

# Cognitive Robotics in Industrial Environments

Stephan Puls, Jürgen Graf and Heinz Wörn  
*Karlsruhe Institute of Technology*  
*Germany*

## 1. Introduction

Industrial robotics is a challenging domain for cognitive systems, especially, when human intelligence meets solid machinery with many degrees of freedom like most of today's industrial robots. Hence, for guaranteeing safety for human workers, safety fences are installed to separate humans and robots. As consequence no time and space sharing interaction or cooperation can be found in industrial robotics.

Some progress has gained in the past to the extent that some modern working cells are equipped with laser scanners performing foreground detection. But with these systems one is not able to know what is going on in the scene and, therefore, could not contribute something meaningful for challenging tasks like safe human-robot cooperation. We are conducting research on reconstruction of human kinematics based on 3D imaging sensors. The resulting kinematical model is tracked and fused with knowledge about robot kinematics and surrounding objects into an environmental model. This allows for efficient risk estimation and subsequent risk minimization through adaption of robot motion. Based on these processing steps, recognition of and reasoning about actions and situations in a human centred production environment is performed. All components and modules are merged into a single framework for human-robot cooperation (MAROCO), in order to pave the way for interactive and cooperative scenarios.

In the following, the framework MAROCO and its components are described and it is shown how the presented approaches contribute to achieve the vision of close productive human-robot collaboration.

In Sec. 2, the state-of-the-art for the major research topics concerning this work is presented. This includes works about human-robot cooperation, human pose reconstruction and research about situation and activity recognition. Afterwards, a system overview is given, which highlights the system architecture of the developed framework. In Sec. 4, theoretical considerations and algorithmic approaches are detailed. The section about experimental evaluation follows, in which all implementations and developments are put on trial and demonstrate their effectiveness. Conclusions are drawn and hints for future work are given in Sec. 6.

## 2. State-of-the-art

The vision of humans achieving a common goal with robot co-workers offers manifold possibilities for robots application. In the past few years several research groups around the

globe contributed to this specific field of robotics research. At first, an introduction of the state-of-the-art for safe human-robot cooperation and interaction is given. Afterwards follows an overview about human pose reconstruction which builds an important basis for the here presented approaches. The elaboration takes into account the work of manufacturers, research institutes, and universities.

## 2.1 Human-robot cooperation

There are just a few camera based vision systems dealing with safe human-robot cooperation. One such system was introduced by the company Pilz in 2007. The system is based on three cameras which are mounted under the ceiling of a robot cell. Stereo vision tools are then applied to the image sequences. The main idea is dividing the robot cell in up to 50 static parts. The recognition capability of the system seems to be foreground detection. Dynamic scenes couldn't be processed efficiently. A meaningful real-time interpretation of the robot cell is not feasible, due to missing means to distinguish between humans and background objects.

The working group Robot Systems of the Fraunhofer Institute IPA from Stuttgart, Germany, incorporated a time-of-flight camera system into the robot cell (Winkler, 2008). This system deals with dynamic safety zones, which are established in a virtual environment model of the working cell. The system defines three types of regions:

- Regions which must provide measurements of the camera system to detect occlusions generated by the robot.
- Critical regions in which no person or objects may appear.
- Areas in which collision detection may not occur.

To reduce the risk for the human co-worker the maximal velocity of the robot can be limited.

A system dealing with direct human-robot cooperation is presented in (Thiemermann, 2005). The research foci are optimizing safety and ergonomics. The robot cell is build up with a SCARA-robot and a CCD-camera based vision system. This scientific work concentrates on hand tracking realised by colour segmentation techniques. Then the shortest distance between the estimated hand positions and the tool centre point of the robot is calculated. The risk recognition part is realized applying a classic fuzzy logic system. The parameters of the fuzzy logic system are trained by an artificial neural network. This work takes also velocities and accelerations into account to finally control the maximal speed of the robot.

Application of CCD-cameras for realisation of such a system seems to be plausible. But there are several open questions regarding stability analysis, robustness against changing illumination conditions, etc. Mere concentration on the co-workers hands can also be restricting.

Another approach for safe human-robot cooperation was published in (Kulic, 2005). The setup of the robot cell is a PUMA robot (type 560). The sensor system is, compared to other approaches, more complex, since several hardware kits like stereo colour vision system, an electrocardiograph or an electromyography are applied. From a scientific point of view, this approach is interesting, but there is little hope that system integrators would spend the

necessary effort in integrating such an amount of sensors. Thus, unfortunately, this approach seems to be too complex and cost intensive.

Another way establishing safe human-robot cooperation was published in the works of (Henrich & Gecks, 2008b). The proposed approach for scene reconstruction is based on an image analysis module originally based on the work of (Henrich et al., 2008a). The vision system tries to identify pixels that belong to the real robot. The system provides some foreground detection with a pixel classification method, which identifies single pixels belonging to the robot, to foreground objects or to the static background. This research group also implemented a dynamic path planning module. But without knowing significant parameter of the human kinematics, path planning is restricted to avoidance of obstacles. Human-robot cooperation is otherwise not feasible.

At a first glimpse, the work of (Knoop et al., 2006) has a similar goal of introducing the human pose which is motivated by service robotics taking into account a humanoid robot and a human co-worker. Significant differences are that the author reported by applying his method for markerless reconstruction of the human body is dependent on hand skin colour detection. The proposed system, called VooDoo, runs in less than 15 frames per second as was reported by the authors (Knoop et al., 2006). Thus, this foregoing is not capable to deal in a safety critical industrial robotic cell. Furthermore, no occlusion detection was reported, which are of great interest especially when it comes to cooperation, due to safety considerations and reasoning about human actions in a blind spot.

An extended version of the VooDoo system was later published in (Lösch et al., 2009). This work concentrates on the time consuming initialisation which is based on a silhouette-based approach. The method proposed argues the negative influences of colour image dependant methods and thus uses the silhouette-approach for the initialisation. But the same author applies the VooDoo system after initialisation of the human kinematical model which is strongly dependant on the skin colour detection.

It is interesting, that all of the authors deal with safe human-robot interaction or cooperation, but only few of the authors are really trying to estimate and calculate significant parameters of the human kinematics. Also, there are approaches that are taking into account hand skin colour detection and simultaneously call these methods markerless.

## 2.2 Human pose reconstruction

In the subsequent section an overview for pure markerless human body tracking approaches will be given. The overview cannot raise a claim to be complete. The papers are presented in chronological order.

The paper of (Fua et al., 2002) presents an implicit surface approach for a generic and robust method handling articulated structures of the human body. The main contribution of this work is the description of a mathematical formalism with simplified and robust implementation of articulated soft objects. The soft object approach is advantageous because of using stereo and silhouette data, providing accurate shape description by a small number of parameters and explicit modelling of 3-D geometry.

The work of (Kehl et al., 2005) proposes a markerless full body pose tracking method which is based on the integration of multiple cues such as edges, colour information and

volumetric data. The human model is reconstructed by applying the stochastic meta descent (SMD) method to super-ellipsoids. The colour information is used to resolve self-occlusions, while edge information provides better accuracy and more robustness.

The work of (Caillette & Howard, 2004a) presents a robust method for real-time visual human body tracking by applying a hierarchical 3-D reconstruction from multiple camera views. Individual body parts are tracked by using 3-D blobs. The blob tracking is based on volume and colour information. The dynamics of the blob model is the highlight of the paper. Self-occlusions and noisy data are also investigated by experiments.

Real-time full human-body tracking based on markerless multi-view image sequences is presented in (Caillette & Howard, 2004b). The full approach is realized taking into account three steps: acquisition, reconstruction and tracking. The main idea of the method is based on reconstructing a 3-D voxel based representation of a person using multiple web cams providing colour images. Self-occlusions are also discussed as well as ambiguous poses. The novelty of the approach is a statistical reconstruction method taking colour features and blobs into account.

The authors of (Jenkins et al., 2006, 2007) present a method for kinematic pose estimation based on monocular image sequences as well as action recognition based on the results of the kinematic reconstruction. The motion primitives are modelled as nonlinear dynamic systems which are applied to predict expected motions. Goal of this paper is the inversion of the estimation process which means estimating motion primitives from measurements of the nonlinear dynamical human body. For these reasons, a particle filter is applied to fulfil this task.

The authors in (Azad et al., 2008) argue that the most challenging problem in human motion capture is the high-dimensional search space. A novel approach presented by the authors is build up on a particle filter framework which combines edge cues and 3-D hand tracking as well as a distance cue for upper body tracking as was proposed by the authors in an earlier paper. To overcome the problem of finding the inverse kinematics for the arm model the authors suggest a solution based on the so-called annealed particle filter approach. Another advantage is that this method does not depend on an initialization method. Proper model alignment is achieved by using fusion method and an adaptive shoulder approach.

The paper (Wan et al., 2008) proposes a method for markerless kinematic reconstruction which is based on voxel information generated from a multi camera set-up and the shape from silhouette method. The volume data is then considered as a Markov random field. A predefined human body model is then matched with the volume data. The matching task is formulated as an energy minimizing function. Thus, the problem is transformed into a 3-D graph construction. The minimizing of the graph problem is achieved by application of max-flow theory. The final reconstruction of the model is calculated using Powell's algorithm.

Based on video streams from a time-of-flight camera, the work of (Zhu et al., 2008) presents a model-based, Cartesian control theoretic approach for human pose estimation. The human body model consists of 17 degrees of freedom and models the upper body. The overall runtime cycle achieves about 10 frames per second. The presented approach is also feature based. Special features are the implemented joint limit avoidance and self-penetration avoidance.

The paper of (Jensen & Paulsen, 2009) is focused on gait analysis using a time-of-flight camera. Thus, an articulated model is fitted in each frame to the data by using a Markov random field. Self-occlusions are treated by smoothing missing data. The created model is cut into cycles, which are then fitted via Fourier method to achieve a cyclic model. The final features that are calculated are speed, cadence, step length and range of motion.

Based on the combination of several particle filters with physical simulation of a flexible body model, the work of (Hecht et al., 2009) describes a new approach for markerless human motion tracking. No inverse kinematics is needed for the physical simulation. Experimental results show that this approach runs with 10 FPS on regular PCs.

The dissertation thesis of (Zhu 2009) presents a computational framework for human-pose estimation from depth image sequences. The approach is feature based and takes kinematic constraints including joint limits and self-collision avoidance into account (see Zhu et al., 2008). Another approach is based on dense correspondence between consecutive frames of articulated human models. Both approaches are coupled via temporal prediction using Bayesian information integration.

The paper of (Mussi et al., 2010) presents a GPU-based implementation of a markerless full-body articulated human motion tracking system. The body reconstruction is based on image sequences from multiple cameras. The tracking task is formulated as a multi-dimensional nonlinear optimisation problem and solved by the particle swarm optimisation (PSO) method. The optimisation searches the best matched between a virtual pose silhouette and the actually pose extracted from the image sequences.

The problem of human pose reconstruction is of great interest and presents a challenging research topic, as exemplified by all presented publications. In the realm of human-robot cooperation and interaction, its purpose follows the higher goal of recognising human actions and situations.

### 2.3 Situation and activity recognition

Recognition of human activities and situation awareness is a premise for advanced safe human-robot cooperation. The most prominent methods used for action recognition systems are based on probabilistic methods, e.g., hidden Markov Models (HMMs) (Krüger et al., 2007; Raamana et al., 2007; Wu et al., 2008). These methods are widely used for application in speech recognition and other domains and, thus, their capabilities have been demonstrated. Moreover their theoretic foundations are well understood and investigated.

Though, according to (Shi et al., 2004), HMM are not suitable for recognition of parallel activities. Thus, propagation networks have been introduced. In these networks each node is associated with an action primitive and embeds a probabilistic duration model. Temporal and logical constraints are enforced by conditional joint probabilities. Similar to HMMs, a multitude of propagation networks are evaluated for approximating the observation probability.

(Minnen et al., 2003) states, that purely probabilistic methods are not suitable for recognition of prolonged activities. Their presented approach implements parameterised stochastic grammars.

The application of knowledge based methods for action recognition tasks is scarce, but work on scene interpretation using logical formalisms has been conducted. In the realm of semantic web, Description Logics are used for defining ontologies and knowledge management. Efficient algorithms have been developed for reasoning with Description Logics. Thus, its application in logics based situation and activity recognition became accepted.

In (Hummel et al., 2007), Description Logics are used for reasoning about traffic situations and understanding of intersections. Deductive inference services are used to reduce the intersection hypotheses space and to retrieve useful information for the driver.

In (Tenorth & Beetz, 2009), a system is presented, which uses Prolog in order to process knowledge in the context of robotic control. It is especially designed for use with personal robots. Knowledge representation is based on Description Logics and processed via an Ontology Web Language (OWL) Prolog plug-in. In contrast to our approach, the Prolog based reasoning system is not used to recognize activities or reason about situations. Instead, it is used to query on its environmental model. Actions and events are observed by the processing framework and used as knowledge facts. The knowledge base can be extended by using embedded classifiers in order to search for groups of instances that have common properties.

Scene interpretation by analysing table covers using Description Logics was conducted by (Neumann & Möller, 2008). Reasoning was based on temporal and spatial relations of visually aggregate concepts. Besides probabilistic information for generation of preferred interpretations, visual evidence and contextual information is used. In (Möller & Neumann, 2008), this work was broadened to cope with general multimedia data.

A comprehensive approach for situation-awareness is introduced in (Springer et al., 2010). This approach includes context capturing, abstraction and decision making. The combined framework manages sensing devices and reasoning components which allow using different reasoning facilities. Thus, logical reasoning can be used for high level decision making.

These last examples including our contributions show that the usage of Description Logics bears great potential. Hence, its adoption in the situation and action recognition task incorporated into the MAROCO framework.

### 3. System overview

The MAROCO framework implements an architecture achieving human centred computing realising safe human-robot interaction and cooperation due to advanced sensor technologies and fancy algorithms. An introduction of an intermediate state of the MAROCO system is given in (Graf & Wörn, 2009a). In the following, the advanced and augmented architecture is presented (Fig. 1). In this section, modules and functions are introduced and linked to Fig. 1 by referencing the given numbers in brackets.

Closing the kinematic chain in an environment with human agents and robots is especially meaningful and a premise in case of contact based cooperation scenarios. Thus a sensor calibration step is part of the framework {1}. The kinematic chain consists of the robot coordinate systems, the coordinate systems of human agents, the environmental model and finally the coordinate system of the 3D camera system.

The sensor system consists of a single depth sensing camera based on the time-of-flight principle which is developed and distributed by the company PMD Technologies. The resolution of the camera system is at the moment limited to 200x200 pixels. The advantage about the used 3D sensor technology is that it provides depth images as well as amplitude images. Amplitude values are a means to evaluate remissions of the active illumination of the camera system. The remission is influenced by objects in the scene and allows for adaption of algorithms towards increased robustness and effectiveness.

Due to this fact, the usage of cheaper sensors like the Microsoft Kinect camera is not feasible. Furthermore, because our sensor is mounted at the ceiling, the included human tracking of the Kinect system would render useless. The installation of the sensor system at the ceiling is meaningful in order to avoid the reach of humans or machinery, thus, allowing for a consistent sensor setup and enforce safety requirements.

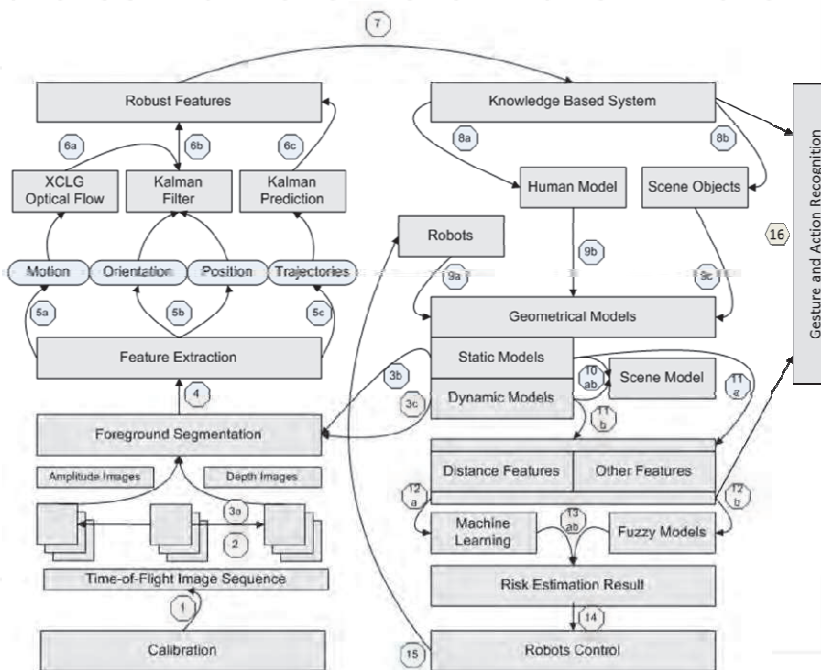


Fig. 1. System architecture of the MARCOCO framework.

In order to isolate relevant information from background clutter [4], background subtraction techniques are used. Our approach is based on Gaussian Mixture Models and advances on works of (Stauffer & Grimson, 2000; Lee, 2005) with adaptations due to requirements of human-robot interaction and the used sensor model. Background modelling incorporates a priori knowledge and can be learned by applying a variety of techniques.

Detection of human presence is done by a decision process depending on selective discriminating features based on foreground information [4]. Therefore, algorithms based on eigenvalue analysis, depth measurements of pixel distributions, the distribution of

connected components and finally motion features generated from optical flow computations {5a} are applied to decide whether the pixel cluster is generated by a human being or not.

MAROCO, the framework realising the system-architecture, provides also a flexible and complex kinematical model for human bodies {8a}. Due to the usage of a single 3D sensing camera mounted at the ceiling, a limited subset of degrees of freedom of the human kinematics is modelled. The kinematic features to be estimated are

- the shoulder and elbow angles of the left and right arm,
- the head position,
- the height of the person, and
- the upper body position {5b}.

These features are processed and generated by means of sequence analysis. Temporal information is incorporated by methods like Kalman filtering, Kalman prediction (Bar-Shalom et al., 2001) and optical flow estimation (Graf et al., 2010b). Thus, robust features are generated out of noisy data. Then all these features are supplied to a pattern recognition module which decides whether the provided features belong to a human model or a scene obstacle {7}. Obstacles are not recognized but represented by their bounding cylinders.

All gathered information and features are then used to construct geometrical models {9a, 9b, 9c}. Static and dynamic objects and agents are merged into an environmental scene model {10a, 10b} (Fig. 2).

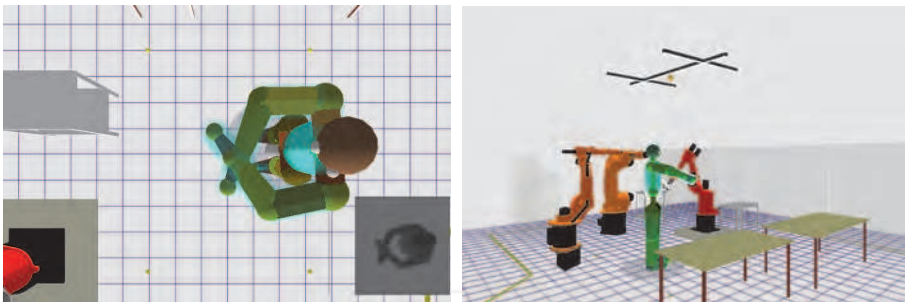


Fig. 2. Reconstructed human kinematics and environment model. Left image also shows depth coded grey scale sensor data in lower right corner.

Working with geometric information rather than pixel-based models results in great benefits concerning runtime behaviour. Using the 3D sensor and applying algorithms purely based on pixel processing (e.g. Graf & Wörn, 2008) is expensive in the meaning of computational time.

The generated robust features are used, besides other distance measurements, to estimate the risk. Feature estimates and distance calculations are then passed to machine learning methods {12a} and to functional evaluation {12b} (Graf et al., 2010a). Risk quantification can be used for influencing robotic behaviour {14} by either reducing motion velocity or adapting the motion path (Graf et al., 2009). This in turn changes representation of robot models {15}.



All information about human and robot kinematics can be used to reason about situations and human activities (Graf et al., 2010c) {16}. This allows recognising actions and drawing conclusions about expectations towards robotic behaviour.

#### 4. Theoretical considerations and algorithms

In this section, more detailed insights into our approaches and implementations are given. First, estimation and computation of robust features is detailed. Afterwards, methods for risk estimation and minimisation are presented. This section concludes with a description of the recognition module of MAROCO which allows reasoning about situations and activities.

##### 4.1 Robust features

In order to model human kinematics many features have to be robustly estimated. One kind of these features is based on motion analysis of the 3D sensor data. A means of motion analysis presents the estimation of the Optical Flow field. This technique is used in image sequence analysis and robotics for a long time (Horn & Schunk, 1981; Lucas & Kanade, 1981). It can be understood as the apparent motion of intensity structures in an image sequence. Our approach of computing Optical Flow fields advances on the combined local and global method (CLG) first introduced by (Bruhn et al., 2005a). The CLG method uses an isotropic Gaussian in order to reformulate the original data term formulated by (Horn & Schunk, 1981).

Our approach extends on this procedure by adapting Gaussians to the underlying distribution of pixels. Thus, it is called XCLG method (Graf et al., 2010b). The Optical Flow is influenced by its neighbourhood and, therefore, pixels at positions of edges or curves need special consideration. Through analysis of image edges, Gaussians are oriented and stretched along the principal axis which is congruent to the edge. The isotropic Gaussian of the CLG method is then substituted by the adapted Gaussian (Fig. 3).

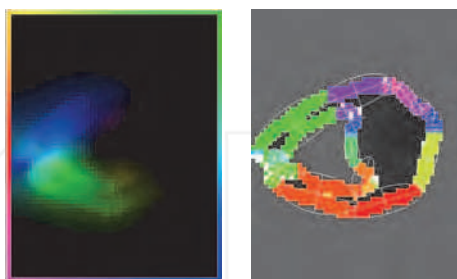


Fig. 3. Optical Flow field and anisotropic Gaussians adapted to underlying edges. The arms are moved towards each other.

Due to the fact that Optical Flow computations are an iterative process, usually, thousands of point wise iterations have to be applied to achieve significant results. For achieving real-time capabilities, application of standard numerical techniques, like Jacobi, Gauss-Seidel or successive over relaxation (SOR), is not feasible. The probably most efficient technique known today solving this kind of equation systems are so called multigrid solvers. They are often applied to sparse equation systems. In (Bruhn et al., 2005b) real time computations of

Optical Flow fields are reported using multigrid solvers. Thus, our approach uses multigrid solvers, which are implemented for general purpose GPU processing. This allows for real-time computations and effective use of motion analysis for robust features.

Other features include estimates about head and body orientation. These are computed through eigenvalue/eigenvector extraction of spatial pixel distributions. For this purpose, the depth images are segmented using additional estimations about body height and body part size relations. The orientations are determined by following assumptions:

- The head orientation is assumed to be the eigenvector corresponding to the larger of the two eigenvalues of the covariance matrix of the head pixel distribution.
- The upper part of the body orientation is assumed to be the eigenvector corresponding to the smaller of the two eigenvalues of the covariance matrix of the shoulder pixel distribution.

Through application of a windowed Kalman filter to past angles calculated from eigenvector analysis, estimations of orientations achieve greater robustness. An adapted Kalman filter is also used to fuse different information sources, such as motion analysis through Optical Flow computations, orientation estimates and arm poses. More details concerning the Kalman filter can be found in (Graf & Wörn, 2009a; Graf et al., 2010b).

The arm poses are also important features. These are estimated through the identification of three key points: shoulder, elbow and hand. Arm segments between these points can be linearly interpolated. In order to estimate the positions of the key points, skeletonisation succeeds a segmentation step. Afterwards the skeleton is mapped onto a graph and the arm poses are determined through path analysis in the graph. This approach takes also occlusions into account. Occlusions can be caused by either arms covering each other or by a robot pose covering human arm segments (Graf, 1010).

## 4.2 Risk quantification

Today's application of robotics in industrial environments is characterized by isolation of robots and humans due to safety concerns. Realising close human-robot collaboration requires evaluation of situations regarding a measure of danger for the human. Risk quantification depending on human and robot kinematics can result in adaption of robot motion and, thus, guarantee safety for human co-workers.

Assignment of a risk value to a situation has to take into account many different parameters of the human and robot kinematics. The main idea is that there is greater danger for a human co-worker, if he is not aware of robot movement. Also the distance between robots and the human agent are of importance.

A method for providing great flexibility in building a knowledge base is the application of two-threaded fuzzy logics (Kiendl, 1997). Two-threaded fuzzy logics allow encoding positive and negative rules in a knowledge base. That reduces the number of necessary rules compared to standard fuzzy logic systems. A detailed description of the implemented fuzzy system and the corresponding rules can be found in (Graf et al., 2010a).

In order to connect the results from the positive and negative rules accumulations, so called hyperinference operators are necessary. In (Kiendl, 1997) a few operators, like a strong and a

weak veto, are introduced. The strong veto operator is defined by (1), where  $\mu(u)$  defines the association function of fuzzy sets,  $\mu^+$  and  $\mu^-$  define the results of the accumulation of positive rules and negative rules respectively.

$$\mu(u) = \begin{cases} \mu^+(u), & \text{if } \mu^-(u) = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Thus, this operator does not respond to the area under the activated positive rule and the negative rule is overly weighted. The great flexibility of two-threaded fuzzy logic systems is bypassed through application of the strong veto operator.

The weak veto operator is defined by:

$$\mu(u) = \begin{cases} \mu^+(u), & \text{if } \mu^+(u) \geq \mu^-(u) \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Therefore, if the area under the negative rule is greater than the one under the positive rule the veto is applied. This action is desirable. On the opposite, if the area corresponding to the negative rule is smaller than the area under the positive rule the veto is not applied. Thus, the area under the negative rule has no influence on the outcome in all those cases. This behaviour is not desirable.

As consequence, a novel operator was implemented which is a trade-off in comparison to the strong and weak veto operators. It is defined as:

$$\mu(u) = \begin{cases} \mu^+(u) - \beta^-(u), & \text{if } \mu^+(u) > \mu^-(u) \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

In Fig. 4, the response characteristics of the proposed operator are presented. The construction of the novel veto operator begins by subdivision of the area under  $\mu^-$  into three parts. At first, the  $\alpha$ -cut of the curve is determined according to the output of the activated negative rule. Then, an orthogonal line is generated as shown in Fig. 4 (bottom row). This defines three parts of the area under the operator. The outer area elements are identical due to the symmetric characteristic of the operator and described by  $\beta^-$ . The adequate output of the veto operator is then generated by  $\mu^+ - \beta^-$ .

This proposed method for risk estimation can be implemented to evaluate a situation in real-time. Furthermore, its effectiveness is demonstrated in the section about experimental evaluation (Sec. 5).

### 4.3 Risk minimisation

As stated in the last section, the risk evaluation is used to influence robotic behaviour in order to guarantee safety for the human agent. In the context of industrial robotics, the efficiency of task performance of robots is very important. Thus, simple adaption of motion velocities does not suffice. A more advanced method is to actually re-plan the robots' path with dynamic safety constraints imposed by the moving human agent.

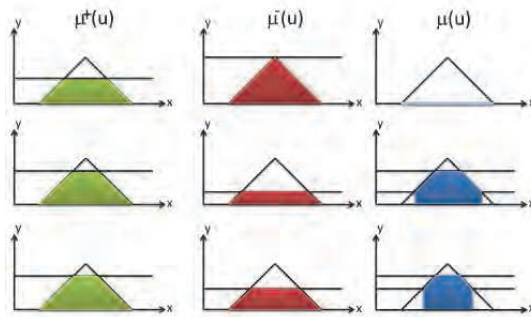


Fig. 4. Response characteristic of the novel veto operator.

The path planning takes place in the robots' configuration space. This space is interspersed with nodes which are connected to a graph structure. Association of risk estimates and configuration space is achieved by evaluation of each node in the graph. The path planning takes these evaluations of configurations into account and returns a safe and shortest path from and to given configurations. It uses a modified A\* search in configuration space to do so. A look-a-head functionality is used to re-evaluate a future path segment and detect impending collisions before they actually occur. In such a case a re-planning is invoked.

The implemented technique allows for fast and responsive re-planning without violating real-time constraints. Details about its implementation can be found in (Graf et al., 2009).

#### 4.4 Situation and activity recognition

All the methods and functionality presented above enable safe human-robot interaction and cooperation. But in order to actually achieve cooperation, situations and human activities need to be recognised and according conclusions about robotic behaviour need to be drawn. As pointed out in Section 2.3. Description Logics are suited for reasoning about context and, therefore, about situations and actions.

In (Graf et al., 2010c), a first approach towards the application of Description Logics for situation awareness is presented. An external reasoning system is used as inference facility. A MAROCO module must, therefore, fulfil at least the tasks of establishing a communication interface with the Description Logics reasoner, managing the knowledge base and managing the reasoner results. An overview of the subcomponents is given in Fig. 5. The communication is achieved through the so called DIG-interface which was defined by the Description Logic Implementation Group. It uses a TCP connection to transmit XML messages. Many reasoners support this interface definition, which allows the separation of application and reasoner by the means of programming language and execution place.

General knowledge and knowledge about individuals in the domain can be distinctly separated and defined in a Description Logic knowledge base. Common knowledge defines the terminology of the domain and, thus, is declared in the terminology box, hence TBox. Declarations about individuals and their properties are centralised in the assertion box, hence ABox. This allows for modular and reusable knowledge bases and, thus, for more efficient coding of knowledge (Hummel et al., 2007).

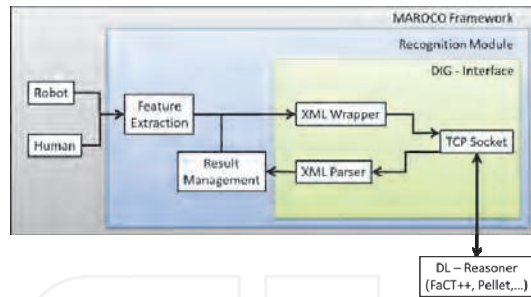


Fig. 5. Components of the recognition module embedded in the MAROCO framework.

The DIG-interface implements a so called *Tell&Ask* (Baader et al., 2010) functionality. The definition of the knowledge base is achieved through *tell* operations. Reasoner results and information can be retrieved through *ask* operations. Modifications successive of *ask* operations are not defined by the DIG-interface. Consequently, the knowledge base needs to be re-established in each runtime cycle in order to incorporate changed sensor data into the recognition process. The differentiation of domain knowledge and assertional knowledge of Description Logics is disregarded by the DIG-interface.

The recognition module handles assertions depending on the current kinematical human model and robot specific parameters and domain specific knowledge. Thus, the distinction of TBoxes and ABoxes is represented internally.

As the assertional knowledge depends on kinematical parameters, a feature extraction component is applied in order to fill the attribute values of the assertions (Fig. 5).

Due to the fact that there is currently no object recognition implemented in the MAROCO framework, objects are included into the situation recognition through means of simulation. Thus, a human agent can hold working tools or measurement devices in his hands. Also, the simulation enables the robot gripper to be holding objects like work pieces. In future works, these purely simulated features will be incorporated into the demonstrator as well. For now, these virtual features enable evaluation of effectiveness and capabilities of the recognition system. Moreover, by incorporating virtual features, the recognition module can reason about probable interactions and generate expectation towards robotic behaviour, e.g., prepare a work piece or hand tools on to a human co-worker. These expectations can be used directly or in context of the recognized actions as input for a possible task planner for realizing concrete close human-robot collaboration. Implementation of a task planning module is a logical consequence and will be done in near future.

Taking temporal information into account during reasoning is accomplished by defining an *after*-role between different actions. This role can be regarded as precondition for actions, because certain actions can only be recognised if certain other actions occurred prior. In order to facilitate temporal dependencies between actions, previously recognised actions are stored and retrieved during knowledge base recreation. This functionality is taken over by the reasoner result management component (Fig. 5).

Furthermore, the knowledge base implements concepts of complex actions which consist of other actions. The temporal relationship includes these complex concepts. Thus, parallel and subsequent occurring actions can be processed and recognized.

Detailed description of the implemented ontologies and knowledge base are given in (Graf et al., 2010c). Evaluation and discussion of effectiveness and capabilities of the presented recognition module conclude the section about experimental evaluation (Sec. 5).

## 5. Experimental evaluation

Due to the application of diverse methods in the framework MAROCO, there is a need for diverse testing and evaluation. In the following sections, especially experimental evaluation of accuracy, efficiency and effectiveness are presented. Also, the capabilities of the proposed methods and their fusion in the framework are discussed.

### 5.1 Robust features and human kinematics

Determination of motion features through computation of Optical Flow fields allows interpretation about direction and apparent motion. These can be identified by representing the Optical Flow field as vector field. In the context of human-robot interaction, rates of changes are of great importance, as they indicate motion intensity. Thus, the vector length plays an important role in the estimation and filtering of robust features. For evaluation purposes, the XCLG method was compared with the CLG method by computing the Optical Flow field of an image sequence and by evaluation of magnitude differences of both vector fields. As shown in (Graf et al., 2010b), the CLG method underestimates vector lengths by 26%-47%. These results demonstrate the greater accuracy of the XCLG method considering vector lengths.

Due to the implementation of the Optical Flow computation using general purpose graphics unit processing, the presented method achieves real-time capability. The computation times are also compared to the CLG method. Each method was implemented with SOR solver and multigrid solver. Different camera systems with differing resolutions were used. Moreover, the publicly available “Yosemite” image sequence was used to verify the results with internationally respected data. The results of these runtime tests are presented in Table 1.

In [ms]	IFM O3D (50x64)			
	CLG		XCLG	
	SOR	MG	SOR	MG
CPU	164	2	172	11
GPU	20	5	23	8
In ms	PMDTec CamCube 2.0 (204x204)			
	CLG		XCLG	
	SOR	MG	SOR	MG
CPU	2150	25	2225	85
GPU	89	20	102	33
In ms	Yosemite			
	CLG		XCLG	
	SOR	MG	SOR	MG
CPU	4020	48	4260	295
GPU	164	32	220	85

Table 1. Overall time for computation of Optical Flow using CLG and XCLG.

Due to lack of ground truth data, evaluation of tracking results in real world applications is challenging. Thus, testing implemented algorithms was done indirectly through examination of overlap of sensor data and tracked kinematics. In order to compare sensor data and tracking results, the human kinematics is projected back onto the image plane. Thus, the cycle from sensor data to tracking data and back again is closed (Fig. 6). Congruency of foreground pixels and back-projection can be interpreted as accuracy of the kinematics reconstruction step and, thus, is a measure of the reliability of the algorithm.

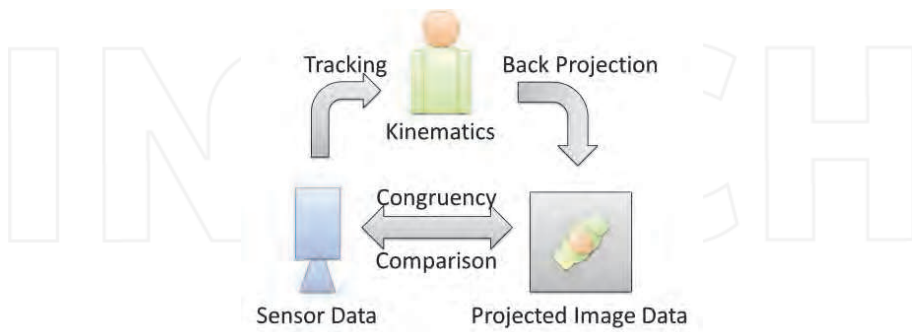


Fig. 6. Data processing cycle for evaluation of tracking data.

In order to analyse the congruency, different human motion sequences were used. Each sequence consists of approximately 600 frames. In Table 2, the results are summarised. The motion sequences include simple motions like forward and backward (1), only arm movements (2), turning around (3), standing still (4), and arbitrary motion (5).

Sequence	Mean	Variance
1	96.60	19.82
2	89.46	14.71
3	90.07	9.67
4	93.13	2.33
5	91.60	14.53

Table 2. Quantification of the congruency rate.

These results show that the reconstruction of the human kinematics is congruent with the observed sensor data to a large degree. Due to the fact that risk estimation is based on the kinematics reconstruction, this degree of congruence has great importance. After all, it influences directly the safety capabilities of the system, because risk estimation is done purely based on reconstructed kinematics.

## 5.2 Risk management

For the evaluation of selected risk estimation methods, different experiments with varying methods have been conducted. These methods include e.g., simple measures like shortest distance between human and robot, methods of differing complexity implemented as Gaussian mixture models and Support Vector Regression.

Compared to simple distance measures and Gaussian mixture models, the two-threaded fuzzy system allows for precise modelling of situations and according risk assignments. For examination purposes the same sensor input sequence was evaluated by the above mentioned methods and risk assignments were compared. The results confirm our assumption about flexibility and effectiveness of the here presented fuzzy method (Fig. 7). Further details can be found in (Graf et al., 2010a).

The conducted experiments also demonstrate that training a Support Vector Regression resulted in unreliable and noisy risk estimation compared to the implemented two-threaded fuzzy system. Thus, the fuzzy system outperforms the Support Vector Regression and is used as preferred risk estimation method. Grounded on the results of the here described fuzzy logic implementation, safety and efficiency for human-robot cooperation is achievable in real-time.

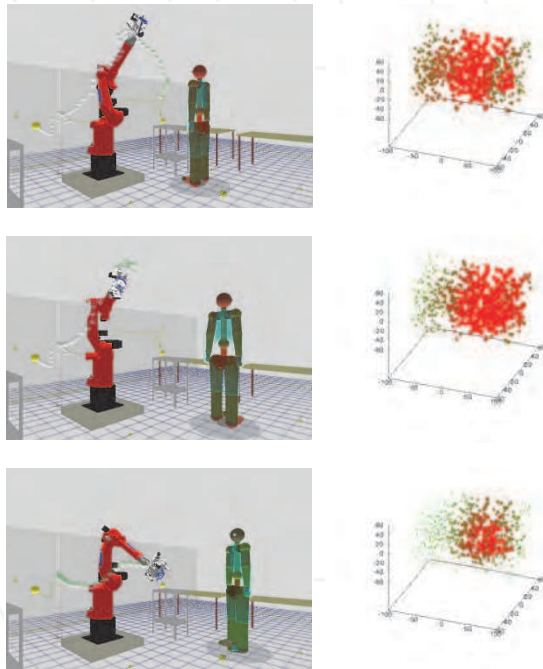


Fig. 7. Selected situations of human posture and corresponding configuration subspace for first three robot axes. Size and colour of each node of the subspace is determined by risk assignment.

For the experimental analysis of the path re-planning technique different scenarios were tested in simulation. Especially, the size of the configuration space graph was subject of evaluation in order to capture scalability of the algorithm. For testing, a sequence of human motion was recorded and played back during simulation (Fig. 7). Thus, arbitrary movements were recorded and thereby the simulation was related to real-world setups. The tested scenarios do not consider human-robot interaction or cooperation, but instead, the robot has a given repetitive task and has to avoid human co-worker in its working area. The



overall hold-up time of the robot reaches about 27% during evaluation. The results are presented in Table 3. Further details are explained in (Graf et al., 2009).

# Vertices of Graph	100	1000	2000
Avg. Path Length	4.62	5.98	7.34
Hold-up Time [%]	63.96%	27.34%	38.96%

Table 3. Results of path planner run-time analysis.

The presented results concerning risk quantification and minimisation demonstrate the effectiveness of guaranteeing safety for human agents in the realm of close human-robot collaboration.

### 5.3 Situation awareness

In order to evaluate the situation and activity recognition module of the MARCOCO framework, different courses of action were executed. On the one hand, efficient analysis of different scenarios requires automated means of feature value setting. Thus, value pre-sets were incorporated into the framework which allows for usage of pre-defined feature vectors. Such pre-sets enable investigation of interesting use cases without capturing sensor data. Also, recognition results can be directly related to defined feature changes through pre-sets. Nevertheless, recognition based on actual sensor data is compulsory in order to evaluate recognition results over time and prolonged actions (Fig. 8).

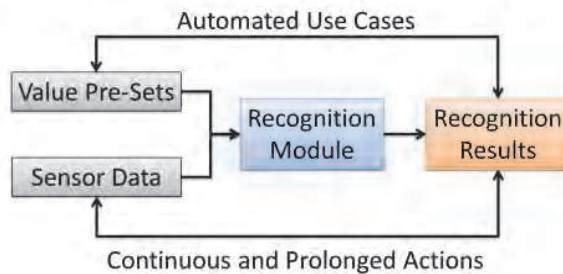


Fig. 8. Usage of feature value pre-sets and actual sensor data.

Based on these pre-sets and on actual sensor data all experiments were conducted. Natural movements and transitions between actions have been tested and special use cases have been investigated.

Table 4 summarises the results of 2140 recognition cycles. The average processing time amounts to approximately 550 ms. The lower bound is less than half of this. There are also casual outliers which take up to 10 seconds. Further investigations based on feature value pre-setting have shown that long duration times might not be directly linkable to the recognition module itself (Graf et al., 2010c). Though, further research needs to be conducted on optimising runtime behaviour of the recognition module.

# Recognition cycles	2140		Max [ms]	9705
Ø Response time [ms]	551.78		# > 1000 ms	17 (0.79%)
Min [ms]	216		# > 5000 ms	4 (0.18%)

Table 4. Results from evaluation of the recognition module.

In Fig. 9, results of the recognition module depending on the human pose are depicted. It demonstrates the capabilities of analysing solely kinematical features of the human agent and its relations to a robot.



Fig. 9. Left: Human agent is watching the robot. Recognized situation: Monitoring. The robot is expected to carry on with its task of following a planned path. Right: Human agent is communicating. The complex action to signal a left turning movement is recognized. The robot is expected to comply with user instructions.

By adapting the virtual features according to the generated expectations the interaction between reasoner results and robotic behaviour can be demonstrated. Thus, the capabilities of the presented approach reach beyond sole activity and situation recognition. By generating expectations towards robot behaviour, an understanding of the situation can be achieved. This induction of relations between concepts can hardly be realized by purely probabilistic methods. The achieved processing cycle time of approximately 550 ms does not allow for safe cooperation based only on the recognition module. Thus, the MAROCO framework uses its implemented techniques and algorithms to enforce safety and real-time capabilities during robot motion.

## 6. Conclusion

The presented framework MAROCO and the incorporated approaches are based on the identification of different modules that have to be taken into account when designing a system for close human-robot collaboration based on a depth imaging sensor. Experimental results give confidence in continuing to strive for true contact based cooperation between robot and human. Thus, our work is a stepping stone for future development.

Thus far, a system was implemented which analysis depth images taken from a 3D camera system mounted beneath the ceiling. Robust features like motion, head and body orientation, position and arm poses are robustly and efficiently estimated. Evaluation has shown that high accuracy is achieved.

All these features are used to reconstruct the human kinematics which is the foundation for risk quantification. A two-threaded fuzzy system with a novel hyperinference operator is

implemented for risk evaluation of situations according to human pose features and relation between human and robots. The system is flexible and effective. In comparison to Support Vector Classification and other means of risk estimation the two-threaded fuzzy system is the most reliable and accurate one.

Results of risk estimation are used for adapting robotic behaviour. Adaption is realised by path re-planning if a look-a-head functionality determines impending collisions of human and robot before they occur. That allows for safe and efficient path traversal and, thus, reduced time of robot hold-up times.

In order to achieve true human-robot cooperation situation awareness and action recognition is necessary. A module for realising this task was implemented using Description Logics for defining appropriate ontologies and for reasoning. The presented system is capable of recognising subsequent, parallel and dependent actions and can generate expectation towards robotic behaviour. Thus, the system reaches beyond sole situation recognition and enables understanding human activities.

Future work will carry on development towards a system that achieves close human-robot collaboration. There are still many open challenges that need to be tackled before this goal is reached. The usage of more than one camera can either widen the supervised work area or enable multi-view capturing of the scene.

Currently, only one human agent can be detected and its kinematics can be reconstructed. Extension of the presented algorithms is needed for multi-human pose estimation. Moreover, the algorithms need to be adapted in order to cope with more arbitrary movements of human co-workers. Some movements are not covered by the current human pose reconstruction process, e.g., stooping down.

Object recognition and semantic mapping of the work area are also important means for modelling interactions of human agents and robots with the surrounding environment. Particularly object recognition will enable more diverse and differentiated analysis of situations. Semantic mapping of objects and places in the robots' work area will allow for recognition of human action plans and, thus, a better understanding of intentions behind human actions.

As pointed out above, implemented virtual features need to be realized for the demonstrator. Moreover, runtime optimisations of the current situation and activity recognition module need to be investigated and implemented. This will allow for evaluation of real-world scenarios of interaction and cooperation. Also, realisation of industrial applications with the MAROCO system will enable evaluation of capabilities and user acceptance. This experimental evaluation can be realised stepwise beginning with simple risk minimisation and collision avoidance, advancing on to telepresence-like systems and concluding in fully autonomous human-robot cooperation.

The MAROCO system emphasises on real-time computation and safety for human co-workers. Nevertheless, the implemented system is a research base and does not permit safety certification. Hopefully, achievements of the human-robot cooperation research community will migrate into applicable industrial systems. Safety regulations and engineers have to adapt to this young field of research.

## 7. Acknowledgment

We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Karlsruhe Institute of Technology (KIT).

## 8. References

- Azad, P., Asfour, T., & Dillmann, R. 2008. Robust Real-time Stereo-based Markerless Human Motion Capture. *Proceedings of the IEEE/RAS International Conference on Humanoid Robots*, Daeheon, Korea, pp. 700-707
- Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. 2010. "The Description Logic Handbook", 2nd Edition, Cambridge University Press
- Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. 2001. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*, Wiley, New York
- Bruhn, A., Weickert, J., & Schnörr, C. 2005a. Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. In: *International Journal of Computer Vision* 61(3), pp. 211-231
- Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., & C. Schnörr, C. 2005b. Variational Optical Flow Computation in Real Time. In: *IEEE Transactions on Image Processing* 14:5, pp. 608-615
- Caillette, F., & Howard, T. 2004a. Real-Time Markerless Human Body Tracking Using Colored Voxels and 3-D Blobs, *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*
- Caillette, F., & Howard, T. 2004b. Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction, *Proceedings of the British Machine Vision Conference (BMVC)*, pp. 597-606
- Fua, P., Gruen, A., D'Apuzzo, N., & Plankers, R. 2002. Markerless Full Body Shape and Motion Capture from Video Sequences, *Symposium on Close Range Imaging*, International Society for Photogrammetry and Remote Sensing, Corfu, Greece
- Graf, J. 2010. Sichere Mensch-Roboter-Kooperation durch Auswertung von Bildfolgen, *Dissertation*, Fakultät für Informatik, Karlsruhe Institut of Technology, Germany
- Graf, J., & Wörn, H. 2008. An Image-Sequence Analysis System with Focus on Human-Robot-Cooperation using PMD-Camera. *Proceedings of Robotik 2008*, München, pp. 223-226
- Graf, J., & Wörn, H. 2009a. Safe Human-Robot Interaction Using 3D Sensor, In: *VDI Berichte 2067 Automation 2009*, VDI/VDE Gesellschaft Meß- und Automatisierungstechnik, Baden-Baden, Germany, pp. 1-12
- Graf, J., Czapiewski, P., Wörn, H. 2010a. Evaluating Risk Estimation Methods and Path Planning for Safe Human- Robot Cooperation, *Proceedings of ISR/Robotik 2010*, München, Germany
- Graf, J., Dittrich, F., & Wörn, H. 2010b. High Performance Optical Flow Serves Bayesian Filtering for Safe Human-Robot Cooperation. *Proceedings of Int. Symposium of Robotics (ISR) and VDI/VDE Robotik 2010*, München, Germany, 8 pages
- Graf, J., Puls, S., & Wörn, H. 2009. Incorporating Novel Path Planning Method into Cognitive Vision System for Safe Human-Robot Interaction. *Proceedings of IARIA Computation World: Cognitive 2009*, Athen, Greece, pp. 443-447

- Graf, J., Puls, S., & Wörn, H. 2010c. Recognition and Understanding Situations and Activities with Description Logics for Safe Human-Robot Cooperation, *Proceedings of the 2<sup>nd</sup> International Conference on Advanced Cognitive Technologies and Applications: Cognitive 2010*, Lisbon, Portugal, pp. 90-96
- Hecht, F., Azad, P., & Dillmann, R. 2009. Markerless Human Motion Tracking with a Flexible Model and Appearance Learning, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, pp. 3173-3179
- Henrich, D., Fischer, M., & Gecks, T. 2008a. Sichere Mensch/Roboter-Koexistenz und Kooperation. In: *Robotik 2008*, München, Germany
- Henrich, D., & Gecks, T. 2008b. Multi-Camera Collision Detection Between Known and Unknown Objects. *Proceedings of 2<sup>nd</sup> ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, Stanford, USA
- Horn, B., & Schunck, B. 1981. Determining optical flow. In: *Artificial Intelligence*, pp. 185-204
- Hummel, B., Thiemann, W., & Lulcheva, I. 2007. Description Logic for Vision-based Intersection Understanding, *Proceedings of Cognitive Systems with Interactive Sensors (COGIS)*, Stanford University, USA
- Jenkins, O., Gonzalez, G., & Loper, M. 2006. Monocular Virtual Trajectory Estimation with Dynamical Primitives. *Proceedings of the American Association for Artificial Intelligence Workshop on Cognitive Robotics (AAAI)*, Boston, Massachusetts, USA
- Jenkins, O., Gonzalez, G., & Loper, M. 2007. Recognizing human pose and actions for interactive robots. In: *Human-Robot Interaction*, chapter 6, pp. 119-138
- Jensen, R.R., & Paulsen, R.R. 2009. Analysis of Gait Using a Treadmill and a Time-of-Flight Camera, *Proceedings of the DAGM Workshop on Dynamic 3D Imaging*, Jena, Germany, pp. 154-166
- Kiendl, H. 1997. *Fuzzy Logik methodenorientiert*, R. Oldenbourg Verlag, München, Wien
- Lee, D. S. 2005. Effective Gaussian Mixture Learning for Video Background Subtraction, *Proceedings of IEEE Trans. Pattern Analysis and Machine Intelligence (27)*, pp. 827-832
- Lucas, B. D., & Kanade, T. 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proceedings of Imaging*, pp 121-130
- Lösch, M., Gärtner, S., Knoop, S., Schmidt-Rohr, S.R., & Dillmann, R. 2009. A human body model initialization approach made real-time capable through heuristic constraints, *Proceedings of the 14th International Conference on Advanced Robotics (ICAR)*, Munich, Germany
- Minnen, D., Essa, I., & Starner, T. 2003. Expectation Grammars: Leveraging High-Level Expectations for Activity Recognition, *Proceedings of Computer Vision and Pattern Recognition*, Vol. 2, pp. 626-632
- Mussi, L., Ivekovic, S., & Cagnoni, S. 2010. Markerless Articulated Human Body Tracking from Multi-View Video with GPU-PSO, *Proceedings of the 9<sup>th</sup> International Conference on Evolvable Systems*, York, UK
- Möller, R., & Neumann, B. 2008. Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval, In: *Semantic Multimedia and Ontologies*, Part 2, pp. 55-98, Springer London
- Neumann, B., & Möller, R. 2008. On Scene Interpretation with Description Logics, In: *Image and Vision Computing*, Vol. 26, pp. 81-101
- Kehl, R., Bray, M., & Van Gool, L. 2005. PHI'05 Workshop in Conjunction with ICCV 2005, Beijing, China

- Knoop, S., Vacek, S., & Dillmann, R. 2006. Sensor Fusion for 3D Human body Tracking with an Articulated 3D Body Model. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, Florida, USA
- Krüger, V., Kragic, D., Ude, A., & Geib, C. 2007. The Meaning of Action: A Review on Action Recognition and Mapping, *Proceedings of Advanced Robotics*, Vol. 21, pp. 1473-1501
- Kulic, D. 2005. Safety for Human-Robot Interaction. Dissertation, Faculty of Graduate Studies, Mechanical Engineering, University of British Columbia, December 2005
- Raamana, P., Grest, D., & Krüger, V. 2007. Human Action Recognition in Table-Top Scenarios: An HMM-Based Analysis to Optimize the Performance, In: *Lecture Notes in Computer Science*, Vol 4673, pp. 101-108
- Shi, Y., Huang, Y., Minnen, D., Bobick, A., & Essa, I. 2004. Propagation Networks for Recognition of Partially Ordered Sequential Action, *Proceedings of Computer Vision and Pattern Recognition*, Vol. 2, pp. 862-869
- Springer, T., Wustmann, P., Braun, I., Dargie, W., & Berger, M. 2010. A Comprehensive Approach for Situation-Awareness Based on Sensing and Reasoning about Context, In: *Lecture Notes in Computer Science*, Vol. 5061, pp. 143-157, Springer Berlin
- Stauffer, C., & Grimson, W. E. L. 2000. Learning Patterns of Activity Using Real-Time Tracking, *Proceedings of IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747-757
- Tenorth, M., & Beetz, M. 2009. KNOWROB - Knowledge processing for Autonomous Personal Robots, *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*
- Thiemermann, S. 2005. Direkte Mensch-Roboter-Kooperation in der Kleinteilmontage mit einem SCARA-Roboter. *Dissertation*, Fakultät für Maschinenbau der Universität Stuttgart
- Wan, C., Yuan, B., & Miao, Z. 2008. Markerless Human Body Motion Capture using Markov Random Field and Dynamic Graph Cuts, In: *The Visual Computer* 24(5), pp. 272-380
- Winkler, B. 2008. Konzept zur Sicheren Mensch-Roboter-Kooperation auf Basis von Schnellen 3-D Time-of-Flight Sensoren. A. Verl und M. Hägele (Eds.), VDI/VDE Gesellschaft für Meß- und Automatisierungstechnik (GMA), Düsseldorf, Deutsche Gesellschaft für Robotik: *Robotik 2008*, pp. 147-151. Tagung München
- Wu, Y., Chen, H., Tsai, W., Lee, S., & Yu, J. 2008. Human action recognition based on layered-HMM, *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1453-1456
- Zhu, Y., Dariush, B., & Fujimura, K. 2008. Controlled human pose estimation from from depth image streams. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-8
- Zhu, Y. 2009. Model-Based Human Pose Estimation with Spatio-Temporal Inferencing, *Dissertation*, The Ohio State University



## **Human Machine Interaction - Getting Closer**

Edited by Mr Inaki Maurtua

ISBN 978-953-307-890-8

Hard cover, 260 pages

**Publisher** InTech

**Published online** 25, January, 2012

**Published in print edition** January, 2012

In this book, the reader will find a set of papers divided into two sections. The first section presents different proposals focused on the human-machine interaction development process. The second section is devoted to different aspects of interaction, with a special emphasis on the physical interaction.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Stephan Puls, Jürgen Graf and Heinz Wörn (2012). Cognitive Robotics in Industrial Environments, Human Machine Interaction - Getting Closer, Mr Inaki Maurtua (Ed.), ISBN: 978-953-307-890-8, InTech, Available from: <http://www.intechopen.com/books/human-machine-interaction-getting-closer/cognitive-robotics-industrial-environments>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

