Online Formal Verification of Robot Trajectories for Guaranteed Safety of Humans

Aaron Pereira[1] and Matthias Althoff[1]

***Abstract*— Guaranteeing safe behaviour of robots is not easy, especially in the presence of humans, whose behaviour is always unpredictable. We show how a frequently-updating, continually-verifying controller can guarantee safety of nearby humans in a formal way while maintaining efficiency of the robot. Two different models of the human behaviour are presented: a model based on extreme human motion as specified in ISO standards, and another, kinematic model parametrised by test data from a range of human subjects performing extreme movements. We implement the controller in an experimental setup, and find that both models allow the robot to work efficiently when the human is safe and stop when the human is in danger of colliding with the robot; the ISO-based model is less cautious and sometimes does not account for fast human motion. The kinematic model, which accounts for all motion, is more restrictive towards robot motion but nevertheless allows the robot to operate when the human is not in danger.**

I. Introduction

Formal methods in robotics are increasingly of interest to researchers, e.g. for planning and scheduling [1], or path planning in compliance with a specification [2]. In particular, their ability to *guarantee* given properties makes them particularly attractive in safety-critical areas.

Robots that can ensure safety of humans working alongside them have potential to make automation safer, more efficient, and more flexible by removing the need for safety cages. Several approaches require the robot to react fast to collisions, but for higher inertia robots or robots carrying dangerous tools it may be desirable to avoid collision altogether. For example, [3] proposes nested safety zones around the robot, with increasingly more conservative robot behaviour as the human breaches these zones. Other approaches, e.g. [4] scale speed or modify the robot trajectory using a measure of danger based on the position of the human. While such approaches work well, they do not *formally guarantee* a safety property of the human-robot collaborative scenario.

To guarantee safety with respect to future human movement, the robot must predict this movement using a model of human behaviour in its environment. Challenges arise from the humans' nondeterminism and speed. In contrast to mobile robots, for which ambulatory movement is most relevant e.g. [5], fixed-base robots working alongside humans at a workstation must consider upper body movement. Human arms can move fast, making accurate prediction even short times in the future difficult. For this reason, most predictions of upper body movement are probabilistic, e.g. [6], [7].

[1]The authors are with the Department of Informatics, Technische Universität München, 85748 Garching, Germany. aaron.pereira@tum.de, althoff@in.tum.de

While these may account for the most likely movement, for industrial application one must guarantee safety *all* of the time – what happens when a human moves unexpectedly, for example when sneezing, grabbing a falling object or recoiling from touching something hot or sharp? We therefore aim to predict the entire set in space that a human could occupy, the *reachable occupancy* [8]. We do this using reachability analysis, a technique for predicting all future states of a system given a set of initial states and uncertain dynamics, which has been used in safety verification of autonomous cars [9]. The reachable occupancy grows very fast in time, so we use it in a continuously verifying controller along with frequent sensor updates to verify only the immediately next section of the robot trajectory. This concept was introduced in [10] for mobile robots and used in [11] for fixed-base manipulators.

In this work, we show that this approach works in practice and that a robot trajectory which is provably safe can be formally verified online, even when accounting for all human motion. We compare two different models for predicting the reachable occupancies: a model based on the assumptions of human motion from ISO standards, and a model based on extreme movement data collected from several test subjects performing fast movements.

This paper is structured as follows: we introduce the concept of online verification in the next section. In Sec. III we present the two models of human movement, which we evaluate in Sec. IV.

II. Concept

The principle of online verification is that the robot *does not execute a movement before verifying its safety*. In a long-term (global) planner, we plan a desired trajectory for the robot; this planner may use probabilistic prediction to avoid likely occupancy of the human. Given this long-term desired trajectory, we verify it one piece at a time and ensure that, at the end of each piece of desired trajectory, we always have a failsafe manoeuvre available to bring the robot to a safe state (in our case, safe means stationary) before collision would be possible. We call the piece of desired trajectory plus the failsafe manoeuvre the "short-term plan"

Figure 2 illustrates the approach: the short-term path from t_k to $t_{s,1}$ consists of the piece of desired trajectory from t_k to t_{k+1} followed by a failsafe manoeuvre until $t_{s,1}$. This is verified safe prior to t_k, so from time t_k to t_{k+1}, this is the plan the robot executes. Simultaneously, it constructs the next short-term plan (the desired trajectory from t_{k+1} to t_{k+2} followed by a new failsafe manoeuvre from t_{k+2} until $t_{s,2}$),

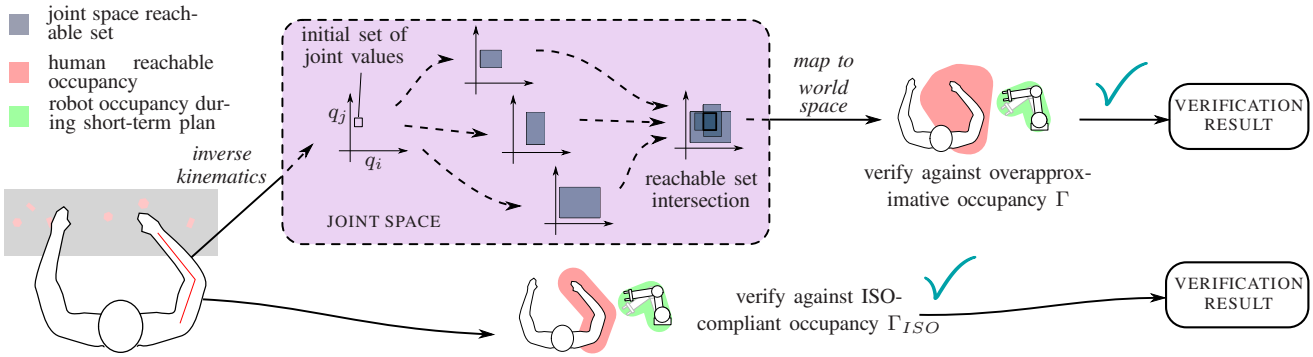

Fig. 1. An illustration of the prediction and verification. When the robot has planned a short-term plan, it predicts the human reachable occupancy until the end of this plan, either using an overapproximative model (above, detailed in Sec. III-B) or with an ISO-compliant model (below, detailed in Sec. III-A). It then checks for intersection of the human's occupancy with its own occupancy, and the short-term path is verified safe if there is no intersection.

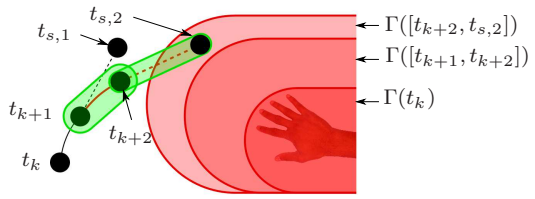

Fig. 2. Verifying safety of a short-term plan. The occupancy of the robot in the next short-term plan (green volumes) is checked against that of the human (red volumes). The failsafe plans are in fact path-consistent with the desired trajectory; here they are shown off-path for clarity of illustration.

calculates its own occupancy over the duration of the next short-term plan and verifies it against the reachable occupancy of the human (from t_{k+1} until $t_{s,2}$). Both occupancies are constructed in an overapproximative way, meaning they enclose all space which could possibly be occupied by the human and the robot during those times, accounting for measurement uncertainty and uncertain dynamics.

If the occupancies do not intersect, then the human could not reach the robot during the next short-term plan; this short-term plan is therefore verified safe and can be executed starting from t_{k+1}. Otherwise, the next short-term plan is rejected and the robot begins the failsafe manoeuvre from the current short-term plan. By keeping the failsafe manoeuvre path-consistent with the desired trajectory, the robot can easily plan future short-term plans even while executing the failsafe manoeuvre, without the need for expensive spatial replanning (the failsafe manoeuvres are shown off-path in Fig. 2 for clarity of illustration).

Several algorithms exist for planning a time-optimal stopping trajectory. In our implementation, we used a modification of the algorithm by [12] to produce path consistent stopping trajectories which satisfied acceleration and jerk limits of the joints. Furthermore, since the movement of the robot is known, calculating the spatial occupancy of the robot along its short-term plan is less of a challenge than calculating that of the human, whose intentions are unpredictable. For example, the method from [13] can be used to quickly and conservatively generate an overapproximation to the robot occupancy.

The main challenge in this approach, therefore, is calculating the future occupancy of the human. We define the *reachable occupancy of the human* $\Gamma([t_a, t_b])$ as the set of all areas in space which could possibly be occupied by the human during time interval $[t_a, t_b]$. Since the exact reachable occupancy is incalculable (as it is the reachable set of a nonlinear hybrid system [14]) we calculate a *tight overapproximation*, i.e. a set in $\mathbb{R}^3$ which encloses the exact reachable occupancy, while excluding as much unreachable space as possible. Nevertheless, the reachable occupancies grow fast by virtue of the fact that humans can move very quickly. The effectiveness of this approach lies in the fact that the reachable occupancies only need to be predicted until little more than the stopping time of the robot, rather than over the whole trajectory, therefore do not grow unmanageably large. In the next section, we focus on how to calculate the human occupancy.

III. Accounting for Human Movement

Humans can move fast and unpredictably. The relevant industrial standards for calculating stopping distances during emergency stops of machinery [15] assume a maximum speed of human movement, which is sometimes insufficient to account for fast movements. Below, we detail one model to predict human occupancy conforming to the ISO standard and next, a model accounting for all human movement. The methods to calculate both occupancies are shown in Fig. 1.

A. ISO-compliant model

In ISO 13855 [15] the maximum speed of a human, for the purposes of calculating stopping distances, is $v_{\max} = 1.6m/s$ for upper-body movement only and $v_{\max} = 2.0m/s$ for full-body movement. We use the latter, and "expand" the human by $v_{\max} \cdot t$ to obtain the reachable occupancy up until time t after the sensor observation.

For representation, we use sphere-swept volumes (SSVs), which are the Minkowski sum ($\oplus$) of a convex hull G of a set of points and a sphere H, defined as $G \oplus H = \{g + h \mid g \in G, h \in H\}$. We call an SSV where the set of points is of order 2 (i.e., G is a line segment) a *capsule*. Of course, an SSV where G is of order 1 is a sphere. The *radius* of the

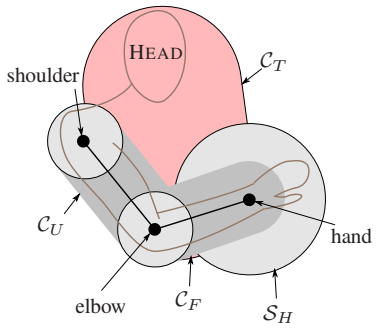

Fig. 3. Modelling of the arm as two capsules $\mathcal{C}_U$ and $\mathcal{C}_F$ and a sphere $\mathcal{S}_H$. Torso and head enclosed in another capsule $\mathcal{C}_T$. Left arm not shown.

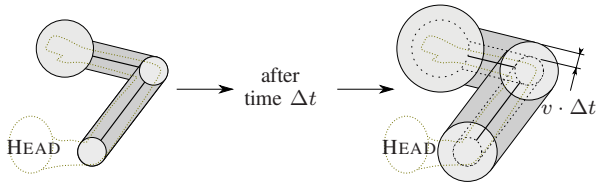

Fig. 4. Prediction Γ_{ISO} based on maximum speed $v = 2m/s$ from [15].

SSV is defined as that of the sphere H, and the *defining points* are the vertices of G.

We obtain the human pose from sensors. The torso and head is enclosed in a single capsule $\mathcal{C}_T$, see Fig. 3. Obtaining the positions of shoulder, hand and elbow, we can then define the arm occupancy. We enclose the arm in 3 SSVs: one capsule encloses the upper arm, with the defining points at the shoulder and elbow, and another encloses the forearm, defined by the elbow and the hand. A sphere encloses the hand. The radii of $\mathcal{C}_T$ is $0.35m$, that of $\mathcal{C}_U$ and $\mathcal{C}_F$ is $0.1m$ and that of $\mathcal{S}_H$ is $0.205m$, taken from measurements of human hands [16].

We then augment the radii by $v_{\max}\Delta t$ as shown in Fig. 4, where Δt is the difference between the end of the short-term plan $t_{s,2}$ and the time of the observation t_{obs} to obtain the occupancy over the time interval $[t_{obs}, t_{s,2}]$ according to the ISO assumptions on human motion, which we call $\Gamma_{ISO}[t_{obs}, t_{s,2}]$

B. Overapproximative model

Humans can, however, move at speeds greater than $2m/s$, especially when executing involuntary or reflex movements like sneezing, catching a falling object or swatting an insect. In [8], we present a model of human movement which is intended to account for all human motion, parameterised by capturing data from a range of humans performing extreme movements. We briefly recapitulate here. The approach consists of offline and online phases. Offline, 38 test subjects (12 female, 26 male, ranging from 18-49 years) performed a range of movements as fast as possible. The motion capture of their arms was fitted to a kinematic parameterisation of the arm using inverse kinematics, to find the ranges of joint positions, velocities and accelerations.

During online operation of the robot, inverse kinematics

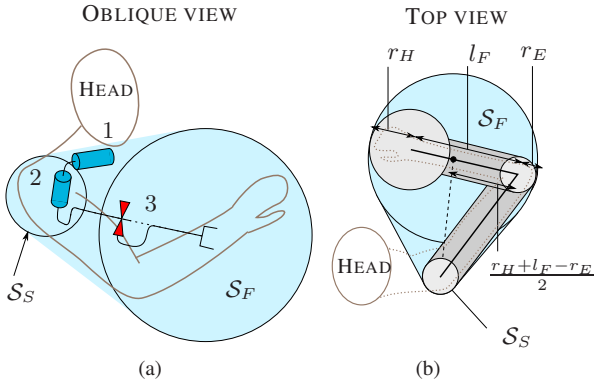

Fig. 5. (a) The kinematics of the 3-DOF model. q_1 and q_2 are joint angles of the first two rotational joints; r_3 is the extension of the prismatic joint. (b) Evaluating r_F the radius of $\mathcal{S}_F$

is used to obtain the joint values of the human arm using the same kinematic parameterisation. With the parameters obtained in the offline phase, reachability analysis is used to find the set of joint values at future time intervals and this is translated into an occupancy in space, against which we verify the short-term plan of the robot as described in Sec. II.

1) Kinematic model: The kinematic model is 3-degree-of-freedom (3DOF) and is shown in Fig. 5a. A point in the middle of the forearm is the end effector. There are two orthogonal revolute joints at the shoulder, but in contrast to more common 4-DOF models e.g. [17], extension of the elbow and rotation of the upper arm around its own axis is replaced with extension of a prismatic joint, and the occupancy is represented as the convex hull of a sphere $\mathcal{S}_\mathcal{F}$ at the end-effector which encloses the forearm, and a sphere $\mathcal{S}_S$ at the shoulder: see Fig. 5b. By the property of convexity, the upper arm is also enclosed. This parameterisation offers the advantage of lower dimensionality and avoidance of the kinematic singularity when the elbow is at full extension.

2) Dynamic model: As previously mentioned, we perform reachability analysis on the initial set of states, which are obtained from inverse kinematics of sensor data and enlarged to account for measurement uncertainty. Let $\mathcal{Q}(0)$ and $\dot{\mathcal{Q}}(0)$ be the sets of initial joint positions and velocities respectively; these are Cartesian products of intervals. We then use 3 models, each individually accounting for joint positions, velocities and acceleration limits, to generate 3 reachable sets. Since all sets are overapproximative, their intersection is also overapproximative and smaller than any of the 3 alone. By using many simple models with linear dynamics, we avoid having to construct one model with complex, hybrid dynamics: the reachability analysis of hybrid systems is more time consuming to calculate and speed is critical in our application. The models used to calculate the reachable set from time 0 to time t are:

1) a 0^{th} order model of maximum joint position: $\mathcal{R}^{(1)}_{\mathbf{q}}([0,t]) = [\mathbf{q_{inf}}, \mathbf{q_{sup}}]$,
2) a 1^{st} order model of maximum joint velocity: $\mathcal{R}^{(2)}_{\mathbf{q}}([0,t]) = \mathcal{Q}(0) \oplus [\dot{\mathbf{q}}_{\mathbf{inf}}\ ,\ \dot{\mathbf{q}}_{\mathbf{sup}}]t$,

3) a 2^{nd} order model of maximum joint accelerations, $\mathcal{R}_{\mathbf{q}}^{(3)}([0,t]) = \text{CH}(\mathcal{Q}(0), \mathcal{Q}(0) \oplus \dot{\mathcal{Q}}(0)t \oplus [\ddot{\mathbf{q}}_{\mathbf{inf}} \frac{t^2}{2}, \ddot{\mathbf{q}}_{\mathbf{sup}} \frac{t^2}{2}])$.

Here, $[\mathbf{q_{inf}}, \mathbf{q_{sup}}], [\dot{\mathbf{q}}_{\mathbf{inf}}, \dot{\mathbf{q}}_{\mathbf{sup}}]$ and $[\ddot{\mathbf{q}}_{\mathbf{inf}}, \ddot{\mathbf{q}}_{\mathbf{sup}}]$ are interval vectors representing position, velocity and acceleration joint limits obtained from analysis of test subjects as previously mentioned, and CH is the convex hull operator. In the last model we enclose the convex hull in a Cartesian product of intervals (the first two models are already obtained as products of intervals). The reachable set in joint space of the positions of the human arm from the observation time t_{obs} to the end of the short-term plan $t_{s,2}$ is $\mathcal{R}_{\mathbf{q}}([t_{obs}, t_{s,2}]) = \bigcap_{i=1}^{3} \mathcal{R}_{\mathbf{q}}^{(i)}([t_{obs}, t_{s,2}])$

3) Conversion to Cartesian space and accounting for moving shoulder: One unsolved problem of the approach from [8] is that it does not account for movement of the base coordinate system of the kinematic chain at the shoulder, which can translate and rotate as the human moves. To model this, a more complex model of the movement of the torso, shoulder complex as well as the ambulatory movement would be required.

Instead, we adopt the approach from [18], where the reachable sets of a higher-order model are accounted for by a lower-order model by enlarging the initial set and adding disturbances. In our case, we account for rotation of the base coordinate system at the shoulder by enlarging the reachable set in the dimensions of the two revolute joints by $0.1rad/s \cdot \Delta t$.

The joint space reachable set $\mathcal{R}_{\mathbf{q}}([t_{obs}, t_{s,2}])$ is converted in to Cartesian space using the method from [13]. We call the obtained set $\Gamma([t_{obs}, t_{s,2}])$. We account for translation of the shoulder using the maximum speed of human movement from ISO 13855 ($v_{\max} = 2.0m/s$) – similarly to the ISO approach from Sec. III-A, we simply add $v_{\max}\Delta t$ onto the radius of the SSV $\Gamma([t_{obs}, t_{s,2}])$.

IV. Evaluation and Discussion

We tested this approach in a setup with a Schunk LWA4P robot controlled over CAN bus by a Speedgoat Real-time Target Machine running Simulink 2015b, operating at $500Hz$. The human was detected using a 6-Camera Vicon Vero 1.3 infrared motion capture system[1] operating at $250Hz$, using retroreflective markers placed on the body. We estimated overall latency of the sensors as $20ms$. We used the GJK algorithm to detect collisions [19]. Our long-term plans were pre-programmed, straight line movements of the end effector. When testing the overapproximative model, we used the model from Sec. III-B to calculate the occupancies of the arm; and when testing the ISO model, we used the model from Sec. III-A. The ISO model was used to account for the torso in both approaches.

The results of two test runs can be seen under `www6.in.tum.de/pub/Main/Pereira/video.mp4` Computations with both models took under $500\mu s$ on average, and the occupancy prediction and verification alone took under $200\mu s$ with one human in the workspace (tracking more than one humans can be parallelised). As expected, we observed that the behaviour using the overapproximative model was more cautious – the robot moved slower around the human (since it verified itself unsafe more often) and only resumed normal operation when the human was at quite a large distance from the robot. Using the ISO model, the robot was able to work more efficiently, but did not always achieve a stop if the human moved very fast. Freeze-frames of the robot performing a failsafe manoeuvre using the overapproximative model is shown in Fig. 6.

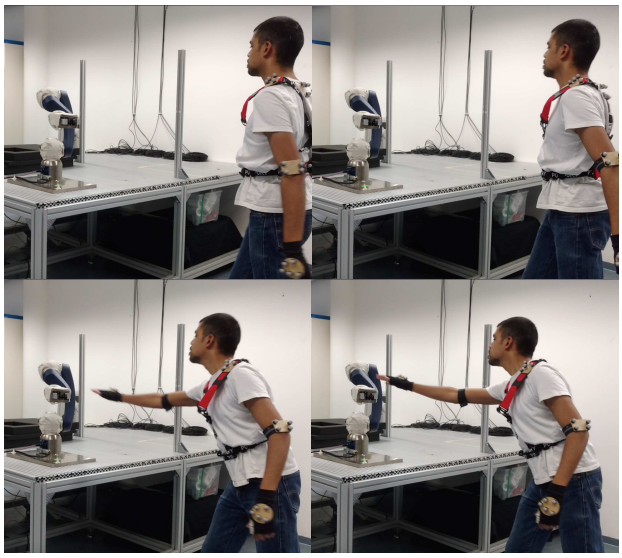

Fig. 6. Freeze-frames at $0ms$ (top left), $167ms$ (top right), $375ms$ (bottom left) and $458ms$ (bottom right) The robot is moving in the top two freeze-frames, but has stopped before the freeze-frame at $375ms$.

Although we used high-end sensors, this method also works with less accurate and fast updating sensors. We tried the method using a Microsoft Xbox Kinect 2 [2], using the provided skeleton-tracking from the SDK and only using the ISO-approach. Performance was worse, due to the reduced frame rate and higher latencies, i.e. the robot behaved more cautiously and was stationary at a greater distance from the human. This is akin to human behaviour when our sensory capabilities are reduced, e.g. we drive vehicles slower in low-visibility weather. In a robot cell, the sensors should have good visibility over the robot workspace and could be combined with other sensing modalities such as pressure-sensitive floors, to ensure that all humans are easily detected. To guarantee safety, the robot would assume the worst case, i.e. that there may be humans just outside of the cameras' field of view; hence where the field of view is limited, the robot would behave more conservatively.

In our implementation, the stopping times of the robot were no more than $200ms$. Since the stopping time determines the prediction horizon for the reachable occupancies, higher-inertia robots, which may take $1s$ or more to stop,

[1] `www.vicon.com/products/camera-systems/vero`, retrieved 15.3.17

[2] `xbox.com/xbox-one/accessories/kinect`, retrieved 15.3.17

may be forced to move more slowly (which reduces their stopping time) or keep more distance from the human. In practice, very high-inertia robots are unlikely to be moving when in close proximity with humans; interaction is likely to be during maintenance, or if the robot is positioning a heavy part for the human to perform dextrous operations upon, like lifting a car chassis for a human to fit the wiring.

In [8] we show that the reachable occupancies Γ are significantly larger in volume than Γ_{ISO}, which should mean that the short-term plans are verified unsafe more often. Though in our evaluation the robot moved more slowly when using Γ_{ISO}, the actual loss of performance in an industrial scenario depends on the exact setup of the robot cell. Further tests are required to compare this formally verified method with e.g. static safety zones. One advantage of our method is that no risk assessment or manual determination of safety zones is necessary: the danger to the human is assessed on the fly. Secondly, using static safety zones, the robot is forced to limit its behaviour whenever the human enters the collaborative area, even if the robot is actually in another area of its workspace. Our method avoids this, since collision risk is checked online.

A further observation was that, although the ISO model did not always account for all movements so the robot did not always stop in time, it might be acceptable if the robot is allowed to be moving during collision. The new Technical Standard [20] defines allowable maximum pressures and impact forces when collisions between humans and robots occur. The continually verifying controller concept can be adapted to a safety criterion where impact forces and pressures are within allowable limits, instead of requiring the robot to be stationary.

V. Conclusion

Online verification is a powerful tool for reactive robot behaviour, which can guarantee a safety property during execution of a trajectory, by continuously predicting the future occupancy of the human and adjusting its behaviour accordingly. We show how the assumptions on human behaviour – using the assumptions from relevant standards or accounting for all human motion – affect the performance of the robot. Even while accounting for all possible movement of the human, the robot can operate efficiently. Such a robot that can replan and verify its own behaviour online has potential to change the face of automation in the 21st century.

Acknowledgment

The authors are grateful to Dario Beckert, Florian Böck and Natalie Reppekus for work on the implementation. This research received funding from the Marie Curie Actions of the EU's 7th Framework Programme under REA grant number 608022. The authors also gratefully acknowledge financial support by the EC project UnCoVerCPS under grant number 643921.

References

[1] A. Orlandini, A. Finzi, A. Cesta, and S. Fratini, *TGA-Based Controllers for Flexible Plan Execution*. Springer, 2011, pp. 233–245.

[2] K. He, M. Lahijanian, L. E. Kavraki, and M. Y. Vardi, "Towards manipulation planning with temporal logic specifications," in *IEEE Int. Conf. Robotics and Automation*, 2015, pp. 346–352.

[3] A. Zanchettin, N. Ceriani, P. Rocco, H. Ding, and B. Matthias, "Safety in human-robot collaborative manufacturing environments: Metrics and control," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 882–893, 2016.

[4] D. Kulić and E. Croft, "Pre-collision safety strategies for human-robot interaction," *Autonomous Robots*, vol. 22, no. 2, pp. 149–164, 2007.

[5] F. Jovan, J. Wyatt, N. Hawes, and T. Krajník, "A poisson-spectral model for modelling the temporal patterns in human data observed by a robot," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2016.

[6] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2013, pp. 299–306.

[7] H. Ding, G. Reißig, K. Wijaya, D. Bortot, K. Bengler, and O. Stursberg, "Human arm motion modeling and long-term prediction for safe and efficient human-robot-interaction," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2011.

[8] A. Pereira and M. Althoff, "Overapproximative arm occupancy prediction for human-robot co-existence built from archetypal movements," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2016, pp. 1394–1401.

[9] M. Althoff and J. M. Dolan, "Online verification of automated road vehicles using reachability analysis," *IEEE Transactions on Robotics*, vol. 30, no. 4, pp. 903–918, 2014.

[10] S. Petti and T. Fraichard, "Safe Motion Planning in Dynamic Environments," in *Proc. of the IEEE-RSJ Int. Conf. on Intelligent Robots and Systems*, 2005.

[11] A. Pereira and M. Althoff, "Safety control of robots under computed torque control using reachable sets," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2015.

[12] T. Kröger and F. Wahl, "Online trajectory generation: Basic concepts for instantaneous reactions to unforeseen events," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 94–111, 2010.

[13] H. Täubig, B. Bäuml, and U. Frese, "Real-time swept volume and distance computation for self collision detection," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2011, pp. 1585–1592.

[14] A. Platzer and E. Clarke, *The Image Computation Problem in Hybrid Systems Model Checking*. Springer, 2007, pp. 473–486.

[15] "Safety of machinery - positioning of safeguards with respect to the approach speeds of parts of the human body," International Organization for Standardization, ISO Standard 13855:2010, 2010.

[16] S. Pheasant and C. M. Haslegrave, *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Taylor & Francis CRC Press, 2006, ch. Hands and Handles, pp. 143–160.

[17] N. Klopčar, M. Tomšič, and J. Lenarčič, "A kinematic model of the shoulder complex to evaluate the arm-reachable workspace," *Journal of Biomechanics*, vol. 40, no. 1, pp. 86–91, 2001.

[18] M. Althoff and J. M. Dolan, "Reachability computation of low-order models for the safety verification of high-order road vehicle models," in *Proc. of the American Control Conference*, 2012, pp. 3559–3566.

[19] E. G. Gilbert, D. W. Johnson, and S. S. Keerthi, "A fast procedure for computing the distance between complex objects in three-dimensional space," *IEEE Journal on Robotics and Automation*, vol. 4, no. 2, pp. 193–203, 1988.

[20] "Robots and robotic devices – collaborative robots," International Organization for Standardization, Geneva, Switzerland, ISO/TS Standard 15066:2016, 2016.