World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:3, No:10, 2009

Worker Behavior Interpretation for Flexible Production

Bastian Hartmann, Christoph Schauer and Norbert Link

***Abstract*—**This paper addresses the problem of recognizing and interpreting the behavior of human workers in industrial environments for the purpose of integrating humans in software controlled manufacturing environments. In this work we propose a generic concept in order to derive solutions for task-related manual production applications. Thus, we are able to use a versatile concept providing flexible components and being less restricted to a specific problem or application. We instantiate our concept in a spot welding scenario in which the behavior of a human worker is interpreted when performing a welding task with a hand welding gun. We acquire signals from inertial sensors, video cameras and triggers and recognize atomic actions by using pose data from a marker based video tracking system and movement data from inertial sensors. Recognized atomic actions are analyzed on a higher evaluation level by a finite state machine.

***Keywords*—**activity recognition, task modeling, marker-based video-tracking, inertial sensors.

I. Introduction

INDUSTRIAL processes in factories require a high degree of efficiency for the production of competitive products. Therefore, production lines always have been constructed with aiming at efficiency by automation of frequently repeated tasks. However, the rather stiff organization of automated production lines brings along the problem of lacking flexibility, which in turn costs a lot of their efficiency in situations of changes. To overcome these issues and to make production more flexible the EU project XPress (IP026674-2) aims at developing a concept for flexible production by using intelligent software agents, based on an approach called “the Expertonic factory”. One of the major aspects in this project is the seamless integration of human workers in software controlled production processes. Thus, flexibility of humans can be used to support automated production still enabling control by intelligent machines.

The Integration of human workers in a software-controlled factory requires a bi-directional interface in order to allow flows of information from machine to human and from human to machine, respectively. The flow of information in direction from machine to human worker is addressed by human-machine-interface devices and methods, ranging from “low level” solutions like textual task descriptions on computer screens up to “high level” solutions such as augmented reality task presentations. Besides design issues and augmented reality problems, these methods mainly deal with the question of converting and presenting computer readable task information in forms, which are understandable by human workers. Designing an interface for the flow of information from human to machine even seems to be a more complex problem, if it is desired that information about human worker activities is acquired automatically in order to have a feedback for the software-controlled factory and to support the worker by real-time observation and guidance of the performed task.

Our work addresses the problem of automatically interpreting task-related human behavior in order to create an interface from human worker to the intelligent factory. The concept for designing solutions for worker behavior interpretation should generally be flexible and not restricted to specific tasks or working processes. Therefore, our concept is focused on a generic description of the process of behavior interpretation that enables us to derive solutions with re-useable components, which may be used for several applications. As a demonstration of the concept we present a functional sample, which is related to a resistant spot welding scenario. In this scenario primitive activities of a human worker are recognized and the resulting behavior, consisting of series of primitive activities, is interpreted automatically.

The remainder of this paper is organized as follows: In the following section we discuss a selection of state of the art work related to the topic of human behavior interpretation and human machine cooperation. Our conceptual approach of interpreting human behavior is presented in section III. The proposed concept has been exemplified by means of a functional sample, which is explained in section IV. In section V, experiments and results from tests with the functional sample are shown and discussed. Finally, conclusion on our work is drawn and an outlook for further work is given in section VI.

II. Related Work

The research field of reasoning about human behavior is drawing increasing attention and comprises a lot of work from other areas of research related to this topic, such as computer

Bastian Hartmann is with the Institute of Applied Research, University of Applied Sciences Karlsruhe, Karlsruhe, Germany (e-mail: Bastian.Hartmann@hs-karlsruhe.de).

Christoph Schauer is with the Institute of Applied Research, University of Applied Sciences Karlsruhe, Karlsruhe, Germany (e-mail: scch1016@hs-karlsruhe.de).

Norbert Link is with the Institute of Applied Research, University of Applied Sciences Karlsruhe, Karlsruhe, Germany (e-mail: Norbert.Link@hs-karlsruhe.de).

vision, wearable and ubiquitous computing, information fusion, machine learning and artificial intelligence. Typical application areas in which human behavior is observed or interpreted are: security, military, medical engineering and healthcare, robotics and human computer interaction. Compared to these typical applications, research projects on industrial applications are still few, but seem to increase lately.

Before reasoning about human behavior, information about humans and their activities has to be acquired. Detecting and tracking humans or parts of the human body is a very broad and active research field. Especially in the field of visual human motion analysis much research has been done and work related to this topic has been well reviewed by [4], [1] and [8]. Besides visual information, measurements from wearable sensor devices are used as well to obtain pose information or to track objects or humans. Applications in this area often use inertial sensors [10], which may be combined with other information from wearable sensors such as ultrasonic distance measures [17]. Another work combines visual information with measurements from wearable sensors by data fusion in order to track industrial tools [11].

Basing on the acquired information about humans or objects, which are utilized by humans, reasoning about human behavior can be performed. In [2] position information taken from video sequences of an office environment was used to recognize the execution of tasks by means of a state machine model. Research on similar indoor environment applications was done by [9] and [3], using an Abstract Hidden Markov Memory Model and a classification method. The work of [12] comprises a system for human behavior understanding in video sequences based on a hierarchical combination of non-parametric database sampling and parametric models.

An approach based on wearable sensor information was presented by [15]. In this work different classifier techniques were used to detect bicycle repair tasks from ultrasonic and motion signals. In [18], activities in a wood workshop were classified from acceleration and sound information. Furthermore, in [16] wearable sensors have been used to track activities in a car manufacturing scenario.

An interesting formal approach has been presented by [14], which incorporates a model of human task-performing processes allowing classification of human material handling tasks for industrial manufacturing. However, this approach does not incorporate how human actions are detected or information about humans is derived.

III. CONCEPTUAL APPROACH

The reviewed work concentrates on the analysis of specific human activities, which results from the fact that recognition and interpretation of human behavior is a very complex problem. However, in our work of integrating humans in software-controlled manufacturing we need to have a more general description of the interpretation of task-related worker behavior, which is able to provide a certain degree of flexibility in form of re-useable components covering different application scenarios in manufacturing. Therefore, we apply a top-down oriented concept for finding solutions for the interpretation of human worker behavior. The worker tasks in industrial production on which our work is focused on fall into the following categories:

- Handling task: Characteristic for this category of tasks is that specific degrees of freedom, such as pose or state, of an object (e.g. a tool or workpiece) are changed over time by the worker.
- Assembly task: In this category, devices are assembled from workpieces in a logic-sequential manner of task steps. Trained workers often perform substeps in these tasks in a gesture-like way.
- Maintenance task: Similar to assembly tasks, tasks of this category are performed in a sequence of actions according to a maintenance plan. However, there are some particular actions which appear to be more abstract than steps in an assembly task (such as visual inspections).

In regard to the state of the art in the field of activity recognition and behavior interpretation and according to our problem, interpretation of human behavior can be described as a hierarchical chain of several processing instances, as it is shown in Fig. 1. The different levels of abstraction on which signals and data are processed in order to reason about behaviour of the complexity of tasks are explained in the following:

- On the lowest level, sensor data has to be acquired from appropriate sensor devices, in order to have physical information for further processing. This physical information can be any information related to worker activities, such as visual information form camera images, spatial movement signals, switches indicating tool usage, etc. The sensor data level includes measuring raw data as well as preprocessing such raw data, thus offering a suitable input for superordinate processing levels. Sensor devices may be subdivided into categories with complementary properties, namely wearable and remote sensor devices. Remote devices (cameras, laser scanners, ultrasound, etc.) are sensors, which provide measurements with a fixed relation to the environment or a world coordinate system, but are in turn prone to erroneous environmental influences such as occlusion or reflections. In contrast to that, wearable devices (accelerometers, gyros, speedometers, switches etc.) do only provide measurements depending on the internal state of the measured object and are not prone to external erroneous influences.
- After acquisition and preprocessing of relevant sensor data, sophisticated signal processing methods may be used to enhance raw data or to derive information, which is not directly available from raw sensor data (e.g. position data from images). Furthermore, data from several sensor sources can be processed in a combined way by using multi-sensor integration or sensor fusion methods [7].
- Activity recognition methods either use raw or processed

sensor signals for recognizing primitive human actions (such as grasping an object or moving to a specific position) in form of symbolic data or probabilities. In our work we distinguish between two classes of methods: time variant methods incorporating temporal variations (e.g. by time series analysis [20,21]) and time invariant methods, which are based on classification of features calculated from subsumed signals (e.g. classification of features derived from sliding window filtering [19]). The major difference between these classes of methods is that time variant methods regard temporal patterns (for instance of gesture-like actions) and that time invariant methods are able to generalize signal parts containing less temporal information. For some applications, a combination of techniques of both categories may be reasonable in order to benefit from their particular advantages. Activity recognition often requires a high degree of adaptation to a specific problem (e.g. training).

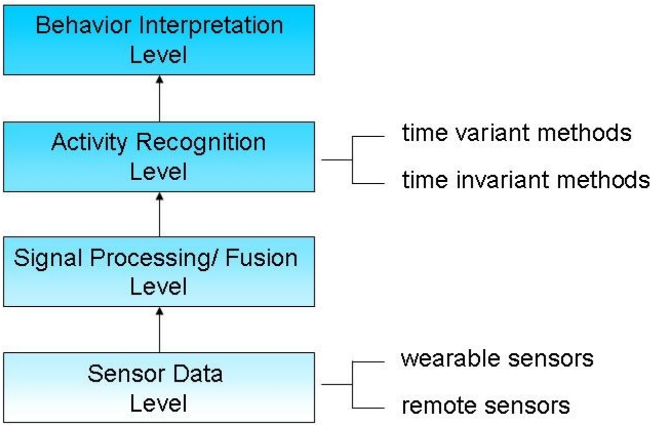

Fig. 1 Process instances for interpretation of human worker behavior

• On the behavior interpretation level, task-related worker behavior is analyzed. The intention is to detect if tasks are executed correctly or if there occur any non-conformities. Thus, behavior on this level represents the complete worker task (such as handling tools for working on work pieces or assembling a product from several parts) and is modeled by interrelated atomic actions from the subordinate level. In our work we define worker tasks as sequences of task steps or atomic actions. According to our knowledge, most of the worker tasks in industry can be described in a graph-like manner (e.g. assembly schedules). Thus, worker behavior can be modeled by using state based methods which may be based on either deterministic models (state machines, Petri nets) or probabilistic models (Hidden Markov Models). As well as on the Activity recognition level, the behavior model strongly depends on the specific task to be observed.

Because of the complexity of the problem of interpreting human behavior, solutions for applications have to be adapted to a specific problem or task to be interpreted. However, since our work addresses flexible production, we use a conceptual formulation, which works as a generic framework for the design of solutions that should provide flexibility in form of a high degree of re-useable components. This generic framework consists of levels of process instances according to Fig. 1 and a top-down oriented concept for the design of task-specific solutions.

Generally, there are two ways to approach a specific human worker behavior interpretation problem. The commonly used bottom-up approach starts with the data available from the worker or process by specific sensor devices. Then appropriate methods how to recognize actions and how to reason about human worker behavior are applied to these data. The alternative way, which we favor in here, is to approach the problem in a top-down-oriented manner. The task to be monitored is first split up into several interrelated task steps, which are essential for the task execution of the task or which indicate eventual non-conformities. All task steps are represented by related atomic actions that have to be recognized. Then, methods for the recognition of these atomic actions and, furthermore, adequate sensors and processing methods for the sensor data have to be chosen.

There are several advantages and drawbacks associated with both approaches. On the one hand, it sounds logically that desired information determines what methods and devices should be used and that a top-down approach is able to provide a higher degree of re-usability due to structured components. However, on the other hand performance of methods and appropriateness of sensor devices determine which information may be derived. Therefore, we treat the problem of finding appropriate techniques for each level of process instances by including both ways of thinking in our conceptual problem formulation for the interpretation of worker behavior, which is illustrated in Fig. 2.

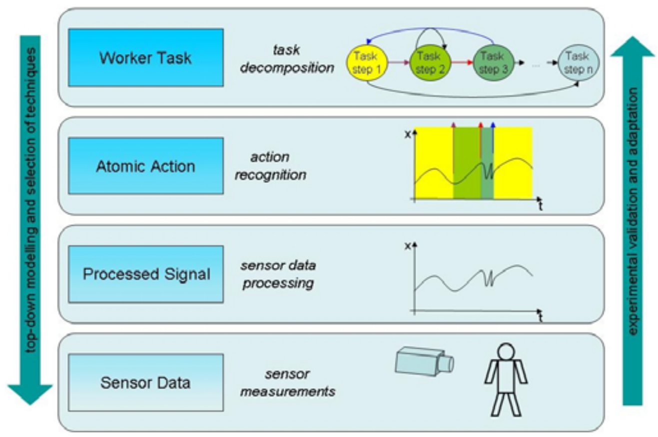

Fig. 2 Approaching the problem of finding appropriate techniques for process instance levels

Here we start in the described top-down manner with modeling of the worker task, defining atomic actions and choosing processing methods and sensor devices. Then eventually modifications of models and methods used in the higher process instances have to be made in the case of the un-availability of required signals or sensor data. Finally, experimental system testing and validation eventually demand further method adaptations at the activity recognition and behavior interpretation level. With this approach we intend to find solutions for specific worker behavior interpretation problems which are flexible in a way that they contain structured components providing a high degree of re-usability.

It should be noted that boundary conditions for designing applications in industrial environments also have to be considered as additional factors. This mainly includes issues such as safety rules for human workers or legal regulations, e.g. data privacy policy.

IV. SAMPLE DEMONSTRATION OF THE APPROACH

After the formulation of our conceptual approach, a so called "functional sample" has been set up, which addresses the scenario of resistance spot welding with a hand welding gun. This handling task scenario represents a common task, which is executed by human workers in the automotive industry at workplaces where no robots are deployed.

For the development of this functional sample the proposed approach is applied in order to create such a behaviour interpretation system with structured and re-useable components.

A. Scenario and Worker Task Description

In our scenario we assume, that a task is being given to the worker, in which the worker should weld a car body door at predefined positions. Fig. 3 shows the tool used in this scenario (hand welding gun) as well as the work piece (car body door). As preconditions we assume that the car door is locked into position and that the positions of the welding spots are known in 3D space. In the scenario, we have five welding spots with defined positions related to the car door as indicated in Fig. 3 b). The worker task is to manipulate the six degrees of freedom (pose) of the welding gun tip (welding gun handling) in order to align the welding gun to the spot poses and pull the trigger on the welding gun to release a welding process. It does not have to be taken care about the spot welding process itself, since this is done by the welding gun controller. The objective of the behavior interpretation system is to observe if the worker executes the given task correctly after he has been instructed. Correct execution of the worker task has been defined as follows:

1) Pick up the welding gun from the storage position.
2) Move the gun to the first welding spot pose.
3) Start the welding process by pulling the welding gun trigger, when the welding gun is aligned to the spot (i.e. at the correct position and gun stable).
4) Repeat steps 2 and 3 for the other four remaining spots.
5) Finish the task by putting the welding gun back to the storage.

Besides observing the execution of the task, non-conformities in form of welding at wrong positions, welding when the gun is still moving and (re-) welding of already welded spots have to be detected and prevented. The quality tolerance for the welding spots has been defined as a maximum deviation of 15mm from the target position.

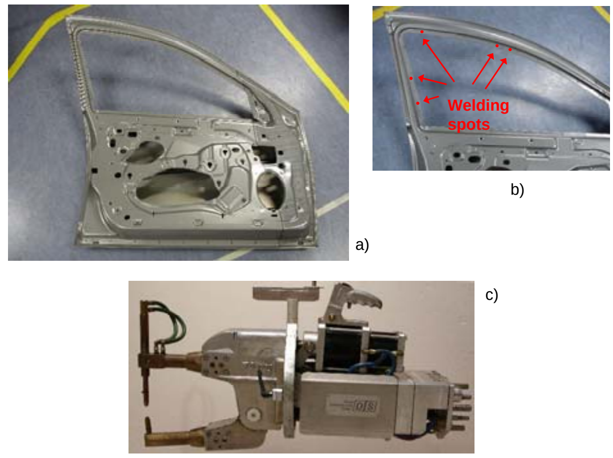

Fig. 3 a) Car body door used as work piece in welding gun scenario. b) Positions of the spots to be welded indicated on the car door. c) Welding gun.

B. Approach towards a Worker Behavior Interpretation System

In order to find an approach for interpreting worker behavior in the welding gun scenario, we proceeded according to our proposal from the previous section by first modeling the scenario in a top-down manner. It should be noted that in our scenario the worker himself must not be monitored (due to legal regulations of the automotive industry). Therefore, the tool rather than the worker has been selected as the target object to be observed for reasoning about worker behavior. Due to the strong relation between task execution and tool usage in this handling task scenario this means no serious limitation.

The decomposition of the worker task resulted in a model which is shown in Fig. 4 by the essential states of its related statechart. This statechart model comprises task steps as states and atomic actions as transitions between states. The initial state (wait for worker) represents the initial situation and indicates that the welding gun is located in the storage. In this state all guard conditions are set to value "true". By recognizing the atomic action that the gun has been picked up from the storage (action "remove gun from storage") it is recognized that the worker has started his task execution and is now moving the welding gun somewhere in the workspace (state "tool moving"). Next it has to be recognized, when the worker aligns the welding gun to the pose of a welding spot (state "tool at spot n"), then if the worker holds the gun stable (state "ready to weld spot n") and finally if the welding gun trigger has been pulled so that the welding is executed (state "welding spot n"). Additionally, incorrect behavior by pulling the welding gun trigger when not in state "ready to weld" is being modeled (state "misbehavior"). The end of the task execution is modeled by recognizing if the welding gun has been returned to the storage position and if all spots have been welded.

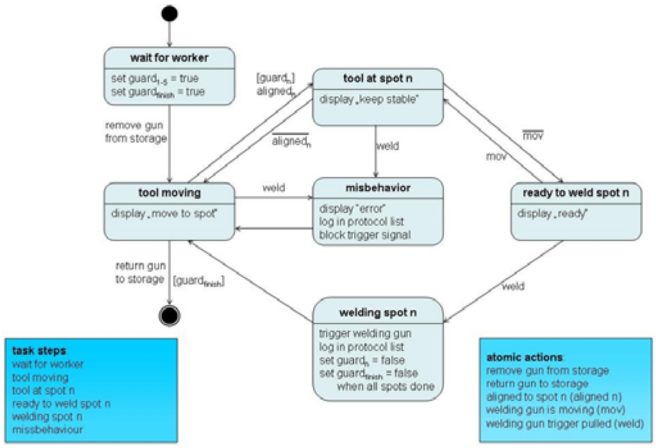

Fig. 4 Human worker behavior model for welding gun scenario

With this decomposition all task steps necessary for the correct execution of the task according to the worker task description have been modeled including the detection of misbehavior. In specific states output information is created by the system either for giving feedback to the worker (display) or for triggering the welding controller to make a weld.

Besides the recognition of atomic actions guard conditions ("$guard_{1-n}$" and "$guard_{finish}$") are used along with a protocol list in order to avoid welding of already welded spots and to ensure completion of the task.

After having decomposed the worker task and modeled dependency relations between task steps, it has to be defined how the atomic actions from Fig. 4 can be recognized. The outcome is straightforward and based on the recognition of position and motion of the welding gun and trigger signals, as listed in the following:

- Action 1: Remove gun from storage: storage trigger signal.
- Action 2: Return gun to storage: storage trigger signal.
- Actions 3.n: Aligned to spot n: in position at welding spot number n.
- Action 4: Welding gun is moving: motion detection.
- Action 5: Welding gun trigger pulled: welding trigger signal.

Due to the simplicity of the action, recognition of trigger values is realized by checking binary information from storage proximity switches and the gun trigger switch. Pose alignment and motion analysis, are carried out by distance measurements with position signals and thresholding of motion signals, respectively.

The remaining instances of the worker behavior interpretation system, which have to be selected, are the appropriate sensor devices for the sensor data level and, optionally, signal processing methods to derive input information for the activity recognition level from raw sensor data. The initial choice was for a marker based video tracking system to acquire position and attitude measurements of the welding gun from video camera data streams. However, video tracking is computationally expensive and provides low sampling rates at high latency. Furthermore, video cameras are remote sensor sources, which are prone to disturbances such as occlusions. Therefore, the idea was to support video tracking by wearable high-frequent measurements from inertial sensors, which measure accelerations and angular rates. This combination should be able to provide remote-sensed low frequent position information and high frequent internal movement information. At last, trigger signal information should be provided by binary switches.

C. *Methodology in Detail*

In the previous section an approach for our scenario has been presented in form of a sketch of methodology. For this methodology a detailed explanation is given in the following:

Sensor Data and Signal Processing: Video-Tracking System

In video-based tracking, position and attitude denoted as six degrees of freedom (6DoF) of a certain object have to solely be determined from video stream data captured from one or more cameras. As a first step, this comprises the recognition, localization, and identification of the object in a camera image. Once the presence of the object has been verified, its position and orientation has to be computed by utilizing known geometric features.

A common approach to support the aforementioned processing steps is to attach fiducial markers to the object. The markers used in our system (see Fig. 5) have black borders on a white background for maximum contrast, to facilitate their detection and localization in a camera image. Furthermore, each marker holds a unique binary code pattern that allows identification of the marker and the associated object.

Given a calibrated camera, the 6DoF of a size-known marker can be estimated by using its corner points, which may be extracted by image processing methods. In photogrammetry, this process is referred to as "pose estimation" and describes the matching of a set of points in 3D space to their projection in 2D image space. The correspondence between those points can be expressed by a set of equations, whose unknowns are uniquely related to the 6DoF of the corner point set, with respect to the camera coordinate system. Though the four corner points from a single marker are sufficient for pose estimation, a considerably higher accuracy is obtained by placing marker clusters on the tracked object. This redundancy allows compensation of measurement errors that are caused by image noise. The same effect can be achieved by combining multiple, synchronized cameras that are directed at the object. Besides this, multiple cameras may be used to increase both the field of view as well as the reliability of the system in occlusion situations.

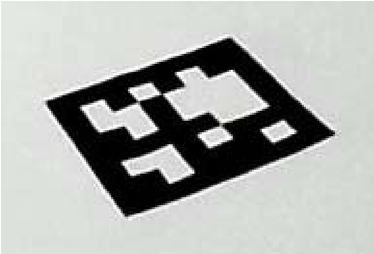

Fig. 5 Square fiducial marker

Much research on marker-based video tracking has been done in the field of augmented reality. A prominent outcome is ARToolkit [6], a software library dedicated to real-time marker tracking. We developed a pose estimation system for rigid bodies based on attached square markers [13]. For this purpose, a robust algorithm for the detection and exact localisation of square makers was developed, which is highly invariant against illumination variation. The accuracy of pose measurement was further improved by the incorporation of multiple camera fusion method, which was embedded in a flexible and dynamic system of pose tracking of multiple objects in complex and freely configurable scenarios.

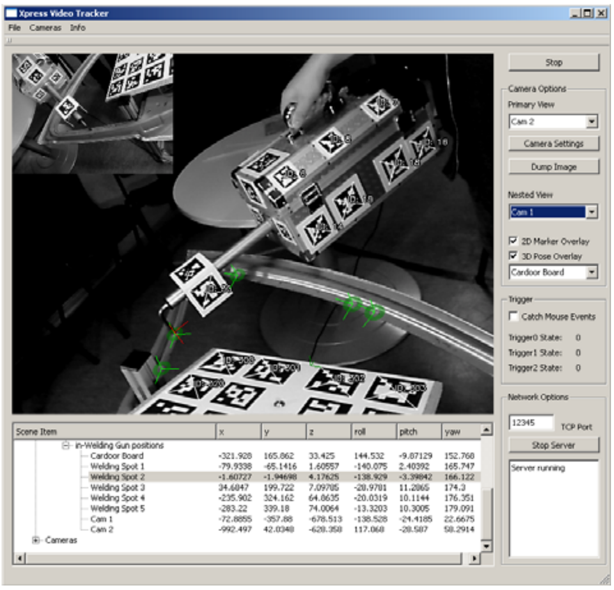

Fig. 6 Video tracking with a welding gun mock-up

For the functional sample, we use two consumer-grade USB-webcams at a resolution of 640x480 pixel. We equipped the welding gun with 22 markers at known positions in a defined welding gun coordinate system (Fig. 6 shows the tracking system with a welding gun mock up). Additional markers placed at the car door span the reference coordinate system for DoF data. Under appropriate illumination conditions, the video system provides measurements within a range of some millimeters of accuracy at a sampling rate of about 10 Hz.

Sensor Data: Inertial Sensors

For the purpose of providing high-frequency acceleration and angular rate signals, measured by a wearable device, a low cost inertial measurement unit (IMU), based on affordable MEMS technology, was utilized. The IMU measures accelerations and angular rates in 3D space from 3 acceleration and gyro sensor modules, each of them perpendicular to each other. From these raw data measurements, high frequent motion signals can be derived without the need of a line of sight in contrast to video sensors. When using inertial sensors it has to be considered for signal processing that sampled sensor signals of the IMU are prone to erroneous influences, of which signal noise and time-varying biases are most important for this particular application. The drift of bias values in the sensor signals makes it necessary that the IMU has to be calibrated before usage. In case of angular rate signals proper calibration can be achieved by measuring bias values when the IMU is in a stable position. However, for measuring offset values of acceleration sensors their attitude must be known, because of the influence of gravity.

In our functional sample, the welding gun has been equipped with the SHAKE SK6 sensor device of the company samh Engineering Services. This device provides 6DoF acceleration and angular rate measurements with a sampling rate of up to 256 Hz, such as further movement-related sensor data (e.g. magnetometer), which may allow reusability in other scenarios.

Sensor Data: Trigger Signals

Trigger signals have been defined for the indication of (simple) atomic actions of the following type of discrete events: “welding gun trigger pulled”, “remove gun from storage” and “return gun to storage”. The acquisition of these signals has been realized straightforward by using binary switches, one attached to the welding gun representing the welding gun trigger and another one indicating the storage position of the welding gun.

Activity Recognition: Spatial Deviation Detection

Basing on position information of the welding gun tip, the spatial deviation of the welding gun to the position of a welding spot can easily be derived by evaluating the euclidean norm d_{spot} of the difference between welding gun tip position $\vec{p}$ and spot position $\vec{p}_{spot}$, according to

$$d_{spot} = \left\| \vec{p} - \vec{p}_{spot} \right\|. \qquad (1)$$

By this it can be checked if the spatial deviation of the welding gun is below the threshold value of the desired tolerance (e.g. 15mm) for all welding spot positions.

The thresholded signal represents a quantization of the continuous position values of *IR³* into discrete position spaces, representing atomic actions of the type of “aligned to spot n”.

Activity Recognition: Motion Detection

Another type of atomic action to be recognized is the motion indicator “welding gun is moving”. Motion of the welding gun can simply be recognized by evaluating, if angular rate and acceleration measurements are within predefined thresholds. Thus, a binary signal (motion/ no motion) with high frequent sampling rates is created, which is used as additional information to video-tracked position values.

In order to avoid the effort that the attitude of the IMU has to be known for calibrating acceleration offsets, we use the length a_{abs} of the vector of the perpendicular accelerations a_x, a_y, a_z for motion detection:

$$a_{abs} = \sqrt{(a_x)^2 + (a_y)^2 + (a_z)^2} \tag{2}$$

Given a_{abs} and the values of the angular rate measurements motion is assumed when any of the four named values is above a defined threshold value.

Behavior Interpretation: Statechart

A statechart is a part of the UML for graphically describing finite state machines with a complex structure. Statecharts have been introduced by David Harel [5] with the intention to extend the formalisms of common state diagrams (e.g. by features of hierarchical structures, concurrencies and communication between states) in order to handle visual descriptions of complex systems.

In our functional sample we use a finite state machine represented by the statechart that we presented in Fig. 4. According to this statechart, the worker task is modeled as a sequence of task steps, related to each other by transitions in the form of atomic actions and optional guard conditions. The transitions result from atomic actions in the form of binary trigger signals, spatial deviation detection information and motion detection results. In most of the states outputs are emitted, comprising signals to enable or disable the welding gun trigger or to provide worker feedback information.

With this functional sample we designed a worker behavior interpretation solution that may be adapted to other tool handling task scenarios as well. Especially the video tracking and inertial sensor components are re-useable in particular, since they provide pose data and internal movement data, which are essential characteristics for a lot of manufacturing tasks.

V. EXPERIMENTS

For testing our functional sample approach, a test environment has been set up with a car door clamped in a fixture, which represents the working place for the welding task (Fig. 7). As mentioned, we use two cameras for our video tracking system, which increases the robustness and accuracy of the position estimation result compared to a single camera system. The sensor box with the IMU is mounted on the welding gun as shown in Fig. 7.

The positions of the welding spots are defined in a world coordinate system with a fixed reference to the car door. By using a calibration pattern, which is placed on a defined position on the car door, the video tracking system can be calibrated, which is necessary when the spatial relation between cameras and car door is changed. Calibration of the IMU hast to be done before every test run, because of the time-varying biases.

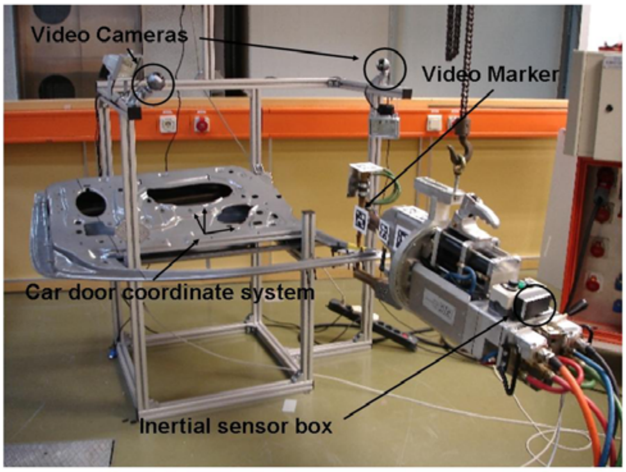

Fig. 7 Functional sample set up

In order to show the operational reliability of our system, tests have been done to show the accuracy of the video tracker, since it is one of the most important components, as well as the performance of the overall system. All tests have been done with a welding gun mock up, which is shown in Fig. 6, because of the easier handling for tests. However, our test results are transferable to the real welding gun, since the measurement principle is the same.

The statistical evaluation result of the video tracker accuracy tests can be seen in Table I. Disregarding single outliers (which rarely occur) we can say that this result is sufficient for the requirement of a maximum tolerance of 15mm. For these tests the welding gun tip has been moved along paths parallel to the axes of the reference coordinate system (so that for each path the reference values of only one coordinate varied). In the table the values of the stable axes have been evaluated. The tests contains 5 repetitions of test runs in which the gun tip has been moved with different alignment to the defined paths. From the table we can see that the deviation of the signal becomes most significant in direction of the y-axis (the reason for that is the geometric arrangement of the cameras). The maximal error created by an outlier has been 16mm and the typical deviation in direction of the y-axis has been 6mm.

TABLE I
POSITION ESTIMATION ACCURACY

Coordinate	Reference[1)]	Mean[1)]	Squared deviation[2)]	Maximal deviation[1)]
x	165,714	165,172	5,576	6,614
y	-57,143	-57,431	34,961	16,233
z	0	1,985	6,959	6,701

[1)] in [mm]
[2)] in [mm^2]

As for the detection of high-frequency motions with inertial sensors thresholding the raw sensor measurements turned out to be a appropriate method to detect motions. Nevertheless, it has to be kept in mind that translational movements at a constant speed can’t be detected by this method. However this means no limitation, because these movements are detected by the video tracker.

The functionality of the complete system has been evaluated in several tests. Fig. 8 and Fig. 9 show a typical test result of detected state sequences for correct worker behavior and the result of a test in which incorrect worker behavior had to be detected. For each test a reference was taken by a human observer (indicated by black values) representing the ground truth for the state values (indicated by red values). In both figures states from Fig. 4 are denoted after: 0 – “tool moving”; 1- “tool at spot n”; 2- “ready to weld spot n”; 3- “welding spot n” (or trigger pulled in case of a reference value); -1 – “misbehavior”. Additional information about the current spot number (i.e. the spot to which the states 1-3 are related) from the reference and from the protocol list is written under related states.

Fig. 8 shows the result of a test in which the four welding spots of our scenario have been welded correctly. As we can see, all states have been detected according to the reference excepting a small inaccuracy when the worker approached the first spot. Such inaccuracies are caused by outliers in the position measurements of the video tracker. It has to be noted that temporal deviations of one second as those which occur at state 3 are to be neglected, because they are an erroneous effect of visual evaluation.

Besides correct worker behavior, tests have been done to evaluate situations of incorrect worker behavior. A typical example is given in Fig. 9, in which misbehavior is related to three situations according to the scenario description, namely: welding at a wrong position, welding when the gun is in motion and (re-) welding of an already welded spot.

From the figure we can see that after spot number 2 had been welded correctly the welding trigger has been pulled without the welding gun being aligned to a welding spot position. After that the welding gun has been positioned in the area of a welding spot, but the welding gun trigger has been pulled when the tool was still in motion. Finally, the welding gun has been correctly aligned again to spot number 2 with the trigger being pulled when the gun wasn’t moving. From the figure we can see that all three situations in which the worker executed the task incorrectly have been correctly interpreted as misbehavior.

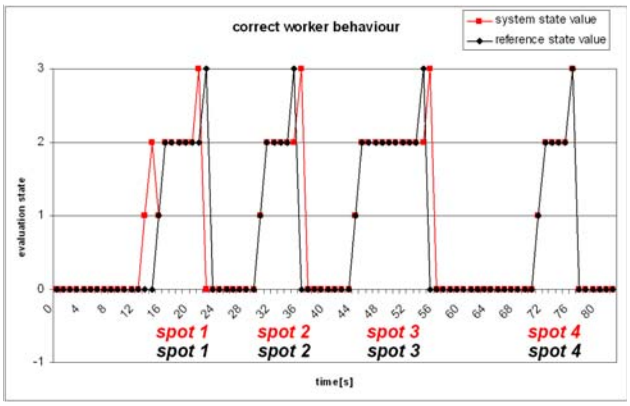

Fig. 8 Evaluation of correct worker behavior

Fig. 9 Evaluation of incorrect worker behavior

VI. CONCLUSION AND OUTLOOK

In our work we introduced a general approach for interpreting human worker behavior with regard to flexible production. The process of interpreting worker behavior according to a given task has been described as a framework of the four levels of process instances of Fig. 1. By approaching given problems in a top down manner (with adaptation after experimental validation) our target is to find flexible solutions with re-useable components that are easy to adapt to new scenarios of interpreting worker behavior.

By means of the welding gun scenario we demonstrated a worker behavior interpretation solution that may also be applied to other tool handling task scenarios. This can be achieved by a new definition of the task model on the highest level of abstraction and some minor adaptations to the other levels. Furthermore, the video tracking and inertial sensor components are re-useable in other scenarios as well, regarding the fact that pose and internal movement information are essential characteristics for a lot of manufacturing tasks. Thus, they are generally not restricted to any specific application

As future work some improvements can be done to the tool handling scenario. Similar to the work of [11] inertial sensor measurements and position data from the video tracking system may be fused in order to improve the quality of tracked positions or deal with video tracking outliers or short time occlusions.

Besides that, our ongoing work is dealing with a scenario of higher complexity, in which the worker task is to manually assemble parts (assembly task). This scenario is of particular interest, since there can be a high variance in the way in which a worker assembles parts. Moreover, the detection of atomic actions in this scenario is more difficult than in the welding gun scenario.

References

[1] J. K. Aggarwal, and Q. Cai, “Human Motion Analysis: A Review,” Comput. Vis. Image Underst., 73, no. 3, 1991, 428–440.

[2] D. Ayers, and M. Shah, “Monitoring human behavior from video taken in an office environment,” Image and Vision Comput., 19, no. 12, 2001, 833–846.

[3] C. Colombo, D. Comanducci, and A. Del Bimbo, “Behavior monitoring through automatic analysis of video sequences,” Proc. ACM International Conference on Image and Video Retrieval CIVR , 2007, 288-293.

[4] D. M. Gavrila, “The visual analysis of human movement: a survey,” Comput. Vis. Image Underst., 73, no. 1, 1999, 82-98.

[5] D. Harel, “Statecharts: A visual formalism for complex systems,” Sci. Comput. Program. 8, no. 3, 1987, 231-274.

[6] H. Kato, and M. Billinghurst, “Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System,” Proc. of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR), 1999.

[7] R. C. Luo, C.-C. Yih, and K. L. Su, “Multisensor fusion and integration: approaches, applications, and future research directions,” Sensors Journal, IEEE 2, no. 2, 2002, 107–119.

[8] T. B. Moeslund, A. Hilton, and V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” Comput. Vis. Image Underst., 104, no. 2, 2006, 90-126.

[9] N. T. Nguyen, H. H. Bui, S. Venkatsh, G. West, “Recognizing and monitoring high-level behaviors in complex spatial environments,” Proc. of the IEEE Comp. Society Conference on Computer Vision and Pattern Recognition, 2003, II- 620-5 vol.2.

[10] L. Ojeda, and J. Borenstein, “Non-GPS Navigation for Emergency Responders,” Int. Joint Topical Meeting: Sharing Solutions for Emergencies and Hazardous Environments, 2006, 12-15.

[11] N. Parnian, and M. F. Golnaraghi, “Integration of vision and inertial sensors for industrial tools tracking,” Sensor Review 27, 2007, 132-141.

[12] N. Robertson, and I. Reid, “Behaviour understanding in video: a combined method,” Computer Vision, ICCV 2005. Tenth IEEE International Conference on 1, 2005, 808-815.

[13] C. Schauer, “Verfolgung von Freiheitsgraden eines Werkzeugs aus Videodaten,“ Diploma Thesis, University of Applied Sciences , 2007

[14] D. Shin, R. A. Wysk, and L. Rothrock, “An investigation of a human material handler on part flow in automated manufacturing systems,” Systems, Man and Cybernetics, Part A, IEEE Transactions on 36, no. 1, 2006, 123–135.

[15] T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, and G. Troster, “Combining Motion Sensors and Ultrasonic Hands Tracking for Continuous Activity Recognition in a Maintenance Scenario,” Wearable Computers, 2006 10th IEEE International Symposium on, 2006, 97–104.

[16] T. Stiefmeier, D. Roggen, G. Troster, G. Ogris, and P. Lukowicz, “Wearable Activity Tracking in Car Manufacturing,” Pervasive Computing, IEEE 7, no. 2, 2008, 42–50.

[17] D. Vlasic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popović, “Practical motion capture in everyday surroundings,” ACM Trans. Graph., 26, no. 3, 2007, 35:1-35:9.

[18] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, “Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers,” Pattern Analysis and Machine Intelligence, IEEE Transactions on 28, no. 10, 2006, 1553–1567.

[19] L. Bao and S. S. Intille, “Activity Recognition from User-Annotated Acceleration Data,” Proceceedings of the 2nd International Conference on Pervasive Computing, 2004 1–17.

[20] M. H. Ko, G . West, S. Venkatesh, M. Kumar, “Using dynamic time warping for online temporal fusion in multisensor systems,” Information Fusion, 9, no. 3, 2008, 370-388.

[21] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, G. Troster, “Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness,” Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on, 281–286.