Making Sense: Experiences with Multi-Sensor Fusion in Industrial Assistance Systems

Benedikt Gollan[1], Michael Haslgruebler[2], Alois Ferscha[2] and Josef Heftberger[3]
[1]*Pervasive Computing Applications, Research Studios Austria FG mbH, Thurngasse 8/20, 1090 Vienna, Austria*
[2]*Institute of Pervasive Computing, Johannes Kepler University, Altenberger Strasse 69, Linz, Austria*
[3]*Fischer Sports GmbH, Fischerstrasse 8, Ried am Innkreis, Austria*

Keywords: Sensor Evaluation, Industrial Application, Multi-Sensor Fusion, Stational Sensors, Wearable Sensors, Challenges.

Abstract: This workshop paper discusses the application of various sensors in an industrial assembly scenario, in which multiple sensors are deployed to enable the detailed monitoring of worker activity, task progress and also cognitive and mental states. The described and evaluated sensors include stationary (RGBD cameras, stereo vision depth sensors) and wearable devices (IMUs, GSR, ECG, mobile eye tracker). Furthermore, this paper discusses the associated challenges mainly related to multi-sensor fusion, real-time data processing and semantic interpretation of data.

1 INTRODUCTION

The increasing digitalization in industrial production processes goes hand in hand with the increased application of all kinds of sensors, whereby the majority of these sensors are exploited for automated machine-to-machine communication only. However, in all human-in-the-loop processes which involve manual or semi-manual labor, physiological sensors are on the rise, assessing the behavioral and somatic states of the human workers as to deduce on activity or task analysis as well as the estimation of human cognitive states.

The observable revival of human labor as an opposing trend to the predominant tendency of full automation (Behrmann and Rauwald, 2016) is associated with the requirements of industrial processes to become more and more adaptive to dynamically changing product requirements. The combination of the strengths of both men and machine working together yields the best possible outcome for industrial production, as humans provide their creativity, adaptability, and machines ensuring process constraints such as quality or security.

In the light of these changes towards men-machine collaboration, it is essential for machines or computers to have a fundamental understanding of their users - their ongoing activities, intentions, and attention distributions. The creation of such a high level of awareness requires not only (i) the selection of suitable sensors but as well needs to solve fundamental problems regarding (ii) handling the big amounts of data, (iii) the correct fusion of different sensor types as well as (iv) the adequate interpretation of complex psycho-physiological states.

This work will introduce the industrial application scenario of an aware assistance system for a semi-manual assembly task, introduce and evaluate the employed sensors and discuss the derived challenges from the associated multi-sensor fusion task.

1.1 Related Work

With the ever-increasing number of sensors, the fusion of the data from multiple, potentially heterogeneous sources is becoming a non-trivial task that directly impacts application performance. When addressing physiological data, such sensor collections are often referred to as Body Sensor Networks (BSNs) with applications in many domains (Gravina et al., 2017). Such physiological sensor networks usually cover wearable accelerometers, gyroscopes, pressure sensors for body movements and applied forces, skin/chest electrodes (for electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR), and electrical impedance plethysmography (EIP)), (PPG)

Gollan, B., Haslgruebler, M., Ferscha, A. and Heftberger, J.
Making Sense: Experiences with Multi-Sensor Fusion in Industrial Assistance Systems.
DOI: 10.5220/0007227600640074
In *Proceedings of the 5th International Conference on Physiological Computing Systems (PhyCS 2018)*, pages 64-74
ISBN: 978-989-758-329-2

sensors, microphones (for voice, ambient, and heart sounds) and scalp-placed electrodes for electroencephalogram (EEG) (Gravina et al., 2017). These wearable sensor types can also be enriched with infrastructural, remote sensor systems such as traditional (RGB) and depth cameras.

Sensor networks are investigated in and employed by industrial applications (Li et al., 2017), specifically in domains such as the Automotive Industry (Marabelli et al., 2017), (Otto et al., 2016), healthcare IOT (Baloch et al., 2018), (Chen et al., 2017) or food industry (Kröger et al., 2016), in industrial use cases as welding (Gao et al., 2016) or CNC-machining (Jovic et al., 2017).

1.2 Contribution of this Work

This work introduces an industrial assistance system which is based on the integration of various sensors which have been applied and evaluated regarding their applicability and suitability in an industrial application. In this context, this work presents an overview of the investigated sensors with reviews and experiences regarding data quality, reliability, etc. Furthermore, this work reports on the key challenges and opportunities which are (i) handling of big amounts of data in real-time, (ii) ensuring interoperability between different systems, (iii) handling uncertainty of sensor data, and the general issues of (iv) multi-sensor fusion.

While Section 2 describes the industrial application scenario, in Sections 3 and 4 the respective sensors are introduced. Section 5 puts the focus on the discussions of challenges and opportunities and section 6 provides a summary and addresses future work.

2 INDUSTRIAL APPLICATION SCENARIO

The industrial application scenario is an industrial assistance system which is employed in a semi-manual industrial application of a complex assembly of premium alpine sports products, where it is supposed to ensure the high-quality requirements by providing adaptive worker support.

The work task consists of manually manipulating and arranging the multiple parts whereas errors can occur regarding workflow order, object orientation, or omission of parts. These errors express in unacceptable product quality differences regarding usage characteristics (e.g. stability, stiffness), thus increase rejects and inefficiency.

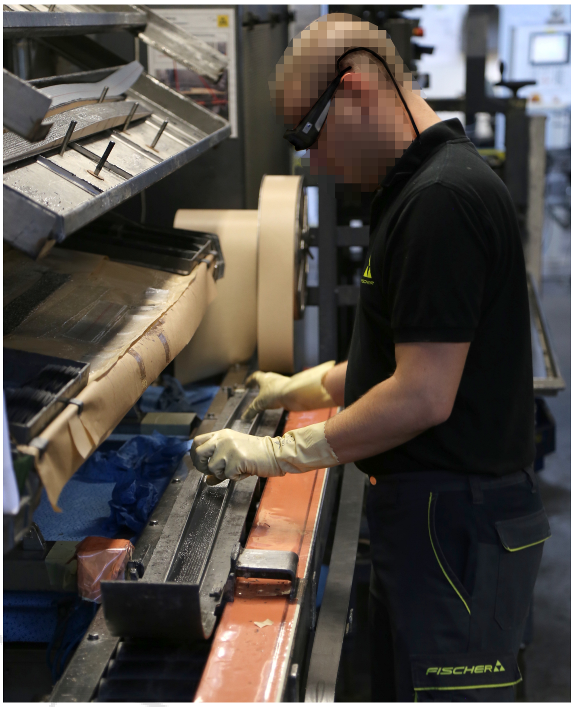

Figure 1: Ski assembly procedure and environment.

Full automation of the process is not feasible due to (i) required high flexibility (minimal lot sizes, changing production schedules), (ii) used material characteristics (highly sticky materials) and (iii) human-in-the-loop production principles enable the optimization of product quality and production processes.

In this context, the sensor-based assistance system is designed to enable the realization of an adaptive, sensitive assistance system as to provide guidance only if needed, thus minimizing obtrusiveness and enabling the assistance system to seamlessly disappear into the background. Furthermore, the adaptivity of the feedback design enables the education of novices in training-on-the-job scenarios, integrating novices directly into the production process during their one month training period without occupying productive specialists.

The assistance system is supposed to observe the task execution, identify the associated step in the workflow and identify errors or uncertainty (hesitation, deviation from work plan, etc.) to support the operator via different levels of assistance (Haslgrübler et al., 2017). The selection of assistance depends on operator skill (i.e. day 1 trainee vs 30-year-in-the-company worker), cognitive load and perception capability to provide the best possible assistance with the least necessary disruption. Such supportive measures range from, laser-based markers for part placement or visual highlighting of upcoming work steps in case of uncertainty, to video snippets visualizing the correct execution of a task, in case of doubt.

3 ACTIVITY SENSING

The most common application of activity and behavior analysis in industrial applications is monitoring of task progress for documentation or assistance applications. The main kinds of sensors and technologies that can be exploited for activity tracking are (i) stationary (visual) sensors and (ii) wearable motion sensors. The different fields of application are introduced in the following, for an overview please refer to Table 1:

3.1 Skeleton Tracking

Mainly stationary visual sensors are employed to identify body joints and the resulting associated skeleton pose. Depending on the application, these sensors address the full skeleton or sub-selections of body joints.

3.1.1 Full Skeleton Tracking

Sensor Description - Kinect v2. The Microsoft Kinect v2 combines an infrared and an RGB camera to track up to six complete skeletons, each consisting of 26 joints. The Kinect uses an infrared time-of-flight technology to build a 3D map of the environment and the objects in view. Skeleton data is provided by the associated Microsoft SDK which is restricted to Microsoft Windows platforms.

In the described application scenario, two Kinect cameras have been installed on opposing sides of the work environment - as a frontal positioning was not possible - to avoid obstructions and enable an encompassing perception of the scene. Based on a manual calibration of the two sensors, the data is combined into a single skeleton representation via a multi-sensor fusion approach as described in Section 5.4. The calibration is achieved via a two-step process: (1) real-world measurement of the placement and orientation angle of the sensors in the application scenario, obtaining the viewpoints of the two sensors in a joint coordinate system and (2) fine adjustment based on skeleton joints that are observed at the same time, at different positions. For this purpose, the head joint was chosen as it represents the most stable joint of the Kinect tracking approach - according to our experience. The overall result of the calibration approach is the localization and orientation of the two sensors in a joint coordinate system, thus enabling the overlay and fusion of the respective sensor input data.

Evaluation. Kinect-like sensors provide unique opportunities of skeleton tracking, thus overcome many problems associated with professional motion tracking systems such as enabling (i) markerless tracking, (ii) fast and simple setup and (iii) low-cost tracking results. However, due to the infrared technology, the depth sensors do not perform well in outdoor settings with high infrared background noise. Furthermore, the cameras require good allocation of the scene, with a full view of the worker for best tracking results.

Overall, the application of Kinect sensors in industrial applications requires careful handling and substantial data post-processing. With the Kinect skeleton data showing large amounts of fluctuations, the Kinect represents a cheap, yet not per se reliable sensor for skeleton tracking.

3.1.2 Sub-Skeleton Tracking

Sensor Description - Leap Motion. Aiming only at tracking the hands of a user, specifically in Virtual Reality applications, the Leap Motion controller represents an infrared, stereo-vision-based gesture and position tracking system with sub-millimeter accuracy (Weichert et al., 2013). Suitable both for mobile and stationary application, it has been specifically developed to track hands and fingers in a close distance of up to 0.8 m, enabling highly accurate hand gesture control of interactive computer systems.

In the introduced industrial application scenario, the Leap Motion controllers are installed in the focus areas of the assembly tasks, trying to monitor the detailed hand movements.

Evaluation. The Leap Motion controller shows high accuracy and also high reliability. Yet, unfortunately, the sensor shows a high latency in the initial registration of hands (up to 1-2 s). In a highly dynamic application as in the presented use-case scenario, this latency prevented the applicability of the Leap Motion sensor, as the hands were often already leaving the area of interaction when they were detected. For this purpose, this highly accurate sensor could not be applied in the final assistance setup, yet they represent a very interesting sensor choice when addressing a very stationary industrial task.

3.1.3 Joint Tracking

Mobile, wearable sensors are used to extract the movement of single body joints, most commonly the

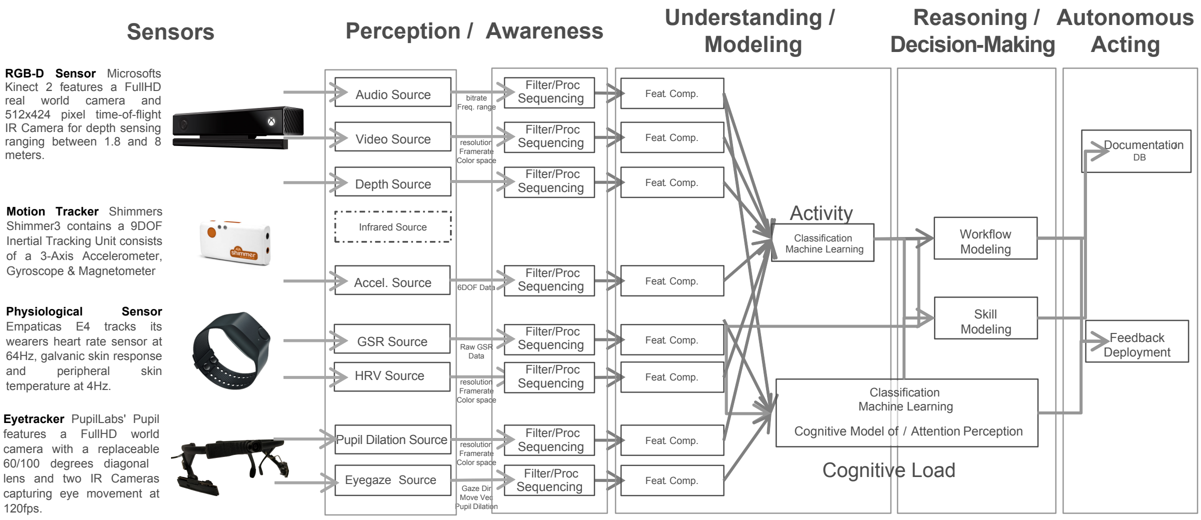

Figure 2: Scheme of the introduced industrial multi-sensor assistance system with the various level of abstractions: Perception, Understanding, Reasoning, Acting. Data from Sensors are processed individually and in aggregated form to perform activity, work-step, skill and cognitive recognition. Reasoning Models are then used to select appropriate assistance measure via different actors.

wrists for inference on hand movement activity. The vast majority of these sensors are based on accelerometers and gyrometers to provide relative changes in motion and orientation for behavior analysis.

Sensor Description - Shimmer. The Shimmer sensors have already been validated for use in academic and industrial research applications (Burns et al., 2010), (Gradl et al., 2012), (Srbinovska et al., 2015) Also, Shimmer research offers the several tools and APIs for manipulation, integration and easy data access. Due to their small size and lightweight (28g) wearable design, they can be worn on any body segment for the full range of motion during all types of tasks, without affecting the movement, techniques, or motion patterns. Built-in inertial measurement sensors are able to capture kinematic properties, such as movement in terms of (i) Acceleration, (ii) Rotation, (iii) Magnetic field.

The updated module boasts a 24MHz CPU with a precision clock subsystem and provides the three-axis acceleration and gyrometer data. We applied a shimmer sensor on each of the worker's hands to obtain expressive manual activity data. The Shimmer sensors provide their data with a frame rate of 50 Hz. In the current scope of the implementation, hand activity data is parsed from respective text/csv-files in which the recorded data has been stored. This accumulates to 6 features per iteration per sensor (3x gyrometer, 3x accelerometer) every 20 ms.

Evaluation. Shimmer sensors provide reliable and accurate tracking data, also in rough industrial environments. The real-time analysis requires a smartphone as a transmission device, yet does work reliably. Overall, when aiming for raw accelerometer data, the Shimmer sensor platforms have proven their suitability.

3.2 Gesture Detection

The introduced Kinemic sensor is closely related to the previously described accelerometer sensors placed on the wrist of the worker. Yet, it does not provide access to the raw accelerometer data but directly provides only higher level gesture detections as result to the system. Due to this reason, the distinction between general joint tracking and hand gesture detection was made.

Sensor Description - Kinemic. The Kinemic wristband sensor for hand gesture detection is a new sensor for which almost no official information is available. It is based on - presumably - 3-axis accelerometer and gyrometer sensor and connected to a mobile computation platform (RaspberryPi) which carries out the gesture detection processes. Currently, 12 gestures are supported, with the goal to expand to customizable gestures, air writing, etc.

Evaluation. The sensors are easily initiated and integrated into a multi-sensor system. The recognition of the gestures works well for the majority of existing gestures. In summary, this sensor with the associated SDK provides a useful solution for people looking for high-level off-the-shelf gesture interaction, without requiring access to raw accelerometer data.

3.3 Behavior Analysis

3.3.1 Gaze-based Task Segmentation

The analysis of gaze behavior also provides interesting insights into the execution of activities, especially the segmentation of subsequent tasks in a work process. Recent work shows that the gaze feature *Nearest Neighbour Index* (Camilli et al., 2008), which describes the spatial distribution of fixations in a dynamic environment (Amrouche et al., 2018). Employing a wearable Pupil Labs eye tracker, this gaze behavior feature was capable of successfully segmenting and recognizing tasks. For the sensor discussion, please refer to section 4.1.1.

4 SENSING OF COGNITIVE STATES

4.1 Visual Attention

Generally, the human eye gaze represents the most efficient and fastest, consciously controlled form of information acquisition with the unique capability to bridge large distances. Intuitively, the human eye is mainly responsible for the positioning of eye gaze, thus represent an expression for stimulus selection, yet, fine details of gaze behavior also show connections to conscious and subconscious information processing mechanisms that allow inferences on internal attention processes.

4.1.1 Gaze Behavior

Sensor Description - Pupil Labs. the PupilLabs mobile eye tracker is realized as a modular and open source solution, providing direct access to all sensors and data streams (gaze position, gaze orientation, saccade analysis, pupil dilation, etc.), rendering the device as more suitable for academic research applications. The PupilLabs eye tracker enables direct access in real-time to all parameters and tracking results. The PupilLabs device provides the eye tracking data for each eye with a distinct timestamp, requiring additional synchronization of obtained data frames.

Evaluation. The PupilLabs eye tracker provides rather simple and encompassing access to basic data streams. As a consequence, the PupilLabs eye tracker is a suitable, low-cost device for ambitious developers that want to develop algorithms based on the raw sensor data. However, the sensor fails in outdoor environments when exposed to scattered infrared light. In the proposed application scenario, the PupilLabs eye tracker is employed for associating gaze orientation to objects in space (hands, task-relevant objects, etc.) via object recognition in the first person video. However, the achieved results are always situated in the user-specific coordinates, which, to be associated with an overall world space of the industrial shop floor requires a complex and detailed localization of the worker, regarding both head location and orientation.

4.1.2 Visual Focus of Attention

The general spatial allocation of attention can also be assessed on a less-fine-grained level via external, infrastructural sensors. The so-called visual focus of attention has found sustained application in human-computer-interaction applications. These differ in application and tracking technology but all use head orientation as the key information for attention orienting (Asteriadis et al., 2009), (Smith et al., 2006), (Leykin and Hammoud, 2008).

Sensor Description - Kinect v2. As described above, the Kinect provides a quite reliable skeleton information on a low-cost platform. It also provides joint orientation, yet not head orientation. To exploit the available data for the estimation of the visual focus of attention, an approximation of shoulder axis and neck-head axis can be employed.

Evaluation. The visual focus of attention data derived from this approach can only provide very rough information on the actually perceived objects and areas in space. However, it directly provides the spatial context, which misses in the assessment via wearable eye trackers, as described above. Hence, the combination of the two sensors, wearable and infrastructural, may help in providing substantial advances in the task of 3D-mapping of visual attention in industrial environments - a task which will be pursued in future work.

4.2 Arousal

In the literature, arousal is defined by Kahneman (Kahneman, 1973) as *general activation of mind*, or as *general operation of consciousness* by Thatcher and John (Thatcher and John, 1977).

Psychophysiological measures exploit these physical reactions of the human body in the preparation of, execution of, or as a reaction to cognitive activities. In contrast to self-reported or performance measures, psychophysiological indicators provide continuous data, thus allowing a better understanding of user-stimulus interactions as well as non-invasive and non-interruptive analysis, maybe even outside of the scope of the users consciousness. Whereas these measures are objective representations of ongoing cognitive processes, they often are highly contaminated by reactions to other triggers, as e.g. physical workload or emotions.

4.2.1 Cognitive Load

Besides light incidence control, the pupil is also sensitive to psychological and cognitive activities and mechanisms, as the musculus dilatator pupillae is directly connected to the limbic system via sympathetic control (Gabay et al., 2011), hence, the human eye also represents a promising indicator of cognitive state. Currently, existing analysis approaches towards analysis of cognitive load from pupil dilation - Task-Evoked Pupil Response (TEPR) (Gollan and Ferscha, 2016) and Index of Cognitive Activity (ICA) (Kramer, 1991) - both find application mainly in laboratory environments due to their sensitivity to changes in environment illumination.

Sensor Description - PupilLabs. The employed Pupil Labs mobile eye tracker provides pupil diameter as raw measurement data, both in relative (pixel size) as in absolute (mm) units due to the freely positionable IR eye cameras. The transformation is achieved via a 3D model of the eyeball and thus an adaptive scaling of the pixel values to absolute mm measurements.

Evaluation. The assessment of pupil dilation works as reliably as the gaze localization with the lack of official accuracy measures in comparative studies. Hence, it is difficult to evaluate the sensor regarding data quality. Overall, the assessment of pupil dilation with the mobile Pupil Labs eye tracker provides reliable data, for laboratory studies or field application. Erroneous data like blinks needs to be filtered in post-processing of the raw data.

4.2.2 Cardiac Indicators

The cardiac function, i.e. heart rate, represents another fundamental somatic indicator of arousal and thus of attentional activation as a direct physiological reaction to phasic changes in the autonomic nervous system (Graham, 1992). Heart Rate Variability (HRV), heart rate response (HRR) or T-Wave amplitude analysis are the most expressive physiologic indicators of arousal (Suriya-Prakash et al., 2015), (Lacey, 1967).

The stationary and mobile assessment of cardiac data is very well established in medical as well as customer products via diverse realizations of ECG sensors. The different sensors are based on two main independent measurement approaches: (i) measuring the electric activity of the heart over time via electrodes that are placed directly on the skin and which detect minimal electrical changes from the heart muscle's electro-physiologic pattern of depolarizing during each heartbeat; and (ii) measuring the blood volume peak of each heartbeat via optical sensors (pulse oximeters) which illuminates the skin and measures the changes in light absorption to capture volumetric changes of the blood vessels (Photoplethysmography (PPG)).

Sensor Description - Shimmer. Shimmer sensors use a photoplethysmogram (PPG) which detects the change in volume by illuminating the skin with the light from a light-emitting diode (LED) and then measuring the amount of light transmitted or reflected towards a photodiode. From this volume changes an estimate of heart rate can be obtained.

Sensor Description - Empatica E4. The E4 wristband allows two modes of data collection: (i) in-memory recording and (ii) live streaming of data. Accessing in-memory recorded data requires a USB connection to a Mac or Windows PC running Empatica Manager Software for a posteriori analysis. Accessing streaming data for real-time analysis of somatic data, the Empatica Real-time App can be installed from the Apple App Store or Google Play Market onto a smartphone device via Bluetooth on which the data can be processed or forwarded. Additionally, a custom application can be implemented for Android and iOS systems.

Sensor Description - Microsoft Band 2. The Microsoft Band 2 is equipped with an optical PPG sensor for analysis of pulse. With the Microsoft Band representing an end-user product, the focus in the provided functionality is not set on providing most accessible interfaces for academic purposes, yet, still, the available SDK enables the access of raw sensor data in real-time. For data access, the sensor needs to be paired with a smartphone device and data can be transferred via a Bluetooth connection for either direct processing on the mobile device or further transmission to a general processing unit.

Evaluation. The Microsoft Band is highly restricted in sensor placement as the sensor is integrated into the wristband of the device and thus measures the skin response on the bottom surface of the wrist. In experiments, the Microsoft Band sensor showed large drops in measurement data, most probably due to a change of contact between the sensor and the skin during device shifts. In contrast, the Shimmer Sensing Platform allows much more freedom in the placement of the sensor with the help of external sensing modules e.g. pre-shaped for mounting on fingers which show the most promising locations for reliable GSR measurements.

Accessing real-time data for the E4 wristband shows similar comfort levels as the Microsoft Band as the device needs to be paired with a smartphone device and data can be transferred via a Bluetooth connection for either direct processing on the mobile device or further transmission to a general processing unit. Being designed for research and academic purposes, the Shimmer platform provides easiest and fastest access via open and intuitive interfaces. Overall, the data from all devices can be accessed in real-time, yet the destinated applications of the products resemble in their applicability in research and development approaches.

4.2.3 Galvanic Skin Response

From the very early 1900s, the Galvanic Skin Response has been the focus of academic research. The skin is the only organ that is purely innervated by the sympathetic nervous system (and not affected by parasympathetic activation). The GSR analyzes the electrodermal activity (EDA) of the human skin which represents an automatic reflection of synaptic arousal as increased skin conductance shows significant correlations with neuronal activities (Frith and Allen, 1983), (Critchley et al., 2000). Hence, Galvanic Skin Response (GSR) acts as an indicator of arousal and increases monotonically with attention in task execution (Kahneman, 1973).

Sensorial Assessment. The accessibility of the raw and real-time data depends on the respective development environment which is provided to support these sensors, ranging from a general limitation to statistical information to access of true real-time data.

The GSR can be assessed via mobile, wearable sensors worn on the bare skin, e.g., as integrated into activity trackers or smartwatches or scientific activity and acceleration sensors. These sensors measure the skin conductance, i.e. skin resistivity via small integrated electrodes. The skin conductance response is measured from the eccrine glands, which cover most of the body and are especially dense in the palms and soles of the feet. In the following, three wearable sensors are explored which provide the analysis of skin conductance response:

Evaluation

E4 Wristband: is a hand wearable wireless devices designated for continuous, real-time data acquisition of daily life activities. It is specifically designed in an extremely lightweight (25g) watch-like form factor that allows hassle-free unobtrusive monitoring in- or outside the lab. With the built-in 3-axis accelerometer sensor the device is able to capture motion-based activities. Additionally, the device is able to capture the following physiological features (i) Galvanic skin response (ii) Photoplethysmography (heart rate) (iii) Infrared thermophile (peripheral skin temperature). The employed Empatica E4 Wristband has already found application in various academic research applications and publications (van Dooren et al., 2012), (Fedor and Picard, 2014).

Microsoft Band 2: offers an affordable mean for tracking a variety of parameters of daily living. Besides 11 advanced sensors for capturing movement kinematics, physical parameters and environmental factors the device also offers various channels for providing feedback. A 1.26 x 0.5-inch curved screen with a resolution of 320 x 128 pixels can be used to display visual messages. Additionally, a haptic vibration motor is capable of generating private vibration notifications.

Shimmer: sensors have already been validated for use in biomedical-oriented research applications. Due to their small size and lightweight (28g) wearable design, they can be worn on any body segment for the full range of motion during all types of tasks, without affecting the movement, techniques, for motion patterns. Built-in inertial measurement sensors are able to capture kinematic properties, such as movement in terms of (i) Acceleration, (ii) Rotation, (iii) Magnetic field.

Table 1: Overview on introduced sensors, grouped according to their sensing category and analysis type, listing the associated technologies and sensor parameters.

	Category	Type	Sensor Name	Technology	Accuracy / Range
Activity	Skeleton	Full Skeleton	Microsoft Kinect v2	Time-of-Flight Infrared	Depth: 512x424 @ 30 Hz FOV: 70° x 60° RGB: 1920x1080 @ 30 Hz FOV: 84° x 53° acc: 0.027m, (SD: 0.018m) depth range: 4m
		Sub-Skeleton	Leap Motion (Potter et al., 2013)	Stereo-Vision Infrared *hand tracking*	FOV: 150° x 120° avg error: $< 0.0012m$ (Weichert et al., 2013) depth range: 0.8m
		Joint Tracking (wrist)	Shimmer	3-axis accelerometer gyrometer	Range: $\pm 16g$ Sensitivity: 1000 LSB/g at $\pm 2g$ Resolution: 16 bit
	Gesture	Hand Gesture	Kinemic	3-axis accelerometer gyrometer	not available
	Behavior Analysis	Gaze Behavior	Pupil Labs (Kassner et al., 2014)	Mobile Eyetracker *gaze feature analysis for task segmentation*	accuracy 91% (Amrouche et al., 2018)
Cognitive States	Visual Attention	Gaze Behavior	Pupil Labs (Kassner et al., 2014)	Mobile Eyetracker *fixations, saccades, gaze features*	Gaze acc: 0.6° Sampling Rate: 120 Hz Scene Camera: 30 Hz @ 1080p 60 Hz @ 720p 120 Hz @ VGA Calibration: 5-point, 9 point
		Visual Focus of Attention	Microsoft Kinect v2	Head Orientation from Skeleton Tracking	not available
	Arousal	Cognitive Load	Pupil Labs	Mobile Eyetracker *pupil dilation*	pupil size in pixel or mm via 3D model acc. not available
		Heart Rate (HRV, HRR)	Microsoft Band 2	PPG	avg. error rate: 5.6% (Shcherbina et al., 2017)
			Empatica E4 Wristband (Poh et al., 2012)	PPG	samping frequency: 64 Hz error rate: 2.14%
		Galvanic Skin Response	Microsoft Band 2		data rate: $0.2/5Hz$ acc. not available
			Empatica Wristband	Empatica E3 EDA proprietary design	data rate: 4 Hz mean cor. to reference 0.93, $p < 0.0001$ (Empatica, 2016)

5 CHALLENGES & OPPORTUNITIES

5.1 Summary

In the previous chapters, several sensors have been described regarding their underlying technology, access to sensor data and evaluation regarding suitability for academic or industrial exploitation. As an overview, a short fact summary of the information is collected in Table 1, including further numerical data regarding the accuracy and range of the sensors, if available.

5.2 Handling Amounts of Data

The first challenge in the analysis of multi-sensor applications is the handling of the amounts of data, usually with real-time requirements. This applies both the required levels of computational performance as well as to further hardware assets as BUS bandwidth or hard drive access speed.

But also the offline handling of the data may represent problems for the design of interactive systems as especially raw video data - when stored - quickly exceeds GigaBytes of data. These amounts of data need to be managed, if possible in suitable database structures to enable efficient further processing of recorded data.

Other than data transfer and storage, also human resources for post-processing of the data represents a substantial challenge. This implies checking, filtering data, extracting relevant segments of data, etc. Especially - when aiming for supervised machine learning tasks - the manual labeling of activities represents an effort which often substantially exceeds the actual time of collected data and needs to be considered in the application setup. This labeling can be improved via suitable software solutions that enable the review and direct labeling of multimodal data streams.

5.3 Interoperability, Interfaces, Operating Platforms

Besides the pure amount of data, the different sources and interfaces represent a further source of problems. Depending on the producer, the analysis of the sensors requires specific supported frameworks and development environments. Mobile sensors are often associated with Android apps for mobile data collection and transfer, or e.g. the Microsoft Kinect sensors require Microsoft Windows platforms for operation, etc.

Creating a multi-sensor industrial application requires the multi-platform capability of development staff and often the creation of distributed systems operating on different native platforms. In the presented industrial application, such a distributed set of platforms is employed, inter-connected with a cross-platform messaging solution, thus overcoming the interoperability issue.

5.4 Multi-sensor Fusion

In many industrial applications, no single sensor is suitable to cover the overall complexity of a situation. Furthermore, no sensor provides perfect data, so redundant sensor designs enable the compensation of sensor failure. However, the handling of parallel, multi-modal datastreams provides several issues regarding data processing and system design, as discussed in the following paragraphs.

5.4.1 Synchronization & Subsampling

The synchronization of different sensor types represents a substantial problem, especially of non-visual sensors (accelerometers, etc.). It is advisable to introduce a synchronizing activity which is unambiguously identifiable in diverse data representations. In the introduced industrial application, a single hand clap has proved to provide useful data for synchronization as it shows explicit peaks in motions sensing and can precisely be timed also in visual and auditory sensors.

However, a single synchronization is usually not sufficient. Different sampling rates from the diverse sensor types require a sub-or re-sampling of data to combine single data snippets into collected data frames which are able to provide an overall representation of the scene over the various available sensors. Sometimes, when recording long sessions (¿1 hour), the differences in the internal clocks of the sensors may also cause significant shifts in the data, making re-synchronization in periodic time ranges advisable.

5.4.2 Dealing with the Uncertainty of Sensor Data

One of the most critical and difficult aspects of multi-modal sensor applications is the evaluation of data quality as this directly affects fusion of different data types. Some sensors directly provide measures of confidence of sensor data, while others require hand-made post-processing for the evaluation of data quality. These can range from rule-based evaluation criteria as application-based plausibility checks (e.g. avoiding jitter in hand tracking data by limiting the maximal distance between consecutive data frames) to statistical measures (check if data lies in the standard value range) or comparison of actual data with predictions from previous frames.

Such evaluation of data quality is required to dynamically select the sensors with the currently best and most reliable sensor data, hence is the main prerequisite for the fusion of redundant sensor data.

5.4.3 Fusion of Redundant Data

Based on an evaluation of incoming sensor data quality, the different data types can be merged via different weights based on the respective sensor data confidence. In the proposed application-scenario, a Kalman-Filter was used to combine skeleton data from two Kinect sensors and an RGB image sensor to calculate a merged, stabilized user skeleton for the adjacent behavior analysis approach.

6 CONCLUSION AND FUTURE WORK

In this paper, various sensors for the analysis of activities and cognitive states are introduced in the specific case of an industrial, semi-manual assembly scenario. The proposed sensors range from image- and depth-image-based infrastructural sensors to body-worn sensors of somatic indicators of behavior and cognitive state. For all sensors, a general description and evaluation regarding the experiences in the described industrial use-case have been provided, trying to help other researchers in their selection of suitable sensors for their specific research question.

The sensor discussion is followed with a general description of issues and challenges of sensors in industrial application scenarios, with a special focus on multi-sensor fusion.

The goal of future work is to realize a truly opportunistic sensor framework which dynamically can add and select sensors which provide the best data for the current application.

ACKNOWLEDGEMENTS

This work was supported by the projects Attentive Machines (FFG, Contract No. 849976) and Attend2IT (FFG, Contract No. 856393). Special thanks to team members Christian Thomay, Sabrina Amrouche, Michael Matscheko, Igor Pernek and Peter Fritz for their valuable contributions.

REFERENCES

Amrouche, S., Gollan, B., Ferscha, A., and Heftberger, J. (2018). Activity segmentation and identification based on eye gaze features. *PErvasive Technologies Related to Assistive Environments (PETRA)*. accepted for publishing in June 2018.

Asteriadis, S., Tzouveli, P., Karpouzis, K., and Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head poseapplication in an e-learning environment. *Multimedia Tools and Applications*, 41(3):469–493.

Baloch, Z., Shaikh, F. K., and Unar, M. A. (2018). A context-aware data fusion approach for health-iot. *International Journal of Information Technology*, 10(3):241–245.

Behrmann, E. and Rauwald, C. (2016). Mercedes boots robots from the production line. Accessed: 2017-02-01.

Burns, A., Greene, B. R., McGrath, M. J., O'Shea, T. J., Kuris, B., Ayer, S. M., Stroiescu, F., and Cionca, V. (2010). Shimmer–a wireless sensor platform for noninvasive biomedical research. *IEEE Sensors Journal*, 10(9):1527–1534.

Camilli, M., Nacchia, R., Terenzi, M., and Di Nocera, F. (2008). Astef: A simple tool for examining fixations. *Behavior research methods*, 40(2):373–382.

Chen, M., Ma, Y., Li, Y., Wu, D., Zhang, Y., and Youn, C.-H. (2017). Wearable 2.0: Enabling human-cloud integration in next generation healthcare systems. *IEEE Communications Magazine*, 55(1):54–61.

Critchley, H. D., Elliott, R., Mathias, C. J., and Dolan, R. J. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *Journal of Neuroscience*, 20(8):3033–3040.

Empatica (2016). comparison procomp vs empatica e3 skin conductance signal.

Fedor, S. and Picard, R. W. (2014). Ambulatory EDA: Comparisons of bilateral forearm and calf locations.

Frith, C. D. and Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological psychology*, 17(1):27–39.

Gabay, S., Pertzov, Y., and Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception, & Psychophysics*, 73(1):123–129.

Gao, X., Sun, Y., You, D., Xiao, Z., and Chen, X. (2016). Multi-sensor information fusion for monitoring disk laser welding. *The International Journal of Advanced Manufacturing Technology*, 85(5-8):1167–1175.

Gollan, B. and Ferscha, A. (2016). Modeling pupil dilation as online input for estimation of cognitive load in non-laboratory attention-aware systems. In *COGNITIVE 2016-The Eighth International Conference on Advanced Cognitive Technologies and Applications*.

Gradl, S., Kugler, P., Lohmüller, C., and Eskofier, B. (2012). Real-time ecg monitoring and arrhythmia detection using android-based mobile devices. In *Engineering in Medicine and Biology Society (EMBC), 2012*

Annual International Conference of the IEEE, pages 2452–2455. IEEE.

Graham, F. K. (1992). Attention: The heartbeat, the blink, and the brain. *Attention and information processing in infants and adults: Perspectives from human and animal research*, 8:3–29.

Gravina, R., Alinia, P., Ghasemzadeh, H., and Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, 35:68–80.

Haslgrübler, M., Fritz, P., Gollan, B., and Ferscha, A. (2017). Getting through: modality selection in a multi-sensor-actuator industrial iot environment. In *Proceedings of the Seventh International Conference on the Internet of Things*, page 21. ACM.

Jovic, S., Anicic, O., and Jovanovic, M. (2017). Adaptive neuro-fuzzy fusion of multi-sensor data for monitoring of cnc machining. *Sensor Review*, 37(1):78–81.

Kahneman, D. (1973). *Attention and effort*, volume 1063. Prentice-Hall Enlegwood Cliffs, NJ.

Kassner, M., Patera, W., and Bulling, A. (2014). Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*, pages 1151–1160. ACM.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance*, pages 279–328.

Kröger, M., Sauer-Greff, W., Urbansky, R., Lorang, M., and Siegrist, M. (2016). Performance evaluation on contour extraction using hough transform and ransac for multi-sensor data fusion applications in industrial food inspection. In *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2016*, pages 234–237. IEEE.

Lacey, J. I. (1967). Somatic response patterning and stress: Some revisions of activation theory. *Psychological stress: Issues in research*, pages 14–37.

Leykin, A. and Hammoud, R. (2008). Real-time estimation of human attention field in lwir and color surveillance videos. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–6. IEEE.

Li, X., Li, D., Wan, J., Vasilakos, A. V., Lai, C.-F., and Wang, S. (2017). A review of industrial wireless networks in the context of industry 4.0. *Wireless networks*, 23(1):23–41.

Marabelli, M., Hansen, S., Newell, S., and Frigerio, C. (2017). The light and dark side of the black box: Sensor-based technology in the automotive industry. *CAIS*, 40:16.

Otto, M. M., Agethen, P., Geiselhart, F., Rietzler, M., Gaisbauer, F., and Rukzio, E. (2016). Presenting a holistic framework for scalable, marker-less motion capturing: Skeletal tracking performance analysis, sensor fusion algorithms and usage in automotive industry. *Journal of Virtual Reality and Broadcasting*, 13(3).

Poh, M.-Z., Loddenkemper, T., Reinsberger, C., Swenson, N. C., Goyal, S., Sabtala, M. C., Madsen, J. R., and Picard, R. W. (2012). Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. *Epilepsia*, 53(5).

Potter, L. E., Araullo, J., and Carter, L. (2013). The leap motion controller: a view on sign language. In *Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, pages 175–178. ACM.

Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Wheeler, M. T., and Ashley, E. A. (2017). Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine*, 7(2):3.

Smith, K. C., Ba, S. O., Odobez, J.-M., and Gatica-Perez, D. (2006). Tracking attention for multiple people: Wandering visual focus of attention estimation. Technical report, IDIAP.

Srbinovska, M., Gavrovski, C., Dimcev, V., Krkoleva, A., and Borozan, V. (2015). Environmental parameters monitoring in precision agriculture using wireless sensor networks. *Journal of Cleaner Production*, 88:297–307.

Suriya-Prakash, M., John-Preetham, G., and Sharma, R. (2015). Is heart rate variability related to cognitive performance in visuospatial working memory? *PeerJ PrePrints*.

Thatcher, R. W. and John, E. R. (1977). *Functional neuroscience: I. Foundations of cognitive processes.* Lawrence Erlbaum.

van Dooren, M., de Vries, J. J. G. G.-J., and Janssen, J. H. (2012). Emotional sweating across the body: comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2):298–304.

Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393.