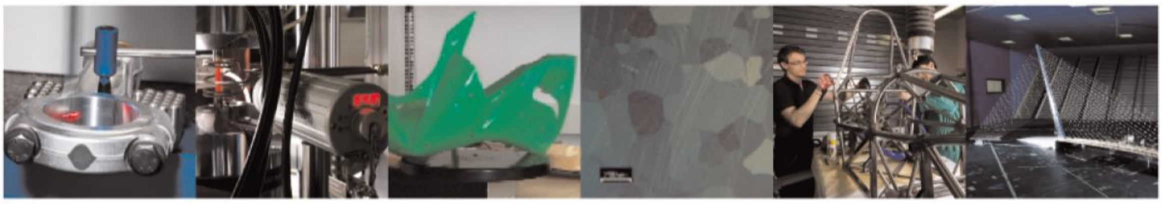

A human modelling and monitoring approach to support the execution of manufacturing operations

Urgo, Marcello; Tarabini, Marco; Tolio, Tullio

This is a post-peer-review, pre-copyedit version of an article published in CIRP Annals. The final authenticated version is available online at: http://dx.doi.org/10.1016/j.cirp.2019.04.052

A human modelling and monitoring approach to support the execution of manufacturing operations.

Marcello Urgo[1,*], Marco Tarabini[1], Tullio Tolio

1 Politecnico di Milano, Department of Mechanical Engineering, Milan, Italy.

Human workers have a vital role in manufacturing given their adaptability to varying environmental conditions, their capability of judgment and understanding of the context. Nevertheless, the increasing complexity and variety of manufacturing operations ask for the exploitation of digital technologies to support human workers and/or facilitate their interaction with automation equipment. The proposed approach uses artificial intelligence for image processing to identify the actions of the workers and exploits the knowledge related to the processes through hidden-Markov models to identify possible errors, deviations from the planned execution or dangerous situations. An application case is provided for assembly operations to assess the viability of the proposed approach in realistic conditions.

Man-machine system, Modelling, Monitoring

1. Introduction and problem statement

In many application areas, humans are a primary source of flexibility through their versatility in working on different and heterogeneous activities, adaptability to varying environmental conditions, capability to learn and improve their skills, capability of judgment and understanding of the context. Due to this, despite the impressive adoption of automated equipment, human workers have an important role in manufacturing, in particular in the execution of processes characterized by a high level of variability (due to personalization and/or high variety of the products) or high complexity/low formalization (e.g., complex Manufacturing-To-Order or high-quality handcraft production). These requirements, constituting relevant but limited market niches in the past years, are becoming some of the most relevant challenges for companies pursuing high added-value manufacturing [1] . Approaches designed to improve the performance in these manufacturing environments constitute a very relevant objective for industrial companies.

Nevertheless, exploiting the high degree of flexibility and adaptability of human workers also increases the probability to introduce errors in comparison to the stability and reliability of the behaviour of automatic equipment. At the same time, the integration of human workers and automatic equipment in the same working space poses concrete and relevant performance and safety challenges [2] . To this aim, monitoring systems can play a role, providing the capability of understanding what a human operator is doing and identify possible deviations from the ideal execution of a given process.

Advanced monitoring approaches are a major need in flexible and high-performance manufacturing systems. Although advanced tools and approaches are available for machines, taking advantage of a wide range of sensors and models [3] , the monitoring of human operators is a less developed area both in terms of research and industrial applications [4] . At the same time, the impressive advancement of digital technologies and artificial intelligence in this field, mainly pursuing applications in autonomous vehicles and security, provides an opportunity for the design of new approaches that were not feasible before [4] .

In this paper we address the design, pilot implementation and testing of a monitoring approach for manually executed manufacturing operations grounding on advanced vision technologies and artificial intelligence. The aim of the approach is to (i) monitor the execution progress of the process to infer whether it has been completed or assessing the fraction already executed and (ii) identify possible errors of the human worker, e.g., operations whose execution has been forgotten, (iii) raising the alarm in case of unsafe behaviour. The intended application is in high-variability manufacturing environments to support the workers in the execution of a wide range of tasks and their interaction with automation equipment, e.g., robots. An important requirement is the possible adoption in real operating environments, being able of coping with the intrinsic uncertainty and incomplete knowledge of real manually executed processes.

The paper is organized as follows: Section 2 presents an analysis of the literature; Section 3 describes the structure of the approach and the associated requirements which are then addressed in details in Section 4; in Section 5, the application to an industrial case is presented while Section 6 reports the conclusions and future developments.

2. Literature review

The modelling of human operators is a challenging research area due the difficulty to formalize and predict their behaviour. The modern trend towards flexible automation in manufacturing processes raises new challenging problems in this area with respect to environments where humans and machines have to operate and collaborate [1] [5] .

Traditional modelling approaches for monitoring manufacturing processes fail when dealing with tracking behaviours that are not deterministic. To overcome this limitation, generative models are defined in terms of stochastic parameters estimated from input data [6] . Hidden Markov Models (HMM) [7] [8] and Dynamic Bayesian Networks are among the most commonly used generative modelling methods. HMM, in particular, are among the most suitable tools for the monitoring of activities performed by humans in different application areas [8] [10] and for gesture recognition [9] . HMMs also provide the possibility of jointly considering the available knowledge (the ideal process) and the information coming from experiments (the output of tracking approaches) enabling mixed generative and discriminative approaches [6] . With regards to tracking the human movements, a very relevant trend is deep learning as a data driven technique for continuous human motion analysis [11] .

Within this increasing corpus of approaches, the tracking of the human pose has shown a rapid and significant development in the last few years taking advantage of deep learning image recognition methods to estimate the human pose in terms of legs, arms, hands and face. Tools exploiting these techniques have been developed by Google [12] within their *GoogleAI* initiative and Carnegie Mellon University [13] with applications to robotics. These advances provide the push towards the possibility of developing new classes of methods exploiting both the good performance of these tools and the higher detail of available information.

3. Monitoring and support human-executed processes

As described in Section 1, the aim of the proposed approach is to be able of tracking a human worker and to compare the results against the correct way of executing the process.

Considering the application in real working environments, the manufacturing process and the associated operations cannot be completely formalized. E.g., the operator could pre-empt operations to perform checks to guarantee the correct execution of the process; or operate additional actions like, picking-up and release tools multiple times, touching his/her head or body, adjusting the glasses, etc.

Since these actions are not necessarily forbidden and do not constitute a significant deviation from the correct execution of the process, a monitoring approach has to be able to operate in these conditions. Thus, being able to infer what the operator is actually doing grounding on a formal description of the process and coping with unformalized actions discriminating those not disturbing the process from real errors.

The reference architecture of the approach is shown in Figure 1. The movements of the human operators are tracked to identify the positions of his/her body and hands. A normal camera is used shooting a video of what is being executed. Relying on a simple video shot allows to avoid the need to wear specific sensors or markers and enlarge the range of applicability of the approach. The video is processed and the coordinates of the key points of the human body obtained.

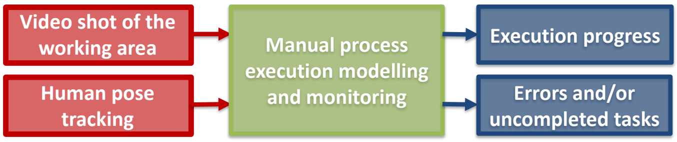

Figure 1. Structure of the proposed approach

The tracked positions are exploited by the monitoring approach that, grounding on a formal model of the process to be executed and the associated uncertainty, evaluates what the human operator is doing with the aim to: (i) assess the execution progress of the process; (ii) identify possible errors in the execution; (iii) identify possible unsafe situations.

4. Solution approach

The proposed approach consists of three steps: (i) the estimation of the human pose and the tracking of the movements in the captured video, (ii) the definition of a model for the execution of the process and the link with the information coming from the tracking approach and (iii) the use of the model to monitor the execution of an ongoing process.

4.1. Human pose estimation and tracking

The position of the operator is estimated by analysing frames of video recordings with OpenPose [13] a software for real-time multi-person tracking that detects human body, hand, facial, and foot key points [13] (Figure 3). The software output are the coordinates of key points for the people detected, together with a detection confidence d_c. This output is processed to censor unreliable data due to the possible presence of false identification of shadows or mirror reflections. The censoring of unreliable data grounds on the confidence provided by OpenPose and coherence constraints related to the human body, going outside the scope of this paper. If multiple operators are in the scene, the continuity of their tracking has to be guaranteed among different frames, since people are detected in random order by the OpenPose algorithm. Thus, a matching approach has been used to find the most likely matching between the different people identified:

1. In the first frame where at least one operator is present, an identifier is assigned to each person;
2. For each subsequent non-empty frame, the matching between the identifier and the people is verified and, if needed, modified to guarantee continuous tracking.
3. If a new person enters the scene, a new identifier is assigned while, if a person leaves the scene, its identifier is released and cannot be used anymore

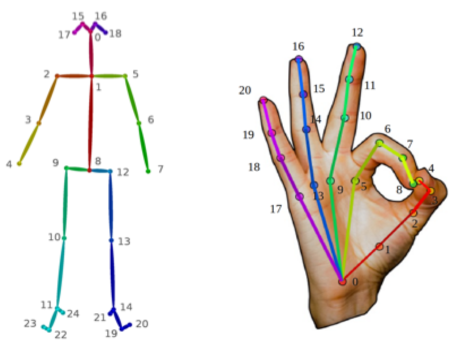

Figure 2. Body and hand modelling schemes for tracking [13] .

The matching of the estimated people in step (2) is based on the computation of the distance between the key points acquired in two consecutive frames. To link the pose estimation with the process to be executed, a set of working regions are defined in the area of the captured frame, hence, the coordinates of the operator's key points are checked against these regions. These regions must be defined in order to be consistent with the position of the hands of the human worker when executing the different operations. Thus, the previous output is translated into a sequence of presence of the key points in the regions.

4.2. Modelling of human-executed manufacturing operations

We consider the execution of a generic process by a human operator. We assume the operator has to follow a given structure of the process thorough the execution of a partially ordered set of operations. This means that the operator has some degrees of freedom in the execution of the process, thus the monitoring has to be able to take into consideration multiple alternative ways of executing it. To match these requirements, a Hidden Markov Model is defined, able to consider a state-based process definition together with the measurements of a set of variables related to the process execution. HMMs are probabilistic models, hence, are able to deal with uncertainties or noise in measured observations [6] linked to the underlying state-based process.

The formalization of the process starts with the definition of the states describing the manufacturing process, modelled through a set of states, $S = \{s, O_1, O_2, \ldots, O_k, t\}$, where states $\{O_1, O_2, \ldots, O_k\}$ represent a finite set of manufacturing operations belonging to the manufacturing process under analysis, while *s* and *t* are dummy states modelling the beginning and completion of the process. Precedence relations as well as the execution times for the operations are modelled through the definition of transitions among the states in *S*. The tracking approach described in Section 4.1 analyses every single frame of the captured video, as a consequence, the output assumes a discretization of the time defined by the number of frames per second in the video. At each time unit, $a_{i,j}$, defines the probability to move from state *i* to state *j* given that the process is in state *i*. An example is provided in Figure 3 showing the modelling of a very simple process consisting of three operations (O_1, O_2 and O_3) that must be executed in series. The model in Figure 3 is a discrete-time Markov Chain whose execution generates a sequence of states in *S*. It is a stochastic model; hence the sojourn time in the states is not deterministic, and it also has the Markov property: the probability of moving to the next state only depends on the current state, independently from the number of time units already spent in it.

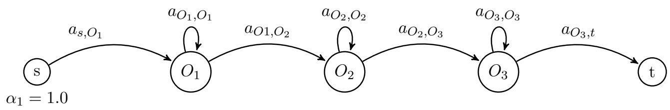

Figure 3. A Markov Chain model of a simple process.

Although the tracking approach (Section 4.1) is able to identify the position of the human body in the scene in terms of the presence of a set of key points in a given area, it cannot provide any information about what the operator is actually doing

Modelling the link between the position of the body (e.g., the hands) and the process in execution takes advantage of the features of HMMs. In a HMM, the states in *S* are hidden, i.e., cannot be observed. On the contrary, they can be monitored through an emission, i.e., observable variables with support $E = \{e_0, e_1, \ldots, e_m\}$. An emission e_i can be a single value or an array of values, providing the possibility to consider a wide range of observable variables in a single emission model. The link between the execution of the operations in *S* and the associated emission in *E* is defined through the emission probability $b_{i,e}$, i.e., the probability of observing emission e_i when the process is in state *i*. The HMM associated to the process modelled in Figure 3 is represented in Figure 4.

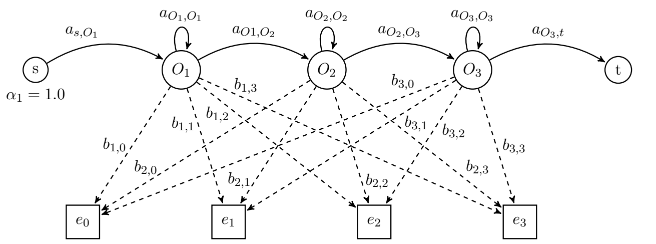

Figure 4. Hidden Markov Model of a simple process.

Hence, for each activity of the process to be executed (represented as a state in *S*), a set of emissions can be defined in terms of the presence of the hands of the operator in a set of areas in the frame. The estimation of the transition and emission probabilities ($a_{i,j}$ and $b_{i,e}$) for a given process to monitor is known as the training of the HMM. The main approach for training is exploiting a set of paired states and emissions sequences coming from correct executions of the process, to estimate the value of $a_{i,j}$ and $b_{i,e}$. Also, the ideal process is exploited for the training.

Starting from the model of the process in Figure 3, together with the estimation of the processing times, ideal executions are sampled together with the associated sequence of generated emissions. These data, together with the ones coming from the experiments, are used to estimate the parameters in the HMM.

4.3. Monitoring a human-executed process

The above-described model can be used to support the monitoring of a process through the observation of the sequence of emissions obtained through the tracking of the human body. This is equivalent to perform a decoding of the emission of an HMM. Thus, given a sequence Θ_E of emissions observed, find the most probable sequence of states Θ_S. This operation is performed using the Viterbi algorithm [7] also returning the probability δ associated to Θ_S.

If δ is low, it means that, although the observed sequence of emissions is compatible with the process, its likelihood is low and, consequently, could signal a wrong execution of the process. If δ is zero, then the observed sequence is impossible or, if the model has been properly defined, identifies the presence of people or their hands in forbidden regions.

5. Industrial application

To assess the viability of the proposed approach we consider an industrial application in the manufacturing of mechanical components and, specifically, the load/unload of a part onto/from the fixtures, whose monitoring is highly recommended to avoid possible errors (e.g., wrong parts mounted onto the fixtures and/or missing clamping operations) in case of high variety of parts. We focus on a set of operations by executing and monitoring them in a realistic environment. The aim is to infer whether the load/unload process has been completed and identify missing operations or unsafe behaviour.

The process under study entails a set of operations, e.g., clamp/unclamp fixtures, unload/load parts from the fixtures, pick-up and release tools, etc. A detail of the fixtures and part for the considered application is shown in Figure 5, together with an example of a possible sequence of operations for the process. The detailed list of operations is reported in Table 1.

Table 1. List of assembling operations and requirements.

Operation	Description	Region Requirements
PC	Pick-up clamping tool	At least one hand in A
RC	Release clamping tool	At least one hand in A
PRP	Pick-up raw part	At least one hand in C
LRP	Load raw part	At least one hand in D
UWP	Unload worked part	At least one hand in D
RWP	Release worked part	At least one hand in B
C1	Clamp fixture 1	One hand in E and one in D
U1	Unclamp fixture 1	One hand in E and one in D
C2	Clamp fixture 2	One hand in E and one in D
U2	Unclamp fixture 2	One hand in E and one in D

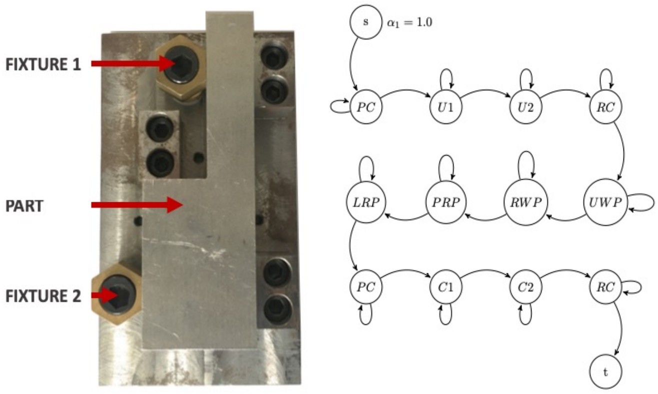

Figure 5. The part type and fixture under study and a possible execution of the load/unload process.

The 25-point body and 21-point hand schemes provided in the OpenPose library [13] are used [13]- For each captured frame, the positions of the key points are given back in terms of their coordinates in the frame itself. Starting from these coordinates, the movements of the worker's hands, namely key point 10 (Figure 2) for both the right and left hand, are mapped onto working regions in the captured frames. An example of the result of the tracking is provided in Figure 6, showing the identification of the key points for the human body and the overlay of the working regions. Thus, the output of the tracking is a sequence of observed emissions for the monitored process in terms of the regions where the hands have been tracked in all the frames, e.g., {A}, {D, E}, etc. Notice that emissions can contain multiple regions for the same hand, e.g., the right hand could be in region D and E at the same time.

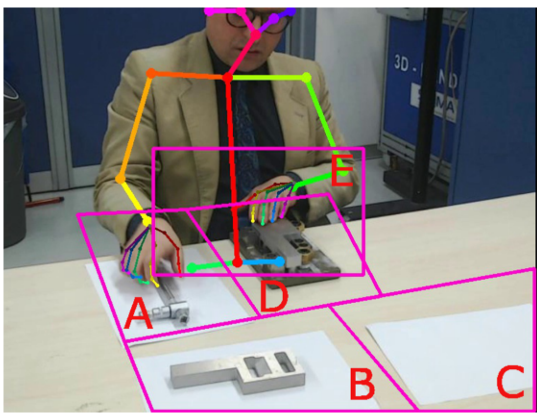

Figure 6. Human body tracking and working regions.

The definition of the HMM (Figure 4) grounds on an ideal process based on the operations in Table 1 and the associated emissions, in terms of the presence of the hands in the working regions. Multiple correct execution modes for the assembly process are taken into consideration by defining alternative sequencing of the operations, to cover the different possible behaviours of the human operators. By monitoring these executions of the assembly process, executed by different operators, an estimation of the emission probabilities of the HMM is obtained. This training of the HMM is further enriched taking advantage of ideal executions of the process, obtained by sampling the processing times for the different operations and deriving the positions of the hands according to the requirements in Table 1.

The testing of the approach has been carried out tracking multiple executions of the same process, both correct and not correct, done by different operators. As anticipated, two different execution modes of the process have been considered and, for each of them, an HMM has been instantiated and trained with 10 correct monitored executions plus 5 ideal executions, and then used to monitor 25 experiments in total.

The identification of potentially dangerous behaviours is not implemented in this realistic use case but can be easily addressed checking for the presence of the hands in forbidden regions.

The results are shown in Table 2 considering three different application scenarios. The first one (A) refers to the monitoring of a correct process, to guess its progress execution and comparing it with the real one. The monitoring capability is good at the beginning of the monitoring, being able of correctly guessing what the operator is doing in 65% of the cases, while the process is at 10% of its progress. Then it goes worse (due to the high variety of the behaviours of the operators) and improves towards the end of the process. Notice that, the decrease in the monitoring capability, between the 95% and 100% of the process advancement, is due to the fact that different executions require different times and, at the end, the operator is still being tracked disturbing the monitoring capability. In one case only, the approach recognized a correct process as completely wrong.

Scenario B refers to the capability of identifying a process deviating from the correct execution at 30% of its progress. In all the cases, the approach was able to identify the problem at about half of the execution. Finally, scenario 3 addresses the monitoring of a completely wrong process. Also in this case, the approach is able to recognize the occurrence of a wrong process at about half of the execution while at the beginning, as expected, no process is labelled as wrong.

Table 2. Results

A. Identify the status of correct process executions (15 tests) B. Recognize an error happening at 30% of the execution (5 tests) C. Recognize completely wrong process executions (5 tests)								
		Progress of the time horizon monitored						
		10%	**25%**	**50%**	**75%**	**90%**	**95%**	**100%**
Success Rate	**A**	0.67	0.33	0.53	0.67	0.80	0.80	0.73
	B	0.40	0.40	1.00	1.00	1.00	1.00	1.00
	C	0.00	0.60	1.00	1.00	1.00	1.00	1.00

6. Conclusions

We presented an approach for monitoring of manually executed processes able to cope with the intrinsic uncertainty and incomplete information of real environments and tested on a real case demonstrating good performance in assessing whether the process has been completed or not as well as raising alarms if required operations have not been executed. Avoiding false alarms is a relevant aspect for the effective adoption in real manufacturing environments.

Future research will address a more extensive training and testing. Beside this, future research directions to improve the features of the proposed approach are the optimization of the best camera position for the process to monitor and the development of more advanced models explicitly considering pre-emption and alternative process executions.

References

[1] Krüger, J., Lien, T.K., Verl, A., 2009, Cooperation of human and machines in assembly lines, CIRP Annals, 58/2,:628-646.

[2] Pellegrinelli, S., Orlandini, A., Pedrocchi, N., Umbrico, A., Tolio, T., 2017, Motion planning and scheduling for human and industrial-robot collaboration, CIRP Annals, 66/1:1-4.

[3] Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K., 2016, Cyber-Physical Systems in Manufacturing, CIRP Annals, 65/2:621–641.

[4] Küpper, D., Lorenz, M., Kuhlmann, K., Bouffault, O., Yew Heng, L., Van Wyck, J., Köcher, S., Schlagete, J., AI in the Factory of the Future - The Ghost in the Machine, 2018, The Boston Consulting Group.

[5] Wang, X.V., Kemény, Z., Váncza, J., Wang, L., 2017, Human–robot collaborative assembly in cyber-physical production: Classification framework and implementation, CIRP Annals, 66/1:5-8.

[6] Hilary Buxton, Learning and understanding dynamic scene activity: a review, Image and Vision Computing, 21(1), 2003, 125-136.

[7] Robert J Elliott, Lakhdar Aggoun, John B. Moore, Hidden Markov Models - Estimation and Control, 1995, Springer-Verlag New York.

[8] Jie Yang, Yangsheng Xu and C. S. Chen, "Human action learning via hidden Markov model," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 27, no. 1, pp. 34-44, Jan. 1997

[9] Yoon, H.S., Soh, J., Bae, Y., Yang, H.S., 2001, Hand gesture recognition using combined features of location, angle and velocity, Pattern Recognition, 34(7), 2001, 1491-1501,

[10] Tran, D. T., Sakurai, R., Yamazoe, H., & Lee, J. H. (2017). Phase Segmentation Methods for an Automatic Surgical Workflow Analysis. *International journal of biomedical imaging*, *2017*, 1985796.

[11] Wang, P., Liu, H., Wang, L., Gao, R.X., 2018. Deep learning-based human motion recognition for predictive context-aware human-robot collaboration, CIRP Annals, 67/1:17-20.

[12] Ganapathi, V., Plagemann, C., Koller, D., Thrun, 2012, Real-Time Human Pose Tracking from Range Data, ECCV.

[13] Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., Sheikh, Y., 2018, OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields, arXiv:1812.08008.