the search for a leakage site, including the following:

(a) Almost always, the correlation function calculated from signals from watchers has several peaks, the largest of which does not necessarily correspond to a real leak. Acoustic interference from mechanisms of pumping stations, elevators, valves, etc. can spread in the pipeline. The operator has to

manually, by the method of subsequent approximation, select the frequency band of signal analysis in order to build up from the crunching signals.

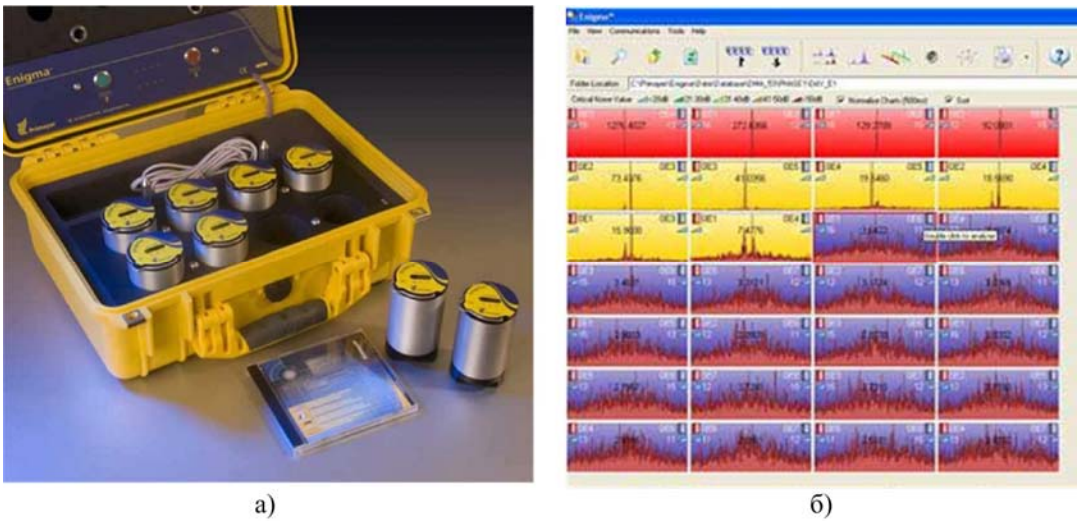
b) The section of the pipeline between the sensors may have bends and taps, due to which the distance between the sensors may differ from that entered into the correlation receiver. Most regulatory documents contain a strong recommendation to trace the surveyed section of the pipeline using a tracer before starting work.

c) The source of the greatest error can be the calculated value of the speed of sound in the walls of the pipeline, which, as a rule, is set in tabular form depending on the material and diameter of the pipe. At the same time, the speed of sound depends on a number of other parameters, especially on the thickness of the pipe walls. The thickness of the walls of worn pipes can differ from the original several times, as a result of which the difference between the real significance of the speed of sound and the given one can be up to 30%! To reduce this error, many correlation leak detectors implement a mode for measuring the speed of the call by solving the inverse problem - determining the delay between signals for a source with pre-known coordinates.

Fig. 3. Appearance of common correlation leak detectors.

Despite the disadvantages listed above, portable acoustic and correlative leak detectors provide acceptable results in most cases and are widely used in organizations operating heat and water supply networks. Less common are more expensive multi-position leak detectors, which include several sensors at once, which can be placed on the pipeline according to a given Schema. The sensors are synchronized with each other and simultaneously record acoustic signals, storing them in their own memory. The data can then be read on the PC for further processing.

A certain fame in domestic organizations received a multi-position leak detector Enigma company Primayer (UK). It consists of 8 self-powered sensors, a case for transportation and programming and software control, which allows you to determine the correlation between the signals of any pair of sensors (see Fig. 4a). The practicals of using Enigma are as follows: during the day, the dutchiks are installed on the pipeline, the sensors are programmed to start at night (when the level of sidely noise is minimal).The next day, the sensors are removed, information is read from them, then the sensors are moved to a new section of the pipeline. The internal power supply of the sensors ensures operation for 5 years. The sensors themselves are water-resistant, can be operated at a depth of up to 10 meters.



а) б)

Fig. 4. 'Enigma' multi-position leak detector and software.

The Enigma software allows you to calculate the correlation functions between the signals of any pair of sensors (a total of 28 possible combinations) and evaluate the number or absence of leakage in the form of the correlation function. Correlation functions with a minimum background level and a distinct maximum are considered to correspond to a leak with a high degree of probability and are marked with a red fill color (see Fig. 4b). if the background level of the correlation function is higher, the fill color will turn yellow (the probability of leakage is average), with a very high background level, the fill color is blue (probably leakage accuracy is small).

An even more expensive, but also more effective method of detecting leaks in the present time is the continuous monitoring of communications using territorially distributed systems, the sensors of which are constantly installed in wells, on hydrants, in thermal chambers, etc. Reading information from such sensors is carried out remotely, as a rule, using GPS/GPRS or LoRaWAN channels. The most widespread such systems have received abroad, where with their help they expect to significantly reduce losses from water leaks.

In 2019, Primayer entered into a £40 million ($53 million) framework agreement with Anglian Water, a utility in east England. Water utilities in England and Wales are encouraged to reduce losses by at least 15% between 2020 and 2025 and halve them by 2050 [4]. As part of the agreement, about 3500 Enigma3hyQ intelligent sensors using multi-point noise correlation technology will be installed on the Anglian Water water supply network. To visualize the locations of the leak, the location of each sensor is displayed on Google Maps and Street View. An example of the installation of Primayer sensors is shown in Fig. 5.

Fig. 5. Acoustic sensor for Primayer geographically distributed monitoring system.

Since 2010, more than 9400 Mlog Radio acoustic sensors from Itron have been installed on the water pipelines of Providence Water (USA). Providence Water estimated that 11.6 % of all water pumped into the water supply goes into various kinds of leaks and does not reach the consumer, which led to annual losses of more than $ 954,000. In addition, the most severe leaks can damage the foundations of buildings and road covering. For example, to eliminate one of the pipe breaks under the highway, the state The repair of the road cost 45350 dollars. Across the U.S. as a whole, it is estimated that U.S. utilities spend more than $900 million annually to fix water leaks and repair pipe ducts.

The capabilities of geographically distributed monitoring systems make it possible to implement more complex and efficient algorithms for determining the facts and locations of leakage. Thus, [1] data on the use in Toronto (Canada) of a system of "smart" sensors and artificial intelligence to determine "anomalous" noise in the pipeline, different from the norm and corresponding to leakage, are given. The server of the system analyzes the data received from the sensors and sends the result to the central control room, which automatically forms a request for repair and sends the repair team.

In general, the above acoustic monitoring tools make it possible to make an introduction that in this area, as in many others, there is a transition from traditional "semi-intuitive" methods of work to the use of intelligent computing systems that can allow:

- Reduce the routine burden on operators;
- Reduce the time to find and clarify the place of leakage;
- Reduce the cost of repairing communications.

To implement intelligent processing algorithms, it is necessary to study the defining characteristics of leakage signals in the time and spectral domain. The obtained characteristics will make it possible to assess the hardware characteristics of the complex of processing devices and the necessary methods and algorithms for decision-making.

3. Experimental Studies and Analysis of the Characteristics of Acoustic Emission Signals

For experimental modeling of various leakage options, a semi-natural stand has been developed and is used, which is a pipe at several points of which ball valves are installed to simulate the leakage of adjustable velicina. At the ends of the pipe there are areas for installing acoustic sensors. Pressurized water is supplied to the pipe from the water supply network. To receive acoustic emission signals, the Iscor-305 correlation leak detector is used, which includes acoustic sensors, amplifiers-transmitters for transmitting signals over a radio channel and a correlation receiver for receiving and processing signals. The input of an external sound card is connected to the line output of the receiver, Connected to a PC. A general diagram of the installation for the experiment is shown in Fig. 6.

In the experiment, signals were recorded alternately at different leakage values under the following conditions:

- in the absence of leakage (there are only noises - transport, industrial, natural);
- Leak source #1 is included;
- Leak source #2 is included.

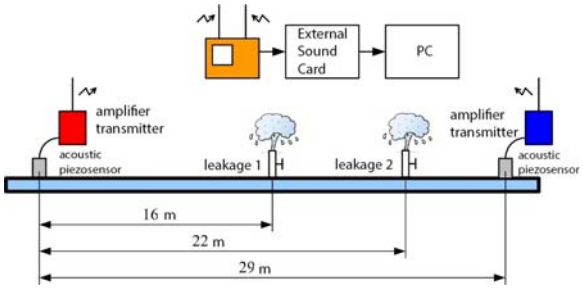


Fig. 6. Installation diagram for simulating and recording acoustic emission signals.

The spectra of the recorded signals are shown in Fig. 7 and have the following features:

1) The spectrum of the non-leakage signal is more uniform and has at least two regions with a maximum amplitude - low-frequency, with a central frequency of ~200 Hz, and a high-frequency one of about 1 to 4 kHz, with the maximum value of the spectrum in the low-frequency region slightly exceeding the maximum value of the spectrum in the high-frequency region;

2) The spectrum of the signal in the presence of leakage is more uneven, in it the level of the spectrum in the low-frequency region is always less than in the high-frequency, and this difference depends on the amount of leakage - the greater the difference, the weaker the leakage.

To assess the change in the spectrum of the signal over time, the spectrograms shown in Fig. 8 are constructed. It is clearly visible that there is no noticeable change in the signal spectra over time, and minor changes in the signal without leakage can be considered as a sign of the presence of short-term interference from sources of transport, industrial or natural noise. It is also clearly visible that when the amount of leakage increases, the leakage signal spectrum shifts to the low-frequency region.

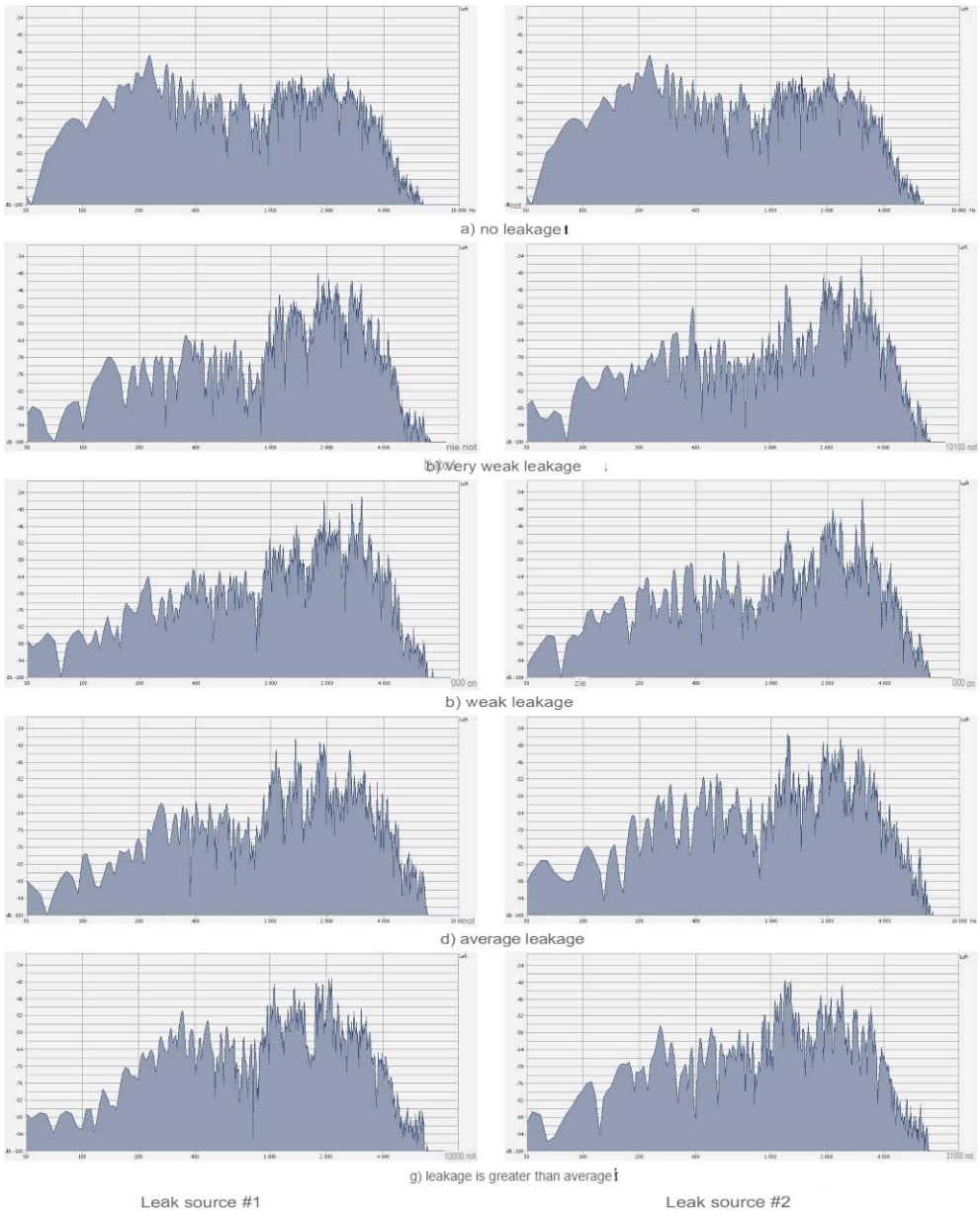


Fig. 7. Spectra of recorded acoustic emission signals.

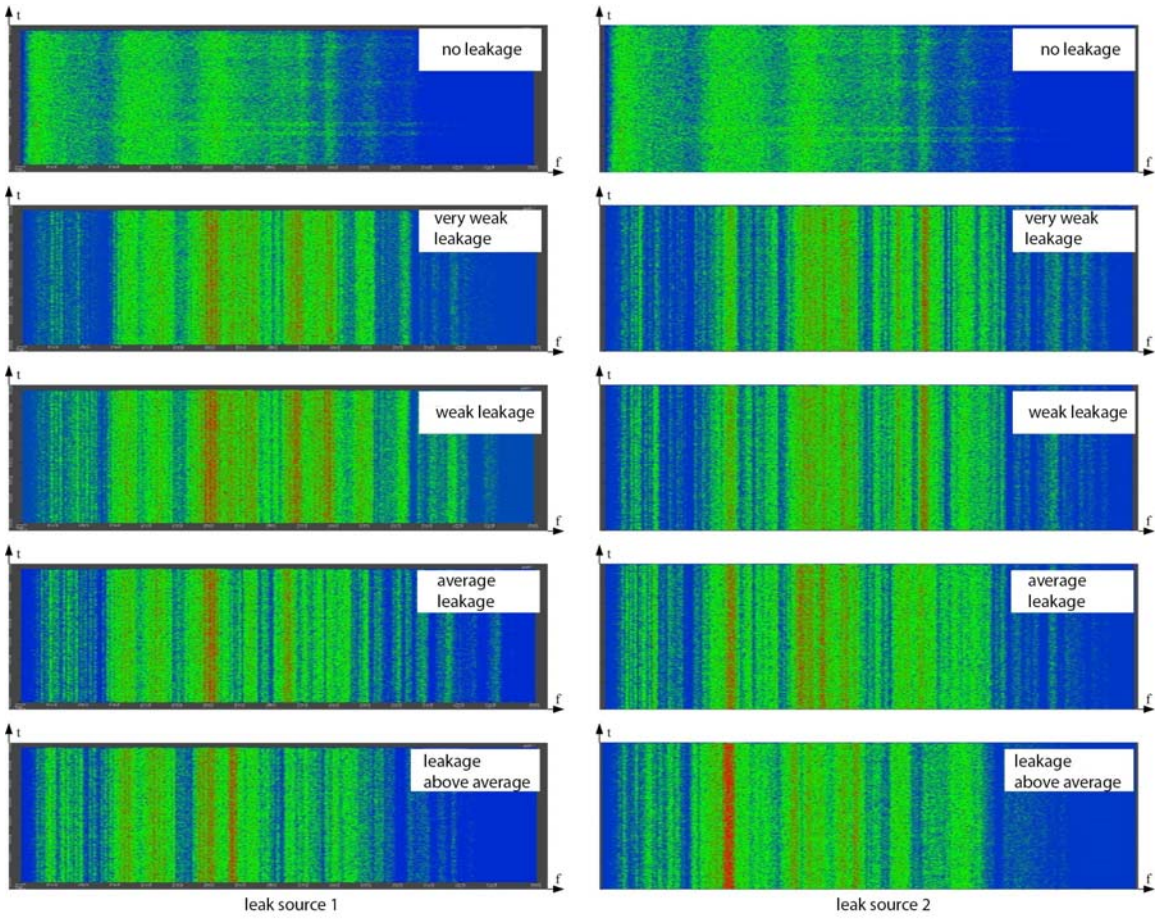


Fig. 8. Spectrograms of acoustic emission signals.

If spectral analysis shows how much the signal is "similar" to a set of sinusoids, then correlation analysis allows you to assess the similarity of the signal with itself, i.e. to what extent it and the physical process that generates it are random. The autocorrelation functions (ACFs) of the recorded acoustic emission signals computed in MATLAB 2012 are shown in Fig. 9.

The following features of the ACF obtained can be noted:

1) In the presence of leakage, ACF has a more periodic structure, with a shorter period for weaker leaks, and a longer period for stronger ones;

2) In the presence of a leak, the first side petal of the ACF (marked in the figure with a red circle) is noticeably higher than that of the ACF signal without leakage. Also, the level of the side lobe of the ACF changes slightly when the amount of leakage changes.

4. Conclusion

The article discusses the main varieties of equipment for acoustic leakage - acoustic, correlation and multi-position leak detectors and geographically distributed acoustic monitoring systems, presents the results of spectral and correlation analysis of various leakage signals and assesses the possibility of using ANNs to determine the fact of the presence or absence of leakage.

Based on the results obtained, it can be preliminarily concluded that the marked features of the spectrum and the autocorrelation function of the acoustic emission signal make it possible to determine the fact of the presence or absence of leakage. When leakage occurs, the spectrum and ACF of the signal acquire a characteristic form, which allows us to talk about the feasibility of their recognition using neural network algorithms.

References

[1]. https://www.everest.ua/ru/v-kanade-potery-vody-yz-gorodskyh-trub-obnaruzhyvayut-s-pomoshhyu-yntellektualnyh-datchykov, 26. 02.2021.

[2]. https://files.stroyinf.ru/Data1/46/46843/index.htm#i21807, View date 26. 02.2021.

[3]. http://www.vstmag.ru/ru/archives-all/2010/2010-4/263-obnaruzhenije-skrytyh, View date 26. 02.2021.

[4]. https://www.waterworld.com/technologies/flow-level-pressure-measurement/article/16190942/revolutionizing-remote-leak-identification, View date 26.02.2021

[5]. https://www.waterworld.com/home/article/14070815/how-to-nix-nonrevenue-water-with-acoustic-leak-detection, View date 2/26/2021.

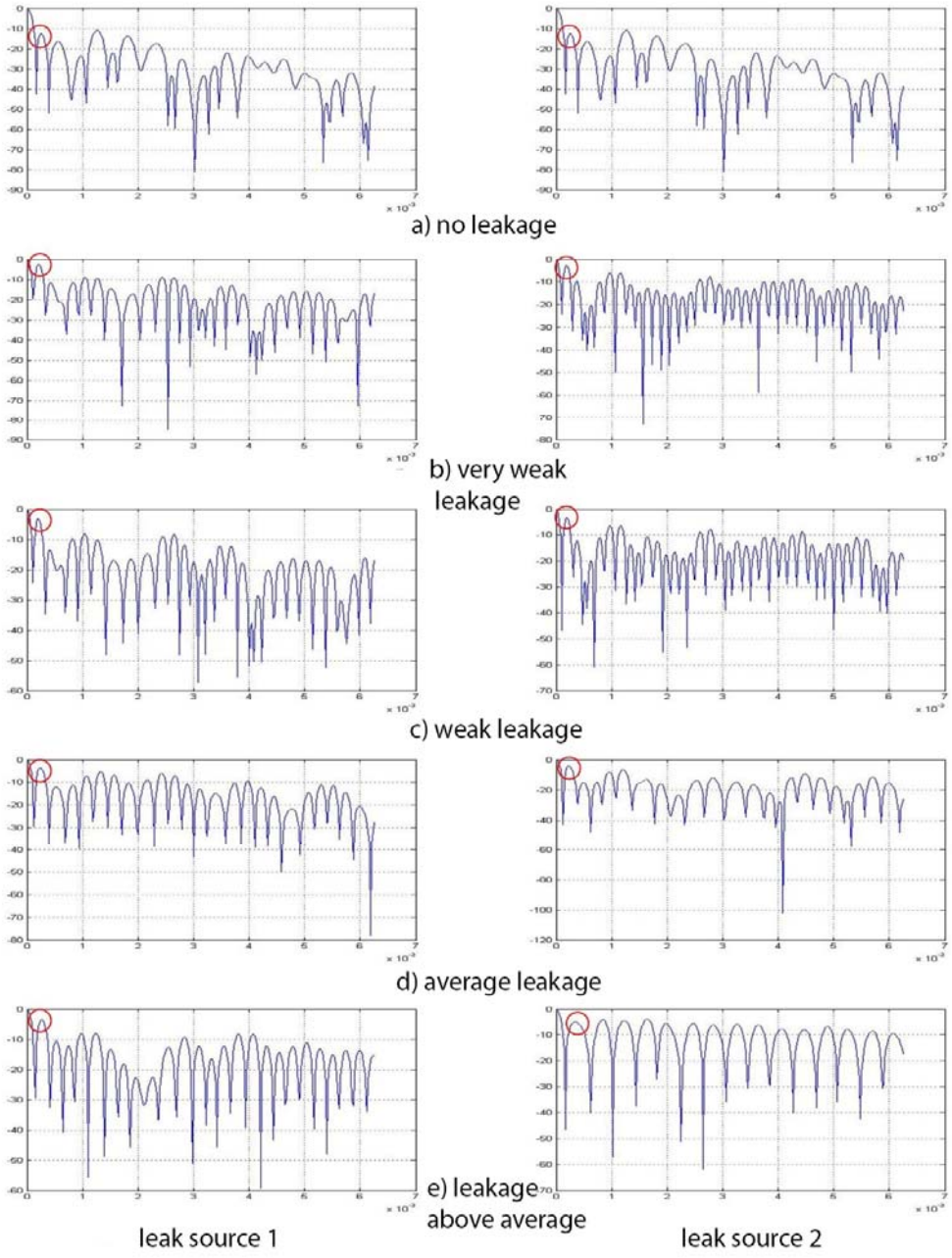


Fig. 9. Autocorrelation functions of acoustic leakage emission signals recorded in the experiment.