INTERACT – Interactive Manual Assembly Operations for the Human-Centered Workplaces of the Future

Grant Agreement Number : **611007**
: **INTERACT**
Project Start Date : **1st October, 2013**

Consortium : **DAIMLER AG (DAIMLER)- Project Coordinator**
ELECTROLUX ITALIA S.P.A. (ELECTROLUX)
INTRASOFT INTERNATIONAL SA (INTRASOFT)
IMK AUTOMOTIVE GMBH (IMK)
EMPHASIS TELEMATICS AE (EMPHASIS)
HADATAP SP ZOO (HADATAP)
UNIVERSITY OF PATRAS (LMS)
UNIVERSITAET ULM (IMI)
DEUTSCHES FORSCHUNGSZENTRUM FUER KUENSTLICHE INTELLIGENZ GMBH (DFKI)

Title	:	**Deliverable 1.2.1 - Requirements on monitoring of manual assembly operations**
Reference	:	**D1.2.1 of Task 1.2, WP1 – Version 1.0, based on DOW of 2013-08-13**
Availability	:	**Public**
Date	:	**2014-03-31**
Author/s	:	**DFKI, EMPHASIS, HADATAP, IMI**
Circulation	:	**Consortium, EC**

Summary:

Deliverable D1.2.1, created in Task 1.2.

This deliverable describes requirements on technologies used within the future INTERACT system for sensing and tracking manual assembly and work operations, both for the pilot cases and generic use.

It also gives an overview of available sensing technologies and evaluates them related to the requirements.

The INTERACT project (611007) is co-funded by the European Commission under the 7th Framework Programme.

This document reflects only authors' views. The Europena Commission is not liable for any use that may be done of the information contained therein.

Contents

1. Executive Summary

The content of this document is the outcome of INTERACT Task 1.2 “Requirements on monitoring of manual assembly operations”. The main purpose of this document is to:

- Asses the current state of the art in tracking technologies suitable for the INTERACT scenario.
- Define requirements on the tracking and sensing subsystem to be implemented in the INTERACT project, including hardware selection and software components.
- Differentiate between generic requirements for potential future INTERACT application cases and specific requirements for the pilot cases that will be implemented in the course of the project.

The document is structured as follows:

- Chapters 2 and 3 give a short introduction and overview to the problematic and pilot cases.
- Chapters 4 and 5 report on the state of the art in potentially applicable tracking and sensing technologies, namely optical sensing, MEMS sensors and RFID technology.
- Chapter 6 summarizes the key functional and non-functional requirements elicitated from the generic INTERACT scenario and the pilot cases, and also proposes technical solutions for the requirements to be considered in WP3. Key aspects of the requirements include:
    - Detail levels and performance measures of tracking humans, objects and tools
    - Additional sensors necessary for identifying key frames
    - Additional sensors necessary for ergonomic evaluation support
    - Software requirements for identifying key frames and sensor fusion support
- Chapter 7 gives an outlook to the upcoming next steps.

The main conclusions and summarized requirements are:

- Optical sensing will likely form the base of the tracking system, especially for human motion tracking.
- The human skeleton data will be tracked in a fine-grained form including hand skeletons, and will be enriched or improved through the data of additional sensors like IMUs.
- It is necessary to allow for arbitrary additional sensors such as switches to identify key frames, as these may not be always inferred from the pure motion data.
- It is necessary to employ additional ergonomic sensors, which measure e.g. forces, to enrich the human model for better ergonomic evaluation
- The sensor system needs to be robust and flexible, as well as easy to setup and calibrate, in order to be adaptable to different scenarios very fast and reliable.
- Data fusion components need to be implemented to deal with different types of incoming sensor data.
- An API for external data e.g. from industrial production systems needs to be defined in order to inject data from third-party tools or systems.

The outcome of this document shows that it is feasible to use tracking and sensing technologies within the INTERACT scenario, and provides input to the actual sensor systems design and implementation in WP3.

2. Introduction and Aim of the Document

This document aims to state the requirements for the tracking of human assembly operations with technological methods. As human assembly operations tend to be complex and manifold in movements and degrees of freedom, there exists a need to combine different tracking technologies to be accommodated to this complexity, as every technology has its own ideal operating conditions. For example, optical tracking solutions often suffer from limitations through occlusion, and thus, in certain situations need to be complemented by non-optical solutions like micro-electromechanical systems or other non-optical sensing systems. Furthermore, in certain environments, there are additional data sources like machines or tools, which already define a given API through which it is possible to access their internal state and data.

The base requirements definition is formed by a high-level description of the possible industrial use cases in chapter 3, which also gives a brief summary of the surrounding technological conditions.

Based upon this, within the following chapter 4, potentially trackable parameters and their specific properties relevant for tracking solution selection are listed, initially without directly referring to the use cases, but rather as a state-of-the-art technology report of current sensing methods for different entities.

In chapter 5, the state of the art in tracking work operations is evaluated with respect to the basic use cases and their general requirements. A variety technologies like optical sensing, MEMS and RFID technologies are considered here, as well as communication technologies and infrastructure types for connecting sensors and computing devices wireless and by wire.

As one of the long term goals of the INTERACT system is a general application within different industries for the tracking of work operations, in chapter 6 a generalized view of the requirements is given, with respect for the basic technological challenges of using multi-sensor systems in an industrial environment. The sections 6.2 and 6.3 bring together detailed pilot case information especially from the other deliverables within work package 1 with the generic requirements. The following section 6.4 points out possible technical solutions for different parts of the use case problem statements.

Finally, chapter 7 summarizes the findings of the deliverable and leads over to the actual work on the sensor systems, which is planned to take place in work package 3, and which has already been started in parallel to the requirements definition. This follow-up work package is supposed to take this deliverable as its main input together with the other requirements elicitated in work package 1.

3. SUMMARY OF INDUSTRIAL PILOT CASES

There are two main industrial pilot cases within the INTERACT project. The first is one is related to digital automotive production planning verification, the second one to warehouse component handling within the area of professional white goods. Both pilot cases will be described shortly in this chapter. For a more complete description please refer to deliverable D1.4.1.

3.1. Automotive pilot case

Within automotive assembly, ongoing development concentrates on cost reduction on the one hand, but on the other hand also on a time-to-market getting continuously shorter. To meet these demands, automotive companies are investing lots of resources for production planning and optimization. To verify this planning, until now mainly physical prototypes are used to carry out assembly tests and process optimizations.

Here, the ongoing developments in the domain of computer graphics, digital human modelling and motion capturing promise the possibility of replacing those prototypes with a digital verification and optimization process, by carrying out various process steps with the help of digital models of parts and workers. This leads to reduced cost for physical prototypes, as well as to shorter round-trip-times when testing out potential process configurations. Additionally, this allows to check for ergonomic issues semi-automatically or even completely automatic.

The results of the INTERACT project should therefore deliver a platform to support this digital verification process, both in terms of software infrastructure, and also in terms of a hardware setup. For example, it is necessary to establish an environment, which can handle various formats of 3D data for parts and tools, but also provides functions of state-of-the art digital human modelling. This allows to recreate the assembly setting in a digital environment, and to generate initial versions of process simulations automatically.

The additional hardware sensing setup allows to easily adapt and constrain movements within process steps by tracking real human movements as an input to the digital human model and ergonomic analysis, but also by tracking tool and part movements to support modification of the according objects in the digital environment.

3.2. Warehouse operator pilot case

In professional white goods production, similarly to the automotive domain a great amount of highly customized products are being manufactured. Anyway, short delivery times and efficient production processes are necessary to stay competitive.

Consequently, the warehousing area needs to be highly optimized to meet these demands. This situation even tightens with the introduction of new parts, which need to be checked in terms of storage position, handling / ergonomics and transport to the assembly lines. Currently, this is accomplished only by manual / physical observations and optimization through expert staff. This normally leads to a working, but often not thoroughly optimized solution.

The results of the INTERACT project should therefore enable a new kind of material handling planning process and ergonomics evaluation. For this, various 3D data of parts, warehouse layout and transport trolley construction need to be integrated to create a digital model of the warehouse setting, which then can be combined with digital human model technology to simulate and optimize the process of part picking and loading. This allows for better efficiency as well as optimized ergonomics for the warehouse operators.

As an input to this system, it is also necessary to assess the initial / current process by tracking the movements of parts and workers in a real world picking situation. This allows for comparisons between actual process and target process created from the digital human model, in terms of different movements and positions.

4. POTENTIALLY TRACKABLE PARAMETERS

4.1. What to track?

4.1.1. Body tracking

Body tracking refers to the acquisition of data that incurs through the presence of one or multiple humans within the sensing area. Trackable parameters range from simple presence detection to the tracking of complex biological features like vital signs or biomechanical features like the state of the musculoskeletal system and movement. This data, together with contextual information, forms the basis e.g. for inferring human activities and other semantically more significant parameters.

4.1.2. Tool tracking

Within a workshop scenario, usually multiple different tools are being used for assembly operations. This might include simple tools driven by muscular strength (e.g. wrenches) as well as complex, electrically driven power tools like electric and pneumatic screwdrivers. While the former only expose a few inherent parameters, the latter may already include various sensors available for tracking on their own, e.g. for forces or orientation, which are accessible via predefined interfaces (e.g. Bluetooth, WiFi).

4.1.3. Part and generic object tracking

Besides humans and tools, there are typically some other trackable entities involved in manual assembly processes: the parts to assemble, the partially assembled product, as well as other helper objects, which may not be included in the final product, for example protective packaging or handling aids. For most of these entities, data acquisition limits itself to identifying such objects, and to track basic parameters like position and orientation.

4.2. Directly trackable parameters

4.2.1. Trackable parameters relating to all tracked objects

Within the INTERACT scenarios, there is a set of basic trackable parameters common to all of the entities mentioned above:

- Presence detection / identification: Presence of an entity within the sensing area, optionally distinguishing different entities (persons, parts, etc.).
- Position in space (3D – X, Y, Z in global coordinates): Based on a reference point in the object's coordinate system, e.g. the center of mass.
- Orientation in space (3D – Pitch, Yaw, Roll in global coordinates): Depends on the objects own coordinate system.
- Dimension or (convex) hull dimension (3D)
- Contact / collision of objects: Includes picking, gripping and handling of tools and parts through human workers, and mounting operations where worker, tools and parts are in contact.
- Simplified contact geometries
- Weight / Material information: elasticity, texture, material

4.2.2. Trackable parameters related to the human body

Human bodies within the scenario additionally expose a large number of characteristic parameters that could be tracked:

- Kinematic model - skeleton data / posture: Position and orientation of body joints. The number and precision of tracking these joints heavily depends on the precision of the sampled data. Typically, only major skeleton joints are tracked, but it may also be necessary to track fine-grained skeleton data like hands and fingers.

- Visual attention and cognitive state: Through head, gaze and eye position as well as facial tracking. Also strongly related to skeleton data.

- Vital data: E.g. heart rate, blood pressure, temperature, electrical muscle activity (EMG, also relevant for skeleton data).

- Forces affecting human body and environment / tools / parts (kinetics model).

4.2.3. Trackable parameters related to tools

Tools in the INTERACT workshop scenario also might expose a few additional parameters besides the basic ones mentioned in 4.2.1:

- Multi-dimensional forces (applied by humans, gravity, self): This includes a range of different types of forces, like shearing, compression, torsion, tension or static/dynamic load.

- Accelerations of tool or parts of the tool, e.g. electric screwdriver rotational acceleration.

- Arbitrary intrinsic tool state: This includes mainly electrical or pneumatic tools, which often expose a number of settings or parameters, for example on/off state, torque limits, target rpm and actual rpm, health state or battery life.

- Arbitrary data transfer from industrial control systems: Industrial, logistic or robotic control systems often also expose a number of data interfaces, which might feed additional important tracking information.

4.2.4. Trackable parameters related to parts and objects

For other entities there may be arbitrary properties that need to be tracked in addition to the basic parameters of 4.2.1. This comprises the same multidimensional forces as explained in 4.2.3, also including accelerations.

In the special case of parts, one might need to identify or track a large number of parts at once to determine if all necessary parts were mounted to the final product. This needs to work in a contactless way, as the mounted parts may not be accessible for visual inspection anymore at the end of the production run.

Another aspect for tracking of objects is their actual movement within the scene. Some objects remain static in their position / rotation throughout the capturing process, they only need to be located once to match the virtual scene. Other objects like parts or tools are free to move during the capturing in the scene; they need to be tracked along with the human movements.

4.3. Resolution and selection of tracking methods and targets

4.3.1. Spatial resolution and synchronization aspects of tracking

Spatial data from tracked entities needs to be transformed to a common coordinate system to be able to successfully compute relations of objects, touch points or collisions. Based on this fact, the resolution of single data sources should at least allow for a similar range of resolution to integrate the values into the common coordinate system.

However, for some entities, it may be sufficient to track a relatively small number of possible discrete values for position / rotation. For example, the usage of a tool may be inferred from the tool not being at its “standby” position, so the tool position tracking might be done via a simple RFID “present” / “not present” sensing.

This also means that most of the tracking quality metrics need to be derived from the application domain itself, as more data does not necessarily implicate better tracking output. Instead, computational complexity and thus overall performance degrade with more (potentially unnecessary) raw data that needs to be interpreted and postprocessed / filtered. This concept also is closely related to the terms of *early* and *late* fusion in multimodal systems [1], but on a more abstract level.

4.3.2. Temporal resolution and synchronization aspects of tracking

To be able to fuse an arbitrary number of tracked data points from different sources, there is also a need for a common time basis as a means to synchronize the different data streams. This holds for static (one-shot) tracking as well as dynamic (continuous) tracking over a longer time. All used tracking subsystems need to be able to either synchronize to an external time base / clock signal, or to keep an internal clock which is sufficient precise and not exposed to significant drift (e.g. NTP[1]-based).

Another important temporal aspect of the INTERACT tracking setup is the temporal resolution of tracked data points. The lower the temporal resolution, the more interpolation between the tracked frames is needed. For tracking human actions, the necessary temporal resolution is typically limited by the maximum velocity of human movements, and ranges between 25 Hz (Cinematic capture of motions for movies) up to 120 Hz (commonly used eye-tracking data rate, also used for TV slow-motion recordings).

However this does not necessarily hold for tool-based actions using e.g. fast-revolving electrical screwdrivers or similar machines.

4.3.3. Selection of tracking targets and methods

At first glance, tracking methods might be selected based on their technical resolution and precision. However, this may lead to unnecessary large datasets containing plenty of information, which is not relevant for the particular domain-level semantic features the system actually needs to track.

To limit the amount of information that has to be processed within a given time slot, one should therefore consider selecting tracking targets and methods depending on the semantic meaning of the data which is expected from that particular part of the tracking setup. This means for example, that if the required information (expressed in the application domain language) only needs to indicate if a particular tool is in use or not, one should avoid to track full trajectories of this tool e.g. via depth cams, and refer to a simple RFID-based presence detection instead.

Moreover, based on the use cases and general project focus, not all possible trackable parameters mentioned in 4.2 are actually necessary to be tracked. As the use cases are normally embedded within a “controlled environment”, it is for example known which persons, tools and parts are present within the scenarios, and thus their presence doesn’t need to be tracked. Nevertheless, the arbitrary state and movement of such items is an important parameter whose tracking is crucial to the whole scenario.

[1] Network Time Protocol

4.4. Synthesized / inferred parameters

On an intermediate semantic level (e.g. not in the application domain semantics), additional parameters can be synthesized or inferred from the directly tracked data. This may complete or fill in missing data for lost data points like occluded skeleton joints, but also generate new higher-level data, e.g. by doing basic, domain-independent activity recognition (walking, standing, etc.).

Another inferred set of parameters may be a logical representation and relation between objects for the tracked human body, for example an automatic detection of tools and objects in the hands.

5. STATE OF THE ART IN TRACKING MANUAL SHOP-FLOOR OPERATIONS

5.1. Optical sensing

Various technologies for optical sensing and tracking are available for use within INTERACT. Within the following sections, technologies are grouped and described by their dimensionality within the spatial domain.

5.1.1. Virtual-1D optical sensing technology

1D sensing technology comprises various basic passive and active photoelectric sensors like ambient brightness sensors, distance sensing via reflection sensors, light barriers, color sensors, bar code scanners, (1D) position and contact sensing. Depending on their type, the output data varies between one digital or analogue value (e.g. distance sensing) and high-level digital data (e.g. barcode scanner). This also leads to a range of different possible processing steps, ranging from A/D conversion to linearization and offset removal.

Most of these sensors also exist with standardized interfaces for data retrieval and configuration, albeit on a mostly low level (e.g. SPI, I2C), and can be integrated in a multi-sensor setup with little additional control hardware.

5.1.1.1. Usage of 1D optical sensing technologies within INTERACT

Within INTERACT, 1D optical sensing technologies are mostly applicable for basic presence, position and proximity sensing, but could also be used for part or tools identification (barcodes) if any reason prevents from using a comparable non-optical technology e.g. based on radio technology.

5.1.1.2. Risks and challenges using 1D sensing technologies

Most of the mentioned technologies are very mature and widely used throughout various sectors of industry, for example in automation or robotics tasks. There exists a lot of experience and detailed knowledge of the application of those systems, so the risks and challenges can be considered as rather minor. Still, one potential challenge is to ensure coexistence with any other actively illuminating devices in terms of extraneous light, modulation frequencies and wavelengths.

5.1.2. 2D optical sensing technology

2D Optical sensing mostly relies on chip-based cameras like CMOS[2] or CCD[3]. These are commonly able to passively capture the light reflected back from a scene in visible and near-infrared spectrums. Also, through years of on-going development and the availability of more capable, fast interface technologies, the achievable resolutions have steadily increased.

2D cameras are available with a range of different specifications related to color space, spatial and temporal resolutions. Common parameter ranges for such systems are:

- **Color depth:** 8 bit grayscale (one channel black and white) up to 72 bit color (three-channel RGB full color, 24 bit per channel)
- **Spatial resolution:** 320x240 pixels, up to > 3840x2160 pixels
- **Temporal resolution:** 15 frames per second up to > 1000 frames per second

Regarding the color depth, it is sufficient for many computer vision tasks to capture black and white imagery, which often comes along with higher resolution both spatially and temporally at the same price, and which may be necessary to track very small or very fast movements and objects.

[2] Complementary Metal Oxide Semiconductor

[3] Charge-coupled device

Also, it is possible to use cameras with additional sensitivity in the near-infrared range, which enables invisible, unobtrusive use of additional infrared illumination or infrared markers as it is the case with most of the common professional marker-based motion capturing systems.

On the other hand, with RGB full-color imagery, it is possible to use color segmentation methods to separate objects of interest from the background (e.g. based on human skin color), or also to be able to better distinguish different objects or persons by their visual appearance.

Concerning temporal resolution, most of the available camera systems already operate within a range where most of the human movements can be monitored with adequate resolution. Special solutions with temporal resolutions higher than usual frame rates of are required for eye tracking systems and also tracking of tool movements faster than regular human movements.

With increasing temporal and spatial resolution and bigger color spaces, there are also stricter requirements for the interface and compression technologies which are necessary to handle and transport the increasing amounts of image data to other systems for further processing. Common solutions and their main benefits nowadays are, amongst others:

- **USB 2.0 / USB 3.0:** Plug and play, well supported and used widely
- **Gigabit Ethernet:** Long cable lengths, PowerOverEthernet support
- **Firewire:** Still used widely, but slightly out-of-date
- **CameraLink:** Common in industrial machine vision scenarios, robust and fast
- **HD-SDI:** Mostly used with TV and cinema applications

There is wide range of commercially available industrial cameras which can be used in INTERACT scenarios as standalone camera or part of a multi-camera setups. Pre-made multi-camera setups are less common, but also exist as professional (PointGrey Bumblebee) or consumer solutions (Leap Motion).

5.1.2.1. Usage of 2D camera systems within INTERACT

Within INTERACT, b/w and color camera systems mainly have two application areas.

First, multiple 2d cameras can be used for reconstruction of scene data in 3D, for example to track human body movements (see section 5.1.4.2). This requires a large amount of processing and data throughput capabilities, as there are usually lots of cameras necessary to gather the scene from enough different angles for sufficient correspondence between the different images.

In professional motion capturing systems, this data bottleneck is bypassed through extensive preprocessing tasks already done in-camera or on special intermediate hardware, which often reduces the camera image to nothing but a compressed image with the recognized features, and thus significantly reduces the data volume that needs to be transferred.

Second, 2D cameras can also be useful to track prominent features of parts or objects, which cannot be seen by depth cameras or other sensors (e.g. colored markers). This very common computer vision task is well supported amongst different software systems and libraries, and allows for a relatively fast and easy tracking of objects with cheap and common 2D cameras.

5.1.2.2. Risks and challenges when using 2D camera systems

Feature tracking with common camera systems have been used and researched for a long time, however the passive nature presents some challenges to the application of such camera systems within INTERACT. Variable and poor lighting conditions have to be taken into account, as well as non-linearities in camera systems like lens distortion. Furthermore, as with 1D optical sensing, extraneous light from other active devices has to be considered when using camera-based solutions.

5.1.3. 3D optical sensing technology

The main goal of the 3D optical sensing technologies specified below is the computation of a so-called depth map, which contains a distance value for every pixel in the map. This information allows for multiple follow-up processing steps, for example to generate a 3D model or imagery of the scene or for using the depth information as an additional input for segmentation and recognition when tracking persons or objects.

With the exception of stereo- or multicamera setups for 3D sensing, most of the systems for depth map acquisition are active devices which need to illuminate the scene themselves. This yields the advantage of equal performance also with low-light conditions, but also the disadvantage that achievable ranges are often limited due to low-power illumination sources. Besides that, additional occlusion can occur when objects are casting shadows from the light source.

5.1.3.1. Structured light sensing

Structured light (SL) depth imaging is an active sensing technology, which relies on projecting known structured patterns onto an area of interest with a laser or regular light projector, and captures the reflection of the pattern with a calibrated camera. From this image and the known pattern, a depth map can be computed through correspondence solving and triangulation.

Early structured light systems only worked in non-realtime by static light pattern projection, and were mainly used for 3D scanning of fixed objects, mostly in industry applications. However, more recent developments using laser projectors and the ever-growing available computing power also with consumer-grade hardware have enabled new areas of use and new devices, of which the Microsoft Kinect Camera was the first SL camera available to consumers at a comparatively cheap price, mainly targeted at the gaming market.

The Kinect (and most of the currently available SL cameras in the lower price range) is based on technology developed by PrimeSense, a company which was recently acquired by Apple Inc., but prior to that, also distributed their own sensor brand named *Carmine.*

The PrimeSense technology works through the combination of an infrared laser projector, which projects a known speckle pattern onto the scene, and a near-infrared camera with a given distance to the projector, which acquires a scene image with the projected pattern. The image processing hardware then computes a depth map, which can be accessed (often along with additional audio and color camera information) as an USB data stream.

Based on the sensor configuration (both in optical and software terms), operation ranges of PrimeSense devices start at 0.15m (Carmine short-range) and reach up to around 7m (Kinect long-range data).

5.1.3.1.1. Advantages and disadvantages of SL systems

The main advantage of current SL sensors is their cheap price and good availability, compared to other 3D machine vision solutions. Through this, it is possible to deploy multiple cameras on a large scale without very high costs. However, operating multiple SL cameras creates multiple other challenges, which are further addressed in 5.1.3.1.2.

Disadvantages are mainly the large size of the devices, as it is necessary to keep a certain amount of offset between camera and projector. This also leads to the aforementioned shadowing effects as another disadvantage, which is inherent to SL technology. However, when compared to motion capturing setups, the dimensions of current SL depth camera systems still can be considered as relatively small.

Another disadvantage common to most of the optical 3D sensing technologies is decreasing performance when used together with specular or mirroring surfaces, because the receiving camera is either not able to capture any of the projected pattern at all, or the pattern is scattered back to the camera via multipath propagation, which also makes it impossible to calculate a valid depth map.

5.1.3.1.2. Potential usage of SL cameras in INTERACT

Within INTERACT, SL depth cameras can be used for various tasks. The first and most common use case for depth cameras is full-body human skeleton tracking. Besides this, it is also possible to track single features or only partial skeletons (e.g. hand skeletons) for later fusion with features tracked in other ways. Through the active illumination technology, it is also possible to track movements in areas which suffer from low light levels without the need for additional light sources.

5.1.3.1.3. Risks and challenges when using SL cameras

Through extensive research with depth sensors during the last few years, there is generally a good understanding of common usage scenarios and limitations of such systems. However, within INTERACT, it is possible that the scenario demands for new ways to deal with SL cameras; for example it might be necessary to implement solutions with multiple SL cameras within the same scene. This leads to some new problems, for which currently no ready-made solution is available, and which are currently still under research:

- **Increasing data rates and volumes:** Depth cameras produce large amounts of data, which need to be transported, stored and processed. Methods are necessary to handle this data, for example through distributed computing approaches.
- **Interferences due to active illumination:** When using multiple depth cameras with overlapping frustums, the projected patterns of the cameras can also be seen by other cameras, which increases erroneous depth samples and thus also decreases the SNR. Possible countermeasures may be optimized camera placement or time / wavelength multiplexing on the hardware side, as well as post-processing algorithms to minimize the impact of interference artefacts on the tracking algorithms. Both approaches have already been researched to a certain extent ([11],[12]).
- **Calibration, coordinate mapping and fusion:** Depth data is generally produced within a camera-local coordinate system, which has no relation to the “global” world coordinate system besides common measurement units (usually metric units). To be able to fuse the data produced by different cameras, it is necessary to establish a reference coordinate system, which may (but does not need to) be real-world / global. In another step, the output of all cameras then needs to be translated to this coordinate system. Several calibration procedures exist to accomplish this steps as mentioned in [13] and [14].

  With this transformation, it is possible to combine information from multiple cameras to gain a more detailed depth view. There are different approaches for this fusion step, mainly differing in their respective *fusion level* – see chapter 5.1.5 for further explanations.

5.1.3.2. Time-of-flight sensing

Time-of-flight (ToF) sensing relies on the known and constant speed of light (c = 299 792 458 m/s). This type of camera also actively illuminates the captured scene, but contrary to structured light methods, the runtime of the emitted light to the target and back to the camera is measured to gain the distance for a given pixel in the camera image. After putting all this pixels together, this results in a depth map similar to the ones that are generated from structured light systems [15].

ToF cameras and methods have been used for many years in research and development, but have not been available to end users until the last few years because of the necessary high-speed circuit technology. In the meantime, there are several different products available on the market which use ToF technology. Examples are long- and short range cameras from Belgian camera manufacturer SoftKinetic, as well as Microsoft’s upcoming successor to the first Kinect, the Xbox One Kinect, which relies on ToF technology.

5.1.3.2.1. Advantages and disadvantages of ToF systems

ToF cameras have several advantages when compared to SL cameras. First of all, there is no need for an offset between camera and projector because there is no triangulation involved. This means that camera and projector can be mounted collinear, which on the one hand allows for smaller devices, but on the other hand also eliminates the shadowing effect known from SL cameras.

Second, depth values are measured quasi-direct in the hardware, without the need for expensive correspondence matching or triangulation in terms of computing effort.

Another advantage is the resolution and overall image quality that can be yielded with ToF technology.

Besides that, it is easier to create multi-camera setups in terms of interference handling, as it is possible to allow coexistence between multiple cameras by changing their modulation frequency. Unfortunately, this is often only possible with more expensive ToF camera models like the Mesa SwissRanger SR 4500 industrial ToF camera.

One downside of current ToF-based cameras is their lower depth accuracy when compared to SL or stereo vision systems [15]. Another problem that ToF cameras share with SL cameras is their sensitivity to specular surfaces and multipath propagation, which can distort depth measurements.

5.1.3.2.2. Potential usage of ToF cameras in INTERACT

ToF cameras can be used within INTERACT as a substitute for SL cameras, in the same ways that SL cameras would be used (see 5.1.3.1.2). Also, it may be possible to add ToF cameras to avoid interference of additional SL cameras.

5.1.3.2.3. Risks and challenges when using ToF cameras

The risks and challenges of using ToF technology only differ in one point from those mentioned for SL in 5.1.3.1.3. As it was already pointed out, it is possible to alter the modulation frequency and modulation type of ToF cameras for interference avoidance, so interference handling is rather uncomplicated with ToF cameras, given they can be reconfigured to other modulations. Unfortunately, this is a feature, which consumer-grade ToF cameras often lack.

5.1.4. Tracking methods

Numerous methods have been developed to extract distinct visual features out of 2D and 3D scene data, and also to track movements of these features over time. The following paragraphs provide an overview of some selected methods possibly suitable for usage within INTERACT.

5.1.4.1. Marker tracking (on 2D imagery)

Tracking of different types of CV-suitable markers is a task, which has recently received increased attention through numerous augmented reality applications especially with the boost in smartphone technology over the last years.

Commonly, markers are simply graphical patterns optimized for computer vision systems by choosing shapes and colors which allow for fast and reliable detection, identification and tracking through common CV algorithms. They expose certain distinct features, like black-and-white square regions, which, on the one hand can be easily detected in camera imagery even under poor lighting conditions, but on the other hand also are able to carry a certain amount of coded information allowing for unique identification.

With four known, non-collinear points like the corner points of such a marker, it is also possible to estimate its position and rotation in 3D space, relative to the camera’s coordinate system. This is commonly used within AR applications to augment additional content within a live camera stream, but can also be used to track the position and orientation of objects that are equipped with such markers.

Another common use case for marker based tracking is also professional motion capturing, which uses non-coded, retroreflective IR markers and additional IR light sources, which makes the markers easily identifiable in camera imagery because of their outstanding brightness. The problem of location and orientation calculation is solved here by using multiple calibrated cameras looking at the scene from different angles.

5.1.4.1.1. Potential usage of marker tracking in INTERACT

Marker tracking can be used within INTERACT to track objects and parts within the scene by using the available 2D camera data. It is also possible to use coded markers to distinguish objects with similar visual appearance, which otherwise could not be separated by object detection and identification methods.

Additionally, within the earlier phases of the sensor system setup, marker-based motion capture systems may be used to produce ground truth data for sensor system performance evaluation as well as for initial movement sample collection.

5.1.4.1.2. Risks and challenges with marker tracking techniques

Most 2D-visual tracking methods suffer from some pitfalls due to the passive nature of the vision components, like poor illumination or noise. However, as the INTERACT scenarios mostly provide controlled industrial environments, it is possible to accommodate for these drawbacks in most of the cases to a certain degree, or to choose sufficient alternative solutions for cases that are suspect to poor performance of marker-based solutions.

5.1.4.2. Skeleton tracking and pose estimation

Skeletal tracking and human pose estimation is well researched yet still challenging within the computer vision area. It comprises tracking of body movements on different scales, ranging from large scale movements of the whole body and limbs down to small scale movements which occur with hands and fingers, or facial expressions.

Results of this research are applied within a wide field of applications, naming movie production, entertainment and gaming, surveillance tasks, medical research and therapy, amongst others.

Early tracking solutions mostly relied on marker-based methods, like they still can be found in professional motion capturing systems. Marker-based tracking solutions deliver detailed and very precise measurements of the various body joints, but also entail a few disadvantages. First, it is necessary to apply markers to the human body, or to wear a special mocap suit with the markers already attached at predefined positions. Second, most of the professional mocap solutions are still very expensive due to the number of high-speed, high-resolution cameras that are necessary to capture a scene in sufficient detail.

More recent tracking approaches realize so-called markerless motion capture. This method operates through direct detection and classification of body parts, respecting their spatial relationships and appropriate anthropomorphic models, which mostly leads to hierarchic models of the human body and its joints, each having certain degrees of freedom, depending on joint type.

Most available depth camera systems are supported by their respective SDKs and software environments, which also mostly provide modules to track human movements, body parts and different numbers of body joints, as this is one of the essential features e.g. for gaming applications, where the whole body acts as an input device to the game.

When tracking human motions, it is possible to identify certain common steps in the various implementations (loosely based on [16]):

- **Preprocessing the source imagery:** Preparing image data through means of contrast, color or brightness enhancements, or through application of algorithms like edge detection, mostly depending on the following processing steps. This step is mostly unnecessary with depth cameras.
- **Segmentation:** Identifying and extract humans within the 2D / 3D imagery. This is accomplished either using spatial features (searching for body silhouettes within the image, based on color or brightness differences) or temporal features (e.g searching for moving objects in front of a static background).
- **Tracking:** Track identified humans across multiple consecutive frames and over time. This is accomplished through finding corresponding points in the frames, often also supported by predictive operators to reduce the search space and thus computational complexity.
- **Pose Estimation:** This step tries to derive a skeletal configuration of the humans that have been tracked in one or more frames. This either involves models of human skeletal systems at different levels of detail, or large sets of training data, especially when used with 2D imagery.
- **Recognition:** Follow-up step often carried out to identify basic human activities depending on the pose over time. In INTERACT, this is considered in later steps, after the sensor fusion with IMU and other data has already taken place.

5.1.4.2.1. Potential usage of skeleton tracking in INTERACT

Skeletal tracking and human pose estimation play a crucial role within the INTERACT project. To recognize and classify the workers movements and actions, it is necessary to realize a fine-grained tracking setup, which delivers results comparable to regular, marker-based motion capturing systems in the situations given through the use case definitions. It is necessary to obtain a skeletal model including finger and foot movements, which ideally should allow for a 1:1 mapping to common digital human models from the ergonomic analysis domain.

To accomplish this, it is necessary to combine existing technologies in order to use their respective strengths. One possible approach might be the combination of short- and long-range depth sensing devices to obtain parts of the human skeleton at different detail levels, and to combine them via spatial fusion algorithms into one skeleton model, which can be further processed and enriched with information from other sensors (e.g. MEMS).

Another approach considered is the combination of multiple cameras or camera sets, to accommodate to larger tracking volumes in industrial environments and occlusion through parts, workers or other objects in the field of view of one or more cameras.

5.1.4.3. Other parameters trackable through vision sensors

Besides tracking objects and humans, it is also possible to make use of other features of the image data under certain circumstances, for example:

- **Biometric identification:** e.g. facial identification
- **Vital data:** e.g. remote heart rate sensing (Kinect 2, Eulerian Video Amplification [17])
- **Eye / Gaze direction**

5.1.5. Fusion approaches

As already mentioned before, in the field of multimodal interaction and multi-sensor fusion, multiple concepts and levels of fusion have emerged. Within multimodal interaction, it is generally differentiated between *early* and *late* fusion. Within the general sensor fusion area, fusion levels range from *raw data* fusion over *feature level* fusion up to *decision level* fusion, with the latter two being roughly equivalent to the fusion terms from multimodal interaction [18].

5.1.5.1. Raw data fusion

Raw data fusion can only be employed with commensurate sensor data. As the fusion only takes place at the raw data level, no semantic information from the datasets is used for fusion. A possible example is the combination of two point clouds from two depth cameras after aligning them through ICP- or RANSAC-based approaches.

5.1.5.2. Early fusion / Feature level fusion

Early fusion refers to a fusion process that only takes recognized distinct features into account which were captured through the different modalities or input devices. Within multimodal interaction, this fusion method is generally recommended for time-synchronization-sensitive applications, like the fusion of recognized lip movements with the output of a speech recognition component.

5.1.5.3. Late fusion / Decision level fusion

Late fusion takes place after determining semantically relevant information from the source data, like position or other attributes. This is often accomplished by weighted decision methods or additional AI components.

5.1.5.4. Focus and potential usage of fusion technology in INTERACT

Within INTERACT, fusion might take place on a few different levels. First, a general scenario-specific fusion for the whole sensor system has to be implemented, as already outlined in task 3.3 of WP3. Second, it might be useful to employ another rather low fusion level to combine e.g. data from multiple depth cams into one point cloud beforehand, so skeleton recognizers can operate on this data before delivering the skeleton data to the high level fusion component.

5.2. RFID, other sensors and infrastructure for location and identification

There are many different standards related to Radio Frequency IDentification (RFID).

Within this chapter, different RFID technology standards are described and compared with regard to their features, reading distances, accuracy and area of applications.

There are many definitions of RFID, but for the purpose of this deliverable, we use the one stated below:

Every time when radio frequencies are used to identify and locate a tagged object, we talk about RFID technology. So RFID technology is each technology based on RF which is enabling the process of identification and / or localization.

We also treat most RTLS (Real Time Locating System) as a specific type of RFID application, but only in the cases when RF are used as a core technology for real time localization.

RFID allows for identification (is the tagged object in reading area?) and for some solutions also for 2D and 3D location (where in the reading area is the tagged object currently located?).

5.2.1. RFID - general description, infrastructure, applications

Basic infrastructure of RFID consists of:

- **Tags:** These are used to tag object that is intended to be identified, the tag is data carrier and in it unique number describing object is stored
- **Readers:** These are used to read data from tags.

Tags and readers are accessible in different forms depending on the use case. There are accessible solid tags as well as inlays, and the physical form is dedicated to use case and specific application. Readers are accessible in a form of mobile readers (PDAs) as well as fixed readers integrated in portals, gates, barriers, kiosks etc. depending on designed process and reading point. Readers could be integrated with antennas, but it is also possible to configure specific readers with different antennas (with certain characteristics) depending on the use case.

Most popular use cases for RFID applications were categorized, analyzed and described as RFID Reference Model [21]. The model also contains such information as frequencies used in different application fields and remarks are related to design challenges. For examples of specific RFID application we refer to the model [21], RFID Journal website (http://www.rfidjournal.com) and RFID Knowledge Database from IDTechEx (where you can find over 4660[4] use cases from different categories such as: Airlines and Airports, Animals and Farming; Books, Libraries, Archiving; Financial, Security, Safety; Healthcare; Land and Sea Logistics, Postal; Laundry; Leisure, Sports; Manufacturing; Military; Oil Gas Extraction, Mining; Research; Passenger Transport, Automotive; Retail, Consumer Goods; Other http://www.idetechex.com/knowledgebase).

Distribution of use cases through application fields listed in RFID Knowledge base is shown in figure 2. It is worth to notice that there are numerous applications in automotive, manufacturing and consumer goods, which leads to the conclusion that RFID technology is capable to operate successfully in such environments (which are similar to pilot cases chosen for INTERACT i.e. workshop and warehouse operations).

To recapitulate, popular applications of RFID in industrial environments are:

- Work-in-progress identification/tracking
  - could be based on RTI (return transport item), pallet or item-level (for valuable objects) tagging;
  - tags could be read when crossing doors, portals, conveyors, workstations etc.;

[4] accessed: 28.01.2014

- benefits could be: better production plans, less overproduction, shorter lead times, improved visibility and transparency of manufacturing operations, reliable data for KPI calculations, decrease of manual identification and data entry activities;
- Warehouse operations;
    - that could be identification of products for in-bound and out-bound operation, also could be based on RTI, pallet or item-level tagging;
    - tags could be read when crossing delivery dock doors, when placed in warehouse locations etc.;
    - benefits could be: better inventory management, better inventory planning, lower stock levels, lower out of stock levels, improved visibility and transparency of warehouse operations, reliable data for KPI calculations, decrease of manual identification and data entry activities;
- Tools managements;
    - tools are tagged and tracked as they are moved and used on shopfloor.

	RFID-Application Field	Subcategories
Mainly Object Tagging	A. Logistical Tracking & Tracing	AA. - Inhouse Logistics; AB. - Closed Loop Logistics; AC. - Open Logistics; AD. - Postal Applications; AE. - Dangerous Goods Logistics; AF. - Manufacturing Logistics
	B. Production, Monitoring and Maintenance	BA. - Archive Systems; BB. - Asset Management (incl. Environmental Monitoring); BC. - Facility Management; BD. - Vehicles; BE. - Airplanes; BF. - Automation / Process Control; BG. - Food and Consumer Goods
	C. Product Safety, Quality and Information	CA. - Fast Moving Consumer Goods; CB. - Electronic Goods; CC. - Textile Goods; CD. - Fresh/Perishable Foods; CE. - Pharmaceutical; CF. - Customer Information Systems
Tagging with Reference or Potential Reference to Individuals	D. Access Control and Tracking & Tracing of Individuals	DA. -Ticketing; DB. - Access Control Systems; DC. - Animal Tracking; DD. - Personal Tracking
	E. Loyalty, Membership and Payment	EA. - Loyalty Cards; EB. - Membership Cards; EC. - Contactless Banking Cards; ED. - Payment and Advertising via mobile phones
	F. eHealth Care	FA. - Assistance for the Disabled; FB. - Hospital Management; FC. - Implants; FD. - Medical Monitoring; FE. - Smart Implants
	G. Sports, Leisure and Household	GA. - Sports Applications; GB. - Rental Systems; GC. - Smart Games; GD. - Smart Home
	H. Public Services	HA. - Public Service Maintenance; HB. - Road Tolling Systems; HC. - Banknotes; HD. - ID Cards and Passports; HE. - Health Insurance Cards

Figure 1: RFID reference model [21]

5.2.2. Typology of RFID

<table>
<tr><th>Criterion</th><th>RFID systems' types</th><th colspan="2">Description</th><th>Frequencies</th><th>Standards</th><th>Typical areas of applications</th><th>Reading distances</th></tr>
<tr><td rowspan="3">Reading rule</td><td>active</td><td colspan="2">tags transmit independently on reader, with own battery</td><td>433 MHz; 2,45 GHz; 5,8 GHz</td><td rowspan="3">see row “frequency” of this table</td><td>RTLS, long distances, sensors</td><td>up to 1000 m</td></tr>
<tr><td>passive</td><td rowspan="2">tags transmit signal when interrogated by reader</td><td>without battery</td><td rowspan="2">LF
HF
UHF (860-960 MHz)</td><td>logistics, access control, registering working time</td><td>1 cm – 10 m</td></tr>
<tr><td>Battery Assisted Passive</td><td>BAP, with own battery</td><td>logistics, long distances</td><td>over a 30 meters</td></tr>
<tr><td rowspan="7">Frequencies</td><td>LF (Low Frequency)</td><td colspan="2">passive, BAP, LF does not allow to read/write of many tags at the same times</td><td>125-134 kHz</td><td>ISO 11784, ISO 11785, ISO 14223, ISO/IEC 18000-2</td><td>access control, tickets, registering working time, animals' identification</td><td>up to 50 cm</td></tr>
<tr><td>HF (High Frequency)</td><td colspan="2">passive, BAP,
R/W of many tags simultaneously</td><td>13,56 MHz</td><td>ISO/IEC 14443, ISO/IEC 15693 , ISO/IEC 18000-3</td><td>access control, tickets, registering working time, libraries, automatics</td><td>up to 60 cm</td></tr>
<tr><td rowspan="3">UHF (Ultra High Frequency)</td><td colspan="2">active, possible communication of tags (meshing)</td><td>433 MHz</td><td>DASH7, ISO/IEC 18000-7</td><td>RTLS, sensors, military</td><td>up to 1 km outdoor</td></tr>
<tr><td colspan="2">passive, BAP, longest read range of all passive frequencies, reading many tags simultaneously</td><td>860-960 MHz (865,6-867,6 MHz EU)</td><td>ISO/IEC 18000-6, EPC Class0, EPC Class1, EPC Class1 Gen2, EPC Class1 Gen3 (BAP)</td><td>supply chain, warehousing, work in progressive, inventory, passive RTLS (pRTLS)</td><td>up to 15 m, (BAP over 30 m)</td></tr>
<tr><td colspan="2">active, possibility of Wi-Fi (ISO 802.11) compatibility</td><td>2,45 GHz</td><td>ISO 18000-4, ISO/IEC 24730-2</td><td>RTLS, assets and personnel tracking and identification in hospitals and mines</td><td>up to 200 m</td></tr>
<tr><td>UWB</td><td colspan="2">active, multi bands</td><td>multiple bands of frequencies simultaneously (3.1-10.6 GHz)</td><td>ISO/IEC 24730-6</td><td>RTLS accuracy up to 30 cm, sensors</td><td>up to 300 m</td></tr>
<tr><td>SHF (Super High Frequency)</td><td colspan="2">active, smaller and more effective than 433 MHz and 2,45 GHz, longer battery lifecycle</td><td>5,8 GHz</td><td>ISO/IEC 18000-5</td><td>RTLS</td><td>up to 200 m</td></tr>
<tr><td colspan="8">Writing data: read only (RO), read/write (RW), write once read many (WORM)</td></tr>
</table>

5.2.3. Description of standards and technologies for RTLS

RTLS - Real Time Locating System, is a system that is capable to determine coordinates of an object that is being tracked.

There are many different technologies used for RTLS. Many of them use RF for location. The choice of the particular technology version is dependent on many factors such as company environment, expected range, accuracy and granularity, expected lifecycle of tag/asset, existing infrastructure and technology portfolio, and others, and of course costs too. Primary technologies used for RTLS are [19] (but are not limited to):

- **Active RFID:** Determination of location via communication with a battery assisted RF-based tags capable of transmitting and/or receiving information independent of the reader (e.g., ISO 18000-7, ISO 24730, UWB);
- **Passive RFID (pRTLS, i.e. EPC UHF):** Determination of location by receiving data of power and time from an array of RF readers that interrogate passive RFID tags, (e.g., EPC Gen2, ISO 18000-6) within an area of coverage; capability of locating by using sectored antennas to determine Angle of Arrival (AoA).
- **GPS (Global Position System):** Determination of location based on GPS infrastructure: GPS devices and GPS satellites;
- **Assisted-GPS (A-GPS):** Determination of laction based on GPS (see bulet above) and cellular infrastructure (towers); more accurate than pure GPS and capable to serve indoor;
- **Wi-Fi:** Kind of active solution (see first bullet); determination of location via analysis of signal strenghts and times from Wi-Fi access points and Wi-Fi enabled devices (e.g. Wi-Fi based active tags;
- **Out-of-Band (OoB) proprietary RF/sensor-based:** Determination of location based on utilization of proprietary solutions and technologies such as ultra-wide band (UWB/IEEE 802.15.4f), infrared and/or ultrasound.

Another example is LIDAR which is based on illuminating an object with laser and calculations / analysis of reflected light.

According to the requirements stated during consortium meeting in Pordenone, the localization system needs to give a location and orientation in space with an accuracy of a few centimeters for distance coordinates. It is not sufficient to determine that an object is exactly in a defined area, but there is also the necessity to determine what is its orientation.. None of RTLS systems is able to deliver both. Use of typical RFID tags does not allow to determine orientation, because none of studied technologies for RTLS deliver data about rotation. The only possible solution is a hybrid solution e.g. merging active RFID RTLS systems with orientation sensors.

Only UWB out of the RTLS-related technologies is able to deliver an accuracy of ca. 30 cm.

Depending on the technology used, one or combination of more different methods (ranging and/or angulating) are used to determine location. These methods could be, but are not limited to:

- Angle of arrival (AoA),
- Time of arrival (ToA),
- Time of flight (ToF),
- Time difference of arrival (TDoA),
- Received signal strength indicator (RSSI).

There are many papers on those methods, e.g. [20].

5.2.4. Potential usage of RFID technologies in INTERACT

Referring to RFID technologies described in ch. 5.2.2 and RFID technologies used for RTLS (see ch. 4.2.3) we sum up possible usage of RFID in INTERACT in below-presented table. For the description of possible trackable parameters refer to ch. 4 and for different RFID technologies refer to ch. 5.2.2.

<table>
<tr><th>Possible trackable parameters</th><th>RFID technology</th><th>Object</th><th>Remarks</th></tr>
<tr><td>• Presence detection / identification</td><td>• Passive LF
• Passive HF
• Passive UHF
• Other</td><td rowspan="2">All tracked objects</td><td>• Choice of frequency depends on environment and expected reading distance (see ch. 4.2.2)
• Passive are most popular when only identification is expected, as they are the cheapest</td></tr>
<tr><td>• Position in space</td><td>• Passive UHF
• BAP
• Active (e.g. WiFi, DASH7, SHF)
• UWB</td><td>• Accuracy of position in space up to 1m (UHF, BAP, active)
• Accuracy of position in space up to 30cm (UWB)
• Z coordinate is problematic and needs special dedicated algorithms and infrastructure configuration
• Meshing possible (DASH7)</td></tr>
<tr><td>• Vital data</td><td rowspan="6">• Active (e.g. WiFi, DASH7, SHF)
• BAP
• UWB</td><td rowspan="2">Human body</td><td rowspan="4">• Proprietary
• Only when integrated with appropriate sensors
• No off-the-shelf tags were found
• Serves only as communication and identification module</td></tr>
<tr><td>• Forces affecting</td></tr>
<tr><td>• Multi-dimensional force</td><td rowspan="4">Tools</td></tr>
<tr><td>• Accelerations of tool or parts of the tool</td></tr>
<tr><td>• Arbitrary intrinsic tool state</td><td rowspan="2">• Need integration with a specific tool and/or system
• Serves only as communication and identification module</td></tr>
<tr><td>• Arbitrary data transfer from industrial control systems</td></tr>
</table>

Trackable parameters vs RFID technologies

RFID / Parameter	Passive LF	Passive HF	Passive UHF	BAP	active	UWB
Presence detection/ identification	X	X	X	X	X	X
Position in space			X	X	X	X
Vital data				X	X	X

Forces affecting						
Multi-dimensional force						
Accelerations of tool or parts of the tool						
Arbitrary intrinsic tool state						
Arbitrary data transfer from industrial control systems						

5.2.5. RF communication

Many standards can be used for wireless communication of sensors or objects (e.g. tools). These are for example:

- Bluetooth, 2,4 GHz, class 2, class 4, LE,
    - range: up to 50 m,
- DASH7, ISO 18000-7, 433 MHz,
    - range: 200 m, up to 1 km outdoor,
- RFID UHF BAP, 860-960 MHz,
    - range: over 30 meters
- Wi-Fi, IEEE 802.11, 2,4 GHz / 5 GHz,
    - range: up to 90 m (300 m outdoor),
- ZigBee, 868 MHz/ 915 MHz/ 2,4G Hz, IEEE 802.15.4,
    - range: up to 100 m
- Long distances as:
    - WiMax, IEEE 802.16,
        - range: up to 10 km,
    - LTE,
    - GPRS and others.

5.2.6. Findings

To gain accuracy of centimeters it would be necessary to apply RFID passive technologies with short range reading distances (LF, HF, near field UHF with reading ranges of centimeters). Then, the RFID reader gives information that the object is within the assigned area with an accuracy of centimeters. Another solution is UWB RTLS system with orientation sensors and an accuracy of approx. 30 cm.

This kind of solution needs to be combined with other technologies to determine rotation and orientation of an object or body part (such as wearable sensor integrated with RFID tags).

This solution needs wires for each and every point where object or body part could be placed. On the other side same effect can be gained with optical sensing systems if cameras cover exactly the area where we want to determine location and orientation of an object or body part.

5.3. MEMS, sensors and RF sensor networks

5.3.1. MEMS technology, sensors and shop-floor applications

As described in section 4.2 the directly trackable parameters that can be used in order to monitor manual shop-floor operation include a variety of parameters related to human body, tools and parts. Besides the use of optical technologies that can be used mainly for motion tracking purposes, the use of other sensor systems, with a focus at MEMS based technologies, that can be used within INTERACT, is being described in the following sections.

5.3.1.1. MEMS technology

According to the "MEMS Industry Group" association [1] a Micro Electro Mechanical System is "an enabling microfabrication technology that uses manufacturing processes similar to that of semiconductors and integrated circuits to create discrete or integrated microdevices such as mechanical structures, microsensors, microactuators, and circuitry on a substrate material including silicon, glass or ceramic."

This technology began as research in the 1970s and became an industry in the 1990s. In the last 15 years with the growing need for small, energy efficient, and low cost sensors and actuators in the handheld devices markets (smartphones, tablets, etc) has led to a big development in the MEMS based industry. Nowadays, it is very difficult to find a product in those markets that does not contain any MEMS based components.

5.3.1.2. MEMS sensors

Based on the MEMS technology the semiconductor industry was able to develop a variety of sensors that are available for various applications in the following areas, such as:

- Inertial sensing
- Environmental monitoring
- Chemical procedures
- Biomedical monitoring
- Communications

As mentioned above the main driving points for developing such sensors were:

- Small sensor size, which makes them ideal for modern application in handheld and wearable systems where space considerations are of very high importance
- Enhanced sensor behavior due to higher sensitivity, better linearity, better responsivity, and dynamic range
- Low power requirements due to the fabrication technology
- Low sensor cost due to batch fabrication
- Ability to integrate sensing, processing and communications in a single chip

Based on the above, a great variety of MEMS sensors are available on the electronic components market. Those solutions can be categorized as described in the following paragraphs.

5.3.1.2.1. Inertial sensors

This category includes sensors used in order to measure parameters such as acceleration as well as inclination (tilt), rotation, vibration, collision and gravity using inertial behavior of electromechanical microstructures on the chip.

The most common devices are *linear accelerometers* that can measure acceleration. They are available in 1, 2 or 3 axis versions and most relay on the measurement of the capacity of variable capacitors due to the impact of linear accelerations on their electrodes.
Accelerometers can be used also in order to measure other parameters, such as *inclination* of an object in respect to gravity, *shock* occurrence and *vibration*.
For the measurements of rotations, *MEMS gyroscopes* are being used, which also can come in 1, 2 or 3 axis versions. The most available gyros use the tuning fork configuration in which the angular velocity is measured using the capacitance change between two oscillating masses due to the Coriolis force as a result of rotation [2].

5.3.1.2.2. Environmental sensors

This category includes sensors which can be used in order to measure environmental parameters such as temperature, barometric pressure and humidity.

The most common approach for MEMS temperature sensing by absorbing the IR energy emitted from objects, whereas the measurement of barometric pressure relies on the use of extra pressure sensitive membranes and additional circuits to convert pressure changes to electrical signals. Humidity follows the principle of sensing capacitance changes in capacitors due to moisture absorbing.

5.3.1.2.3. Inertial Measurement Units

Inertial Measurements Unis (IMUs), also referenced as 6- or 9-axis accelerometers, are combinations of MEMS sensors in single packages which can be used for measuring the movement of objects in the 3D space.

An IMU combines information from a 3-axis accelerometer and/or a 3-axis gyroscope and/or a 3-axis magnetometer in order to provide information of movement of the sensor. The combined sensor output includes rotations, linear accelerations, headings, quartenions, and gravity.

5.3.1.2.4. Chemical Sensors

Using MEMS technology a variety of sensors for chemical procedures has been developed. The major measurements include: chemicals detection, leakage detection, pH measurement, etc.

The implementation of the above sensing procedures is being done with the use of Chemical Sensors arrays which can simultaneous detect and measure the amount of specific substances (more in [3]).

5.3.1.2.5. Biometrical Sensors

MEMS based biometrical sensors, or Bio-MEMS, have been developed in the last years in order to support applications in the areas of genomics, proteomics, point-of-care diagnostics, tissue engineering, and implantable microdevices [4].

5.3.1.2.6. MEMS in communication systems

Following the requirement for small board designs and low power consumption the communication components industry has also benefited from the use of MEMS technology. The major objective till now has been the development of RF MEMS based circuits with high-Q components in order to replace board consuming circuits such as RF filters and resonators.

5.3.1.3. Potential use of MEMS sensors in INTERACT

As described in section 3 one of the major objectives of the INTARCT project is the tracking of manual shop-floor operations. This can be achieved by tracking the objects that participate in a manual operation and which are humans, parts and tools. The various trackable parameters for those categories are presented in par 4.2.

Due to the nature of those objects, which (i) have limited space available for integrating sensors on them in order to track their operation, and (ii) are moving inside the shop-floor, so the dimensions and weight of the desired sensors to be integrated plays significant role, the requirements for the use of sensors in order to track the desired parameters are in accordance to those that are addressed by MEMS based technologies.

5.3.1.3.1. Trackable parameters related to the human body

As described in par 4.2.2 various parameters of the human body during a can be tracked during a manual operation.

For tracking the Kinematic model - skeleton data / posture the use of IMUs is a potential solution which can give effective results in the case of the INTERACT system. The movement of various joints of the human body can be tracked with IMUs which will produce data relative to this movement. The same could also stand for the worker cognitive state since the head position for example can be also tracked with the use of an IMU.

Inertial sensors and/or IMUs can be used in order to track forces and other parameters such as vibrations, shocks, and collisions/contacts with other objects that affect the worker during the use of tools or parts.

In the case of the worker’s vital data MEMS bases biomedical sensors can be used in order capture his condition during operations.

The use of the above sensors can be also combined with other data inputs such as RFID and optical sensing in order to construct the best available human body model.

5.3.1.3.2. Trackable parameters related to tools

Tracking parameters regarding to the use of tools during an operation depends on the structure of the tool and can be done using two different approaches:

1. For parameters that are accessible through an available tool interface (serial, TCP/IP, etc.) the data collection can be achieved by implementing the specific interfaces with the tools using the necessary HW and SW infrastructure.
2. For parameters that are not available through an existing interface the adaptation of sensors is necessary. As in the case with the human body the use of MEMS based sensors can be used in order to collect the desired parameter data. Accelerations and multi-dimensional forces applied can be measured using inertial sensors or MEMS pressure sensors.

5.3.1.3.3. Trackable parameters related to parts and objects

For tracking arbitrary properties of parts or other objects such as multi-dimensional forces or contacts/collisions can be measured using MEMS inertial sensors (accelerometers, shock sensors).

5.3.2. RF sensor networks

5.3.2.1. Sensor networks

After decisions an what sensors are required in order to best tracking the desired parameters in every application the most important design issue is the deployment method of those sensors. This procedure can be treated as a hierarchy of tasks [5], as follows:

- detecting the relevant quantities
- tracking and collecting the data
- assessing and evaluating the information
- formulating meaningful user displays, and
- performing decision-making and alarm functions

The most common approach to this requirement is the design and deployment of a Sensor Network, which includes a data acquisition network which includes the sensors and all necessary components (hardware and software) in order to collect the sensor data, and a data distribution network which includes all components which are used in order to process and deliver the processed sensor data to the outside world. In the general case of deploying a Sensor Network the main key design parameters as presented in [6] include:

- **Sensor acquisition and data rates**, which are critical for the design because the desired bandwidth of the sensor network and the technologies required to implement it relies on them
- **Cost of network deployment**, which is critical for the feasibility of deployment and also for the choice of technologies that will be used for it.
- **Network size and density**, which can lead to decisions on technology use, network topology and operation and cost per node limitations.

Apart from the above parameters and according to the application requirements, supplementary design parameters can be considered at deployment phase, such as:

- **Data communication latency**, which is critical for networks with real time performance requirements
- **Shop floor environment and spatial structure**, which include the various conditions (environmental, signal propagation, network coverage area, etc) and which play significant role in the network design
- **Data security**, for applications where sensor data has operational value that demands specific security treatment such as encryption.
- **Sensor System and network reliability**, according to the application criticality and need for sensor data availability

5.3.2.2. Wireless Sensor Networks

In the last years the most common approach is the design of Wireless Sensor Networks (WSN) in which the sensor nodes communicate with the Base Station using wireless technologies. The biggest motivation for the use of wireless media for data transfer and communications is that there is no requirement for deploying a structured wire infrastructure. Especially in sensor applications where human wearable sensor systems or sensors attached to moving assets (tools, parts, vehicles, etc) are required the system deployment has to follow design parameters, such as:

- **Monitored subject movement freedom**, which is very critical for human and small asset sensors application where movement of the subject (human, tool, etc.) must not be affected by the sensor system installation
- **Sensor system size and weight**, for applications where space constrains affect the available size and weight for the sensor system
- **Sensor system energy efficiency**, which is a vital parameter which can be treated as a resultant of various parameters referenced above (data rates, sensor acquisition, data security, sensor system size, area coverage etc).
- **Wireless network technology**, which is used by the distributed components in order to communicate. Various technologies have been developed in order to address issues such as low power consumption and small board size. RF technologies combined with communication protocols and routing methods provide integrated communication solutions for WSN implementations. The most widely used are 802.11 WLAN, ZigBee, ZigBee PRO, Bluetooth, and TCP or UDP over GPRS.

5.3.2.3. Wireless Sensor Network Architectures

The architecture of Wireless Sensor Networks has been a significant research subject since the 1990s. It must be noted that the majority of the techniques described have been adopted from computer network theory. The different approaches proposed are based on:

- **Network topology**, which describes the way the nodes are connected to each other in order to create the network. Various categories, like fully connected networks, mesh networks, star networks, ring networks, bus networks, etc. are used usually in references.
- **Data communication protocols**, which describe the techniques and architectures that are being used by the network nodes in order to communicate with the rest of the network, like switching protocols, multiple access protocols, the OSI architecture, etc.
- **Data routing methods**, which describes the algorithms in order to device the path that will be followed by a packet of information from a source node to a destination node in a network where multiple paths exist between those two nodes. Several methods for routing like fixed (i.e. pre-planned), adaptive, centralized, distributed, broadcast, use of hierarchical models etc. have been described in bibliography.
- **Power management,** which with the advent of geographically distributed and wearable sensors has been a subject of major focus. Various parameters have been considered for the optimization of power usage and expansion of life cycle for wireless sensor nodes. The use of low power components such as MCUs , MEMS sensors, RF circuits together with the implementation of techniques for low power consumption during operation of the network regarding to sensor acquisition, data processing in the nodes, RF communications, etc, are all aspects that are being examined in order to expand the network autonomy.

5.3.2.4. Sensor Networks - the “Internet of Things” approach

The significant developments in data communications during the past decades followed by significant achievements in the Wireless Sensor Network field has driven the development of systems that allow the use of the Internet protocol in Wireless Sensor Networks, and moving away from traditional sensor network implementations which led to costly, market specific, difficult to expand solutions. This fact led to the introduction of the term “Internet of Things” in 2009. Since then various definitions have been used in order to describe it ([7], [8], [9]). Despite the different approaches of the term one aspect is common to all: the handling of sensors as unique entities with unique identification and with common access methods.

In order to achieve such an approach [9] the *WSN communication stack* must be structured in a way so that it can lead to unique addressing scheme of the sensor nodes. This scheme is combined with mechanism referenced as *middleware* which allows the access to heterogeneous sensor resources in a unified way. This layered structure can be used as a basis for the implementation of data aggregation and sensing layers on top of it.

Besides principles that have been proposed for the development of the next generation Internet [10], a set trends that apply to the IoT approach of wireless sensor networks. Those trends, analytically presented in [10], are:

- Scalability of communications and management
- Cooperative Communications and networking
- Energy efficiency
- Convergence of the Sensor Network and Data Analytics

The "Internet of Things" approach has been proposed for several real world applications, where its implementation can lead to significant benefits. Those applications include:

- Transportation and logistics
- Smart environments
- Personal and social networking
- Healthcare

The case of tracking manual operation on the shop-floor environment in the INTERACT system can be considered as a smart environment case.

5.3.3. Sensor networks in INTERACT

As described in par 5.3.2.1 all sensors that will be used for tracking the manual operations on the shop-floor must be integrated as sensor network that will deliver the data to the other modules of the INTERACT system for further processing. Those sensors do not include only the categories described in 4.2, but also other systems such as RFIDs, RTLS, optical sensing, tools, etc.

Based on sensors and network related information presented in the previous paragraphs, a possible structure of this sensor network can be implemented following principles presented below:

- Sensors, where possible, will be MEMS based and organized in Wireless Sensor nodes that can be integrated on the tracked objects. These nodes will be designed according to the small board dimensions and power consumption requirements. This will be achieved by use of smallest possible lowest power components, minimal HW design for the sensor nodes and appropriate power preserving SW techniques.
- All network nodes, wireless and wired, will be organized as a sensor network. The wireless network technology, network topology, communication protocols and routing methods are subject of development in order to achieve the optimum solution in respect to all parameters referenced in 5.3.2.
- Due to the conditions that exist in the shop-floor the sensor network the deployment of the network should be very easy. The first implementation, network expansions, changes in network size and density, changes in network setup, and other procedures will be easy to imple-

ment. Also the sensor network operation will not affect shop-floor infrastructure including communication networks, tools, and other components.

- The sensor network will have a generic structure in order to be able to include heterogeneous sensor systems. This can be accomplished by the adaptation of the methodologies described in the IoT approach (par. 5.3.2.4). This design method could also be probably used for the creation of a basis for spatial and temporal resolution and synchronization tasks as described in chapter 4.
- In order to be able to adjust the system setup according to the application use, all sensors must be easy to reconfigure and administer. This is also a principle that can be implemented using the IoT approach since the creation of a platform that will handle sensor heterogeneity will give the ability to create common and easy IoT based parameterization procedures.

6. REQUIREMENTS ON TRACKING AND POTENTIAL SOLUTIONS FOR THE INDUSTRIAL USECASES

6.1. Common requirements for INTERACT applications

6.1.1. Generic use case and system description

The chart in the following section 6.1.2 shows the two main operation modes of the INTERACT system, and points out where tracking and sensing of movements is necessary.

The first step in an INTERACT system consists of capturing human motion samples of basic motions (e.g. walking, grabbing) for the creation of a statistical Motion Graph++ [21] model, which describes the movements in a statistical way as Gaussian mixture models. These models are stored in a database.

From this database, an initial digital human model scenario can be synthesized, given a work task sequence from the application domain formulated in machine-readable controlled natural language. This model can serve as initial input for ergonomic analysis as well as for process evaluation of the work tasks.

In the second step, participants in an ergonomics or process optimization workshop use the initial version from step 1 as a starting point to optimize certain key parameters of the movements to gain improved processes or ergonomics. For this, the work tasks are carried out by a workshop participant, whose movements are captured and temporally mapped to the work tasks through cutting the movement trajectories. Predefined values of the movements are then filtered, and used as an input to further parameterize and constrain the synthesized digital human model scenario, which leads to an improved synthesized output. This can in turn again be an input for process evaluation and ergonomic analysis, and also can be carried out repeatedly to move towards an optimized solution iteratively.

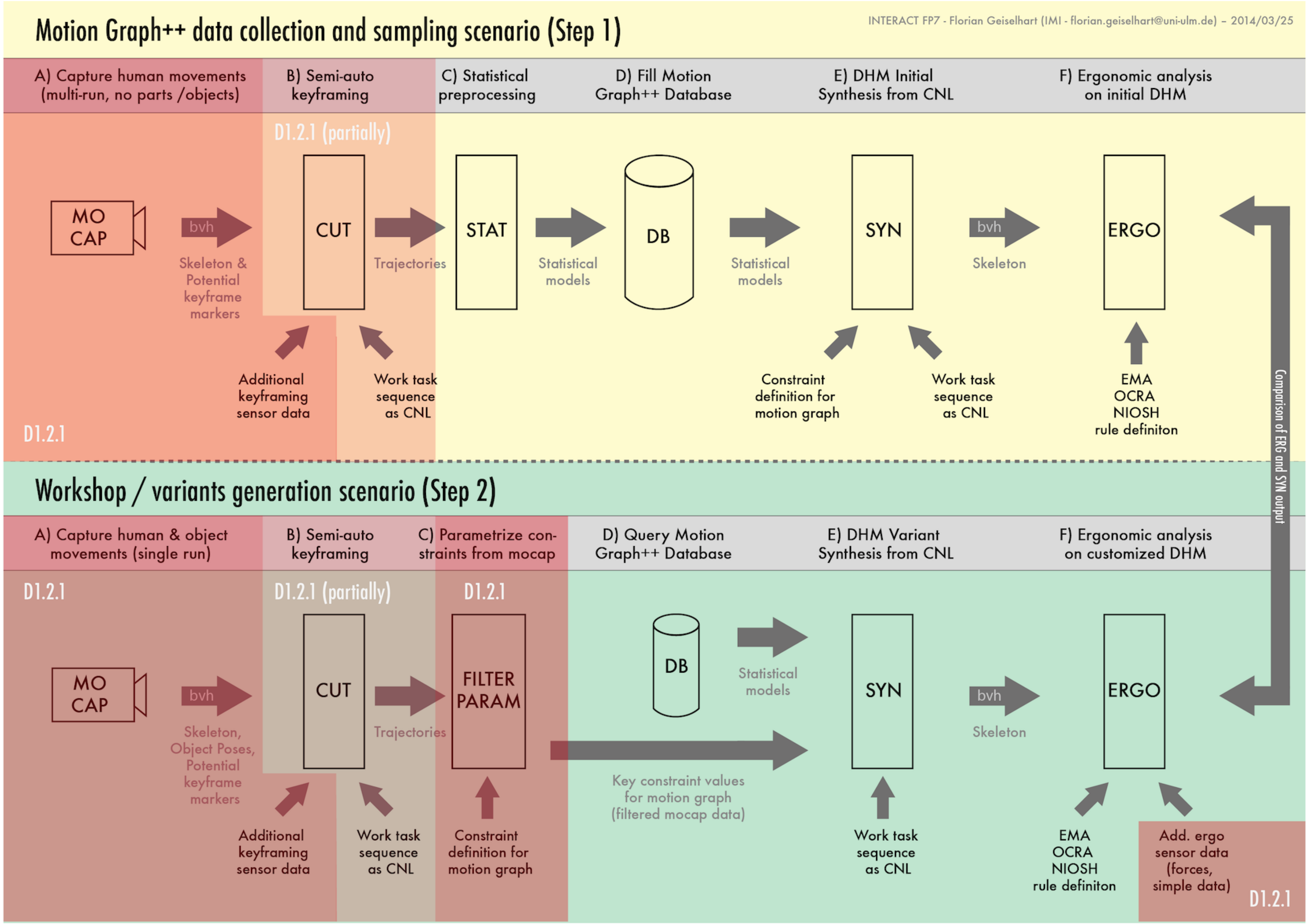
For additional improved ergonomic analysis, there is also a need to integrate real-world measured ergonomic sensor data into the synthesized model, for example to evaluate forces which affect the workers body.

The main application scenario for both pilot cases and the system as a whole is the usage of the system within planning workshops, which corresponds to a semi-controllable environment. This means that certain aspects of a regular industrial environment do not play a role, but nevertheless the system needs to provide a certain robustness against different interference factors like occlusion, changing light conditions or electromagnetic stray radiation.

Coming from this system description and application scenario, there is a need for the tracking of human skeleton movements throughout the whole system, as well as an additional selective need for additional sensors to either augment or improve the skeleton sensing under difficult circumstances, or to track additional data points like force measurements, object or part locations or the state of tools as an input to key frame identification algorithms and ergonomic analysis.

The following sections within this chapter describe the detailed requirements on these different tracking subsystems as well the overall non-functional needs.

6.1.2. Overview of INTERACT backend system architecture



6.1.3. Requirements on sensing and tracking subsystems

6.1.3.1. Functional requirements regarding human motion tracking

As already pointed out, human motion tracking plays a central role in the INTERACT system. Human motions are the input both to the motion graph data collection and the workshop variant generation. This not only includes a full body skeleton tracking, but also fine-grained tracking of finger movement, for example to differentiate between grasping types.

The following table lists the requirements related to tracking humans within the INTERACT system.

Requirement	Description
Markerless tracking	Tracking of human movements should not include additional visual markers which need to be attached to the body at least in the workshop scenario (step 2). However, for certain joints, it is acceptable to have additional wireless sensing units (e.g. IMU units) attached above clothing or to be worn e.g. as a glove to enable or improve certain tracking aspects. In step 1, it may be acceptable to use a marker based motion capturing system due to the higher demands that need to be met in terms of precision and noise for motion sample collection.
Skeletal model detail level	Skeletal model details need to be captured according to chosen reference skeleton of Rocketbox' character model (see appendix for model details): • 25 joints for full body skeleton • 20 additional joints for hand skeleton model
Operator movement detail level	The following types of movements need to be measured or inferred from the skeletal tracking data: • Calculation of upper body bending • Calculation of upper arm orientation • Calculation of Hand Position • Examination of back support existence • Height of hip evaluation • Calculation of upper body orientation • Measurement of time spent in a specific position • Measurement of frequency of a specific position • Trunk rotation calculation • Trunk lateral bending • Measurement of far reach postures
Contact and object relationships	• The movements have to be tracked when the human is carrying objects, which also may occlude parts of the body. • The contact detection of the human limbs with arbitrary objects has to be supported through the body tracking.

6.1.3.2. Functional Requirements regarding part and object tracking

In step 1 of an INTERACT system setup, normally no parts or additional objects need to be tracked. However, in step 2 additional object tracking may prove useful for keyframing, synthesis and also ergonomic evaluation.

In the general INTERACT scenario, information about present tools, objects or parts is a-priori knowledge. This information is contained in the work task sequence data, which means that no direct object identification is necessary, but rather a tracking of known objects or markers applied to them.

The following table lists the requirements related to part and object tracking.

Requirement	Description
Object tracking	Objects have to be tracked in 3-dimensional space including their full orientation.
Pulling / pushing support	Additional sensors may be used to support detection of object pushing or pulling.
Contact and body relationships	• The objects movements have to be tracked when the human is carrying objects, which also may occlude parts of the objects. • The contact detection of the human limbs or other objects with any arbitrary objects has to be supported through the object tracking.

6.1.3.3. Functional Requirements regarding additional sensors for key frame identification support

To allow for better recognition of possible keyframe markers for movement type changes, additional sensors may be used to provide useful input to the cutting subsystem. A list of movement types that need to be supported can be found in deliverable 1.1.1.

Requirements for this specific sensor types can be found in the following table.

Requirement	Description
Keyframing sensor support	The system should support and provide an interface to different types of low-dimensional sensors to enrich the tracked data with markers for potential keyframes. Such sensors may be: • Switches (Boolean 0/1) • Distance Sensors / Proximity Sensors • Arbitrary sensors that output a numeric or boolean value
Keyframing pipeline support	The system should be able to deliver semantically meaningful keyframing metadata to the consumers of the tracking data.

6.1.3.4. Functional Requirements regarding additional sensors for ergonomic analysis

For ergonomic analysis in step 2, real-world measurements of certain parameters need to be taken in order to support ergonomic analysis.

Requirement	Description
Force measurements support	The system should support measuring forces and their various parameters at selected points to allow for better ergonomic evaluation: • Force vector/ orientation • Applied Force measurement • Time of applied force measurement • Frequency of applied force measurement
Body forces support	The system should be able to measure forces on a whole-body level as well as on finger / fingertip level.
Force measurement range	Force measurements should be possible in a range from 0,1 N to 10 N.

6.1.3.5. Functional Requirements regarding sensor data filtering

For step 2, only certain data points are being used to parametrize the constraints which are used to synthesize the human motions from the database. For this, the tracking system has to provide a spatial and temporal filtering possibility. See the following table for the requirements of this component

Requirement	Description
Filtering rules input	The filtering rules are provided by the application framework in the form of constraint definitions, which need to be satisfied by the filter. The filtering component should accept this input and configure its filters accordingly.
Temporal filtering	The filtering component should be able to filter tracking data temporally, e.g. through dropping captured frames with upstream smoothing.
Spatial filtering	The filtering component should be able to filter tracking data spatially, e.g. by just forwarding the data for one or some joints of the human body.

6.1.3.6. Non-functional requirements

In terms of system performance and quality, several important factors are relevant for the tracking subsystems of the INTERACT system. See the following table for a list of non-functional requirements the system has to implement.

Requirement	**Description**
Occlusion robustness	Occlusion in vision sensors can be caused by: • Worker • Parts, objects or tools • Surrounding infrastructure Tracking must not be lost during phases of occlusion. The occlusion situation can be assessed in advance, sensors can be placed accordingly.
Tracking robustness	The tracking devices need to be robust against normal levels of disturbing parameters, namely: • Changing light conditions • Electromagnetic stray radiation • Magnetic field anomalies, e.g. through large metallic parts or magnets within motors or tools.
Tracking levels	Tracking of workers skeleton joints, parts and objects has to take place with 6 degrees of freedom to capture the full scale of possible movements.
Tracking precision	A sufficient tracking precision is crucial for motion synthesis, especially in step 1. This will usually require a mm-level precision at least for certain operations. In step 2, more coarse tracking might be sufficient to recognize the motion keyframes, but for collision testing with parts or other objects, additional measure may need to be taken to enable for precise collision sensing.
Tracking noise and stability	• Sufficient for motion synthesis in step 1 (low noise necessary) • Sufficient for cutting and keyframing in step 2 (medium noise level acceptable)
Skeleton nomenclature	The skeleton joint naming and identification should be flexible enough to follow common conventions, especially respecting related projects like the currently running project ARVIDA, which is funded by BMBF and whose goal is the definiton of a virtual technology reference architecture. The currently envisioned model for this is the H-Anim Standard provided by the Web3D consortium [22].

Tracking speed / frames per sec	To our current knowledge, human body tracking with a rate of 50 Hz is sufficient for motion synthesis purposes, while synthesis quality gradually degrades going down to 25 Hz, where data becomes unusable for synthesis. Slighlty lower rates than 50 Hz may be acceptable in certain situations (especially in step 2), though. Arbitrary sensor data (e.g. IMU data) should at least reach a refresh rate of 50 Hz to accomodate for faster movements e.g. of tools. These values are derived from the requirements of the motion synthesis system, as well as from knowledge about the speed of typical human movements (see Ch. 3 for details). These values also provide a guideline for the communication and storage technologies to be considered, as the data from all sensors needs to be transported and stored within this time frame.
Setup and calibration flexibility	• Short setup and calibration times (< 1h) are necessary because of frequent relocations of the setup. • Different kinds of sensor setups should be supported in order to accommodate the sensor system to different work task situations. • It should be easy to add or remove additional sensors to quickly accomodate to scenario-specific challenges.
Tracking volume	The system should be able to track a space which is large enough to accomodate the workplace setup. With the current pilot cases, the approx. current maximum dimension of the capture volume is 15m by 6m by 3m (W x D x H).

6.2. Requirements specific to the assembly line pilot case

6.2.1. Summary of pilot case description

Tracking of various parameters as described in section 4.1 (human body, tools, parts, etc.) which are part of assembly line operations can give significant information related to the operations: process time, process type, tasks order, ergonomics, design for assembly, station layout (incl. vehicles, bins, pallets, …), materials deposition. Each task could be evaluated by workers, planners and managers responsible for topics such as ergonomics, logistics, scheduling, design etc.

Tasks performed by worker at assembly line should be tracked. Tracking tools, parts, units, bodies etc. parameters should be sufficient to deduct what kind of movement was performed and what were the parameters of the task (e.g. force applied by worker). Ergonomic assessment should be also possible. This means that various technologies should be applied together.

Activities that can be used as use case for the INTERACT approach are presented in the following table. It should be noted that the term "predefined" is being used for assets whose physical parameters (dimensions, shape, weight, material, position, etc.), that are necessary in order to fully specify the process, are well defined prior to the process.

1. ***Use Case Summary 1: Assembly task execution***

Description	Assembly of part into subunit or subunit into unit
Steps for one worker (Workers can perform tasks in parallel.)	1. Worker goes to predefined storage location. 2. Worker grabs predefined asset(s) (part/subunit/unit/tool/screw). 3. Worker moves with grabbed asset(s) to workstation and puts it down temporarily. 4. Worker may grab more than one item (e.g. screwdriver in one hand, screw in the other). 5. Worker may repeat steps 1-3 for many assets that he needs. 6. If no repetition of steps 1-3, worker goes with asset(s) to workstation and starts assembly task. 7. Storage location could be sub-location of a predefined workstation. 8. Worker executes assembly task. 9. Captured data on assets and bodies (steps 1-8) are sent to app which deduct types and order of performed tasks.
Trackable parameters	1. Body movement and rotation throughout the whole process. 2. Movement and rotation of shoulders, hands, fingers. 3. Assets movement and rotation throughout the whole process. 4. Contact of assets and bodies. 5. Internal states of used tools (e.g. state, forces). 6. Collisions of assets and/or bodies (eg. screwdriver hits screw and screwing can be started).

6.2.2. Specific requirements for assembly pilot case

In the further requirements elicitation process involving the industrial pilot case partner, specific requirements were defined. Together with the information from the use case description (see previous chapter), additional requirements for this pilot case were found.

6.2.2.1. Tool / External API support

Within the assembly pilot case, different tools are being used, which maintain intrinsic states, as already pointed out in chapter 4. Those states may either be accessible through external APIs or communication channels, or need to be sensed through additional sensing points (e.g. switches for on/off states).

For this, the INTERACT system has to provide a possibility to insert data from external APIs, as well as support for additional simple data sensors, which can be assigned to a predefined tool state (similar to the keyframing support sensors).

Furthermore, this ability of the system can also be used in combination with the generation of keyframe markers already mentioned above.

6.2.2.2. Support for ergonomic evaluation methods

As ergonomic guidelines differ from country to country and also from industry to industry, different ergonomic scales and evaluation methods need to be applied. The INTERACT system should be able to support multiple different ergonomic assessment methods through its sensing flexibility and detailed tracking level.

For the assembly operation use case, a possible ergonomic evaluation metric is EAWS.

6.2.3. Further information on the pilot case

For detailed information please refer to deliverable 1.4.1.

6.3. Requirements specific to the warehouse operator pilot case

6.3.1. Summary of pilot case description

Tracking of various parameters as described in section 4.1 (human body, tools, parts, etc.) which are part of warehouse operations can give significant information related to:

1. The ergonomic analysis and evaluation of the operations
2. The safety analysis and evaluation of the operations
3. The spatial analysis and evaluation of the operations
4. The temporal analysis and evaluation of the operations

The above information can be used, by feeding appropriate tools, in order to optimise the process design and efficiency by means of ergonomics, process time, process space, part and tool structure and design.

Warehouse operator processes that can be used as use cases for the INTERACT approach are presented in the following tables. It should be noted that the term "predefined" is being used for assets whose physical parameters (dimensions, shape, weight, material, position, etc.), that are necessary in order to fully specify the process, are well defined prior to the process.

Use Case Summary 1: Part Fetching and loading on trolley

Description	Warehouse operator fetches part from a predefined warehouse shelf and stores it in a predefined location on trolley
Steps	1. The operator goes to the predefined shelf 2. The operator grabs the part 3. The operator carries the part to the predefined trolley 4. The operator loads the part in the predefined location on the trolley
Trackable parameters	1. Operator movement throughout the whole process 2. Operator body physical strains as a result grabbing, carrying and loading the part 3. Part movement throughout the whole process 4. Trolley location in the warehouse area

2. ***Use Case Summary 2: Performing multiple part fetching-loading cycle ("spaghetti-chart")***

Description	Warehouse operator fetches multiple predefined parts from predefined warehouse shelves and loads them on predefined locations on the trolley
Steps	1. The operator goes to the predefined shelf 2. The operator grabs the part 3. The operator carries the part to the predefined trolley 4. The operator loads the part in the predefined location on the trolley 5. The operator repeats the process steps 1-4 for all predefined parts
Trackable parameters	1. Operator movement throughout the whole process ("spaghetti

	chart") 2. Operator body physical strains as a result grabbing, carrying and loading the parts 3. Parts movement throughout the whole process 4. Trolley location in the warehouse area

3. Use Case Summary 3: Trolley load tracking

Description	Warehouse operator fetches part from warehouse shelf and stores it on trolley
Steps	1. The operator goes to the predefined shelf 2. The operator grabs the part 3. The operator carries the part to the predefined trolley 4. The operator loads the part in the predefined location on the trol-ley 5. The operator repeats the process steps 1-4 for all predefined parts
Trackable parameters	1. Trolley location in the warehouse area 2. Part location in trolley

6.3.2. Specific requirements for pilot case

6.3.2.1. Support for ergonomic evaluation methods

As ergonomic guidelines differ from country to country and also from industry to industry, different ergonomic scales and evaluation methods need to be applied. The INTERACT system should be able to support multiple different ergonomic assessment methods through its sensing flexibility and detailed tracking level.

For the warehouse operator use case, possible ergonomic evaluation methods include OCRA and NIOSH.

6.3.3. Further information on the pilot case

For detailed information please refer to deliverable 1.4.1.

6.4. Potential technical solutions for the requirements

In order to fulfil the requirements presented in the previous section several potential solution will have to be evaluated during the design phase of INTERACT. The following chapter describes a rough selection of technologies from different fields that will be considered.

It must be noted that the all potential solutions will be evaluated also in respect to the various parameters presented in chapter 5 as general guidelines for the sensor systems that will be implemented during the development phase of the INTERACT project.

6.4.1. Optical sensor based solutions

Optical sensors were evaluated as the basic technology especially for human motion tracking. Within the INTERACT system, the use of state-of-the art depth cameras and skeleton recognition algorithms will likely be used to track human position, motions and skeletal body configuration across the sensing area. Through multi-camera setups and data fusion, precision and coverage of these devices can be improved to conform to the INTERACT requirements.

For the large amounts of data generated by the depth cameras, a distributed approach will be necessary, which offloads computer vision processing to separate units.

In terms of technology, the INTERACT should support different types of depth cameras with Structured Light technology as well as Time-of-Flight technology in a pluggable way, to be able to exchange one or more cameras with more mature technology in the course of time.

Additional 2D camera systems (which may already be integrated in the depth camera units) will be considered for object tracking through the use of markers or object recognition.

For key frame identification applications, additional optical sensors like distance sensors, light barriers or light sensors may be used as simple data points.

There is also a potential need for an interface to common MoCap systems for the motion sample collection in step 1, which could be achieved through the support of standard protocols like VRPN.

6.4.2. MEMS based solutions

The use of MEMS sensors, but also of other sensor technologies, can contribute in meeting the use case requirements. For the various requirement categories potential solutions are given below.

1. For fulfilling requirements arising for *operator movement* MEMS based Inertial Measurement Units, also referenced as 6- or 9-axis accelerometers, can be used in various points of the operator body in order to enrich the skeleton data from the optical sensing system. This enrichment can be applied at skeleton areas like the hands that cannot be tracked sufficient with optical systems, or at points for which the optical systems provide unreliable data. An IMU could be also combined with a barometric pressure MEMS sensor for giving height measurements where applicable. The current state in MEMS technology with very small chip sizes (3x3 mm), high accuracy sensors, digital interfaces, and power consumptions of less than 5mA can lead to small, power efficient, accurate and cost effective sensor systems.

   Special attention should be paid for data collections from workers' hands during assembly operations. For this purpose data gloves which integrate large number of sensors for measuring finger flexion, finger and wrist abduction, the palm-arch, and wrist flexion can give valuable information during assembly processes, especially where small parts are being handled.

2. IMUs can be also used for capturing *part movement and location/orientation* data. For a part of known 3D model the application of a 9-axis IMU, which measures linear accelerations, angular rates and orientation, in a specific location (e.g center of gravity) can provide data in order to synthesize the movement of this part during an operation. Of course the limitation of

the IMU size in respect to the part size should be taken in account, since this method is only applicable for parts which have a size and structure that can support an IMU installation.

3. The case of measuring *physical strains* applied to the human body during operations, which usually are lifting and carrying of parts, fixing of parts, and operation of tools can have different approaches, according to the type of process. The first approach is to directly measure strains on specific points of the body with sensors. Small size force sensors attached to those specific points can measure the forces applied during an operation, or force measuring glove for measuring the dynamic pressures and forces hands and fingers apply while grasping, gripping, holding, moving, lifting parts. Bending and stretching sensors will be also evaluated during the INTERACT project for their capability of providing information related to body strains related to upper body bending or far reaching postures for example.

   A second approach, applicable to lifting and carrying parts operations, is to calculate the strains by combining information of the part physical parameters (mass, center of gravity, 3D model), its position (location and orientation), and the operator position. This way the strain on the operator body can be calculated in upper platform levels.

   Regarding strains coming from the use of tools which do not operational information through digital interface, like forces or torque, the approach depends of the type of tool used, since its size and the effect of adapting a sensor on it must be considered. The use of multi axis force/torque sensors will be evaluated during the development phase of INTERACT.

6.4.3. RFID sensor based solutions

Due to the accuracy restrictions coming along with RFID-based solutions, RFID was evaluated as mostly unsuitable for a major part of the INTERACT sensing requirements.

However RFID might be used in the context of simple data point sensing for keyframing support, but more as a non-mechanical replacement for position switches or similar tasks. This however makes no use of RFID-specific features besides its non-mechanical, contactless operation.

Another potential use might be the UWB-based localization of trolleys in the warehouse scenario, for example to initially reduce the search space for a more precise vision-based localization in a later recognition and fusion step.

6.4.4. Data fusion solutions

To deal with the multi-sensor-architecture, as well as occlusion and keyframing requirements, it will likely be necessary to implement different levels of data fusion in the INTERACT system. Besides that, it will be necessary to implement a strict time synchronization and timestamping throughout the tracking system for different sensor systems, as this is not only a precondition for a working fusion component, but also for the further processing in keyframing, synthesis, and ergonomics analysis.

On a low level, a possible fusion approach that will be evaluated is the fusion of multiple depth camera systems for improved occlusion robustness and precision. Different depth cameras may watch the scene from different angles, which leads to multiple (partial or complete) instances of the same human skeleton, possible in their own coordinate space. In order to fuse these skeletons into one improved skeleton, after transformation into a common coordinate space (through calibration measures), quality measurements of single joints will need to be taken into account to compute a fusion weight quantifier which defines the influence of a single-sensor value onto the fusion result.

On an intermediate level, data from different sensors like already fused skeletons and IMUs needs to be combined through fusion methods. This is for example necessary to accomplish the IMU-based improvement of vision-based skeleton tracking data. For this, different types of data need to be fused, so it is crucial to preprocess the data to a fusable common format first. In the case of IMU + skeleton/vision fusion, this means that the IMU needs to be assigned a semantically meaningful position first within the full body skeleton, as well as there is a need for a common coordinate system. After this, the same weighted methods for fusion may be applied as with the low level fusion.

On a high level, the system also needs to provide a kind of fusion process for the usage with the keyframe markers. Here, different data sources need to be combined to infer potential keyframes / cutting points for the further processing. This not only involves pure mathematical combination of different numerical values, but might also require machine-learning or AI-based components which need to be evaluated in the further systems design.

6.4.5. Key frame identification support solutions

As the tracking system should be able to provide markers for potential motion key frames for motion cutting as a metadata add-on, it is necessary to preprocess and analyze incoming sensor data for this functionality.

Possible sources of keyframing information are the key frame related sensors that were already mentioned in the requirements. Normally, this information only needs to be preprocessed, to be assigned with semantically meaningful information and to be forwarded to the cutting subsystem.

Another possible source of keyframing information is inference for example from body or object tracking values. In a very simple case, this means that given threshold rules in position or rotation values lead to the addition of a marker for a potential key frame (e.g. the smoothed X-Z coordinates in the capture space which determine a possible transition from standing to walking).

Depending on the motion graph primitives definition, more complex scenarios might also be possible, where a combination of multiple values leads to the inference of a key frame marker. This could be for example a position value of a tool, which has contact with the human body and a screw, and whose state changes to ON, which could lead to the inference of a "start screwing" marker.

7. CONCLUSIONS

The sensor devices described in this document can be deployed in a wide variety of applications in different application domains. The three areas identified in Section 4.1 – Body tracking, tool tracking, and part and generic object tracking – are sufficient for a large number of applications, including the ones for INTERACT. Even though they are rather different at first sight – i.e. assembly vs. loading tasks – the generalized requirements in chapter 5 show that the same types of sensors are usable, with only very few special cases for the pilot cases which need to be considered.

The generic common goal which was identified, is process optimization and ergonomic evaluation.

It is a also major objective of INTERACT to reduce costs of production planning in the setup phase of a production line. Through the use of flexible sensor system with different kinds of sensors, it might be possible to cut down these costs to a fraction of the current costs for real prototypes and evaluation time.

The already mentioned flexibility and naturalness of interaction with such a sensor system will also improve the experience for users in workshops which have little time and only few experience with virtual technologies.

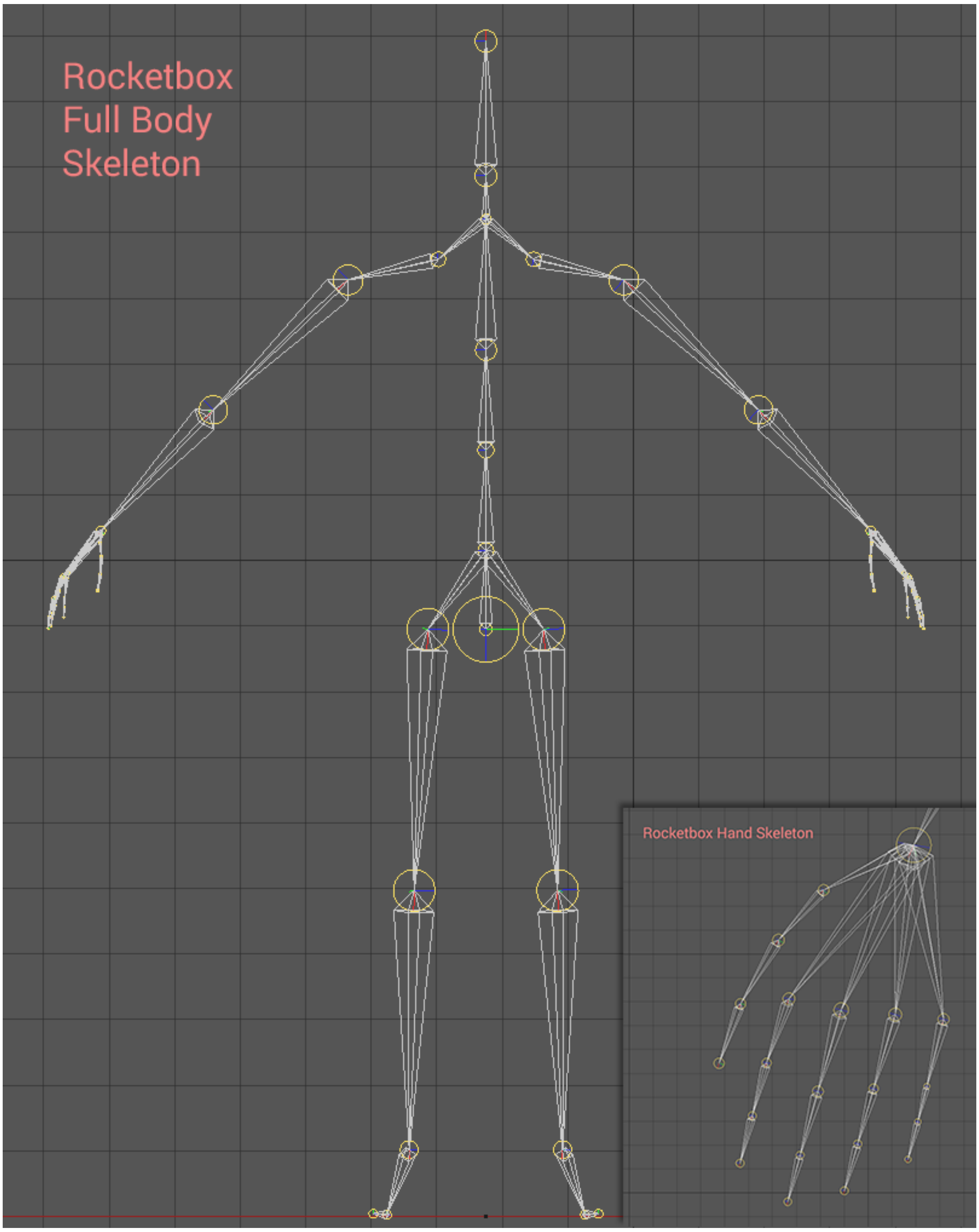
The aim of this document is not to specify exactly what will be measured in each use case. Obviously for the different use cases, as for other applications, the sensors applied will differ in number and kind. Tracking in INTERACT will be used for several purposes, including economic and ergonomic criteria. Thus, a wide range of applications may profit from realizing the stated tracking requirements, but also puts higher demands on flexibility and robustness of the prospective sensor systems.

All functional and non-functional requirements have been derived in chapter 6. The parameters to be tracked combined with the stated requirements will be the decision basis for technology to be applied, e.g. sensor technologies and other system components in WP3 and the other following work packages.

8. Appendix

8.1. INTERACT reference skeleton model - Rocketbox skeletal model

Note: The H-Anim standard ISO/IEC FCD 19774 specifies a systematic system for representing humanoids in a network-enabled 3D graphics and multimedia environment[5]. The Arvida pro,ject, in which INTERACT partners are involved, works on a specification of a standard for network-enabled simulation systems. Arvida builds on the H-Anim standard with respect the representation of digital humans and defines an interface for accessing digital humans in 3D simulations. With the resources of INTERACT it is not possible to implement such standards and therefore INTERACT plans to use an avatar representation which is in wide practical use. However, INTERACT will follow the Arvida specifications and will follow them where possible with reasonable effort.



[5] http://h-anim.org/Specifications/H-Anim200x/ISO_IEC_FCD_19774/

9. GLOSSARY

AoA	Angle of Arrival
AR	Augmented Reality
API	Application programming interface
CCD	Charge-coupled device
CMOS	Complementary Metal Oxide Semiconductor
CV	Computer vision
EAWS	Ergonomic Assessment Worksheet
EMG	Electromyography
EPC	Electronic Product Code
GPRS	General Packet Radio Service
GPS	Global Positioning System
HF	High Frequency
HW	Hardware
I2C	Inter-Integrated Circuit
ICP	Iterative closest point
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
LF	Low Frequency
LIDAR	Light Detection and Ranging
MEMS	Micro Electro Mechanical System
NIOSH	National Institute for Occupational Safety and Health
NTP	Network Time Protocol
OCRA	Occupational Repetitive Action
pRTLS	Passive RTLS
RANSAC	Random sample consensus
RF	Radio Frequency
RFID	Radio Frequency Identification
RSSI	Received Signal Strength Indicator
RTLS	Real Time Location System
SL	Structured Light
SPI	Serial Peripheral Interface
SW	Software
TDoA	Time Difference of Arrival
ToA	Time of Arrival
ToF	Time-of-Flight
UHF	Ultra High Frequency
UWB	Ultra Wide Band
VRPN	Virtual Reality Peripheral Network
WSN	Wireless Sensor Network

10. REFERENCES

[1] P. K. Atrey, M. A. Hossain, A. E. Saddik, and M. S. Kankanhalli, "Multimodal fusion for multimedia analysis: a survey," *Multimedia Systems*, vol. 16, no. 6, pp. 345–379, Nov. 2010.

[2] http://electroiq.com/blog/2010/11/introduction-to-mems-gyroscopes/

[3] Dr. L. Fuller, Microelectromechanical Systems (MEMs)- Chemical Sensors

[4] Steven S. Saliterman (2006). "Fundamentals of bio-MEMS and medical microdevices". Bellingham, Wash., USA: SPIE—The International Society for Optical Engineering. ISBN 0-8194-5977-1

[5] E. Gaura, L. Girod, J. Brusey, M. Allen, G. Challen, "Wireless Sensor Networks: Deployments and Design Frameworks", Springer, Sep 14, 2010

[6] D.J. Cook, S.K. Das, John Wiley, "Smart Environments: Technologies, Protocols, and Applications"

[7] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, 2010, pp. 2787–2805. [Online]. http://www.sciencedirect.com/science/article/pii/S1389128610001568

[8] D. Miorandi, S. Sicari, F. D. Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," Ad Hoc Networks, vol. 10, no. 7, 2012, pp. 1497–1516. [Online]. http://www.sciencedirect.com/science/article/pii/S1570870512000674

[9] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Future Generation Computer Systems, 2013, pp. 1645-1660 [Online]. http://www.sciencedirect.com/science/article/pii/S0167739X13000241

[10] A. Merentitis, F. Zeiger, M. Huber, N. Frangiadakis, K. Mathioudakis, K. Sasloglou, G. Mazarakis, V. Gazis, Z. Boufidis, "WSN Trends: Sensor Infrastructure Virtualization as a Driver Towards the Evolution of the Internet of Things", *UBICOMM 2013*.

[11] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2011, pp. 137–146.

[12] F. Faion, S. Friedberger, A. Zea, and U. D. Hanebeck, "Intelligent sensor-scheduling for multi-kinect-tracking," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 3993–3999.

[13] R. Sivalingam, A. Cherian, J. Fasching, N. Walczak, N. Bird, V. Morellas, B. Murphy, K. Cullen, K. Lim, G. Sapiro, and N. Papanikolopoulos, "A multi-sensor visual tracking system for behavior monitoring of at-risk children," in *2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1345–1350.

[14] Y. M. Kim, D. Chan, C. Theobalt, and S. Thrun, "Design and calibration of a multi-view TOF sensor fusion system," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008. CVPRW '08, 2008, pp. 1–7.

[15] M. Hansard, S. Lee, O. Choi, and R. P. Horaud, "Time of Flight Cameras: Principles, Methods, and Applications," Oct. 2012.

[16] T. B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," *Computer Vision and Image Understanding*, vol. 81, pp. 231–268, 2001.

[17] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian Video Magnification for Revealing Subtle Changes in the World," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 65:1–65:8, Jul. 2012.

[18] D. L. Hall and J. Llinas. „An introduction to multi-sensor data fusion". *IEEE*, 85(1):6--23, 1997.

[19] M. Liard. „UNDERSTANDING RTLS: What it is, How it's Used & What You Need to Know before Deploying". *RFID Journal*, 2012, available at: http://www.rfidjournal.com/whitepapers/download?41, accessed: 3.01.2014.

[20] M. Laaraiedh, Y. Lei, S. Avrillon, B. Uguen. „Comparison of Hybrid Localization Schemes using RSSI, TOA, and TDOA". *Wireless Conference 2011 - Sustainable Wireless Technologies (European Wireless), 11th European*, pp. 1-5, 27-29 April 2011

[21] J. Min and J. Chai, "Motion Graphs++: A Compact Generative Model for Semantic Motion Analysis and Synthesis," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 153:1–153:12, Nov. 2012.

[22] http://www.web3d.org/x3d/specifications/ISO-IEC-19774-HumanoidAnimation/