Research Statement

Heng Yang

1 Vision: Safe and Trustworthy Autonomy

Recent decades have witnessed a surging interest in deploying *robots* and *autonomous systems* to address grand societal challenges in transportation, energy, healthcare, agriculture, manufacturing, and space exploration, to name a few. As more robots enter our society and some of them start operating in close proximity to humans and assets (*e.g.,* in *safety-critical* applications such as autonomous driving, or *high-integrity* applications such as space robotics), a key question that the general public will ask before accepting robotics technology is: *are robots safe and can they be trusted to perform what they are designed to do*? Unfortunately, the current answer to this question seems to be *negative*. For example, a recent survey showed that "*nearly* 3 *in* 4 *Americans say autonomous vehicle technology is not ready for primetime*" [15]. Moreover, it is not rare to see reports about self-driving vehicles making fatal mistakes [10].

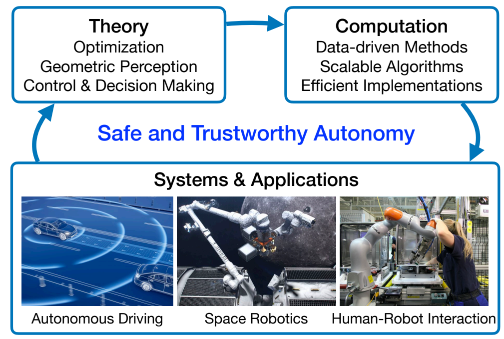

Figure 1: Safe and trustworthy autonomy requires an advanced toolbox that combines theory, computation, and system validation.

My research aims to enable safe and trustworthy autonomy for a broad range of high-integrity applications, by designing tractable and provably correct algorithms that enjoy rigorous performance guarantees, developing fast implementations, and validating them on real robotic systems (*cf.* Fig. 1).

My past research focused on robot (visual) perception and has led to the development of the first *certifiable outlier-robust geometric perception toolbox* (Section 2). A modern perception pipeline typically consists of a (deep) feature matching *front-end* and a geometric estimation *back-end*, where the back-end often solves a difficult *mathematical optimization* problem to estimate relevant geometric models (*e.g.,* Fig. 2 shows an example of estimating the 3D orientation and location of an incoming vehicle given a 2D image observation). In stark contrast with traditional geometric estimation algorithms that are fast but can *fail without notice*, my work has established the foundations of *certifiable algorithms* that can certify success and detect failure: they return an optimal estimate with a *certificate of optimality* for the majority of problems, but declare failure and provide *a measure of suboptimality* on (rare) worst-case instances [21, 3]. I built this new paradigm for the geometric estimation back-end by leveraging a combination of rigorous theory (*e.g.,* graph theory, robust estimation, semidefinite relaxation) and principled scientific computing (*e.g.,* large-scale convex optimization solvers, fast implementations), and demonstrated the resulting algorithms on safety-critical applications such as self-driving and space robotics. Algorithms in this toolbox also have broader impacts on robust estimation (*e.g.,* by providing a general methodology for optimal estimation in the presence of outliers) and mathematical optimization (*e.g.,* by showing how to solve large-scale degenerate semidefinite relaxations). I was honored to be selected as an RSS Pioneer, and also to talk about the impact of certifiable algorithms on safe autonomous driving on MIT News Spotlight [12].

An autonomous system, however, interweaves different modules (*e.g.,* perception, control, planning) and ensuring *system-level safety* requires much more than optimality of the geometric estimation back-end. **Therefore, my future research aims to expand the trustworthiness and performance guarantees beyond geometric estimation** (Section 3). **First**, I plan to guarantee safety of the entire visual perception pipeline (*cf.* Fig. 2), which not only includes the geometric estimation back-end, but also a learning-based front-end. This line of research will develop the theory and practice necessary for us to warrant that the solutions certified as *optimal* by the back-end correspond to a *correct* and *safe* understanding of the world, *e.g.,* in Fig. 2 to guarantee that the optimal pose is close to the groundtruth pose of the incoming vehicle. **Second**, I plan to integrate safe perception and safe control/planning in the context of *active perception* [2]. Although it is common in robotics to treat perception as an upstream task before planning and control, humans often *move to better see*, making perception an active and adaptive process. I believe designing a *feedback* mechanism between perception and action under a proper theoretical

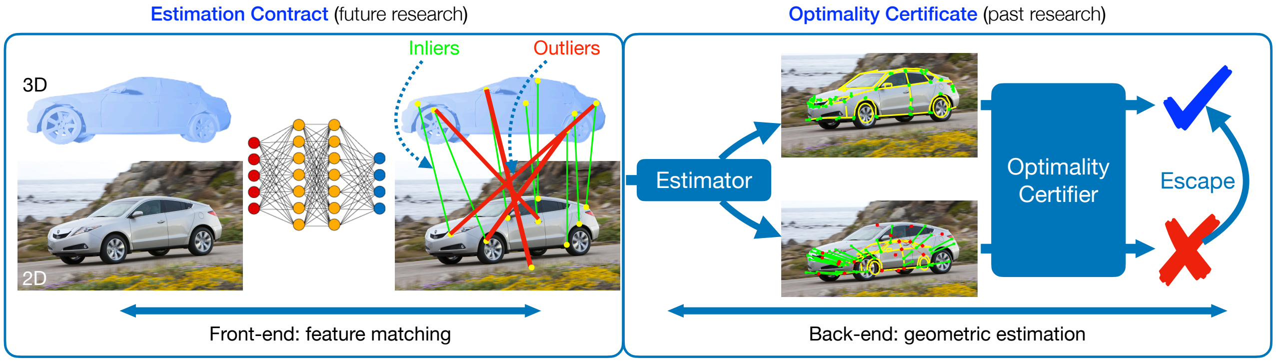

Figure 2: Example of a safe and trustworthy robot perception pipeline for vehicle pose estimation. Given a 2D image and a 3D model of an incoming vehicle, the pipeline first learns a neural network to detect and match semantic keypoints (such as mirrors and wheels) from 2D to 3D (typically referred to as the feature matching *front-end*), and then seeks the best 3D rotation and translation (*i.e.,* pose) to align the 3D model to the detected keypoints by solving a *mathematical optimization* (typically referred to as the geometric estimation *back-end*). Safety of the entire pipeline can be guaranteed by computing an *optimality certificate* from the back-end and verifying geometric conditions of the matches outputted by the front-end (what we call an *estimation contract*). My past research developed the first toolbox for computing *certifiably optimal* solutions for the back-end (Section 2, right-hand side in the figure), while my future research aims to investigate the design and verification of the estimation contracts (Section 3, left-hand side in the figure).

framework, while offering safety guarantees, is a stepping stone towards human-level intelligence. My past research equipped the perception module with *certifiability*, and provides a promising component for the feedback mechanism.

While I am deeply interested in contributing to the *theoretical* and *computational* foundations of robotics, I also recognize that dealing with *real robots in real-world complex environments* is one of the most important and intriguing aspects of robotics research, especially in the context of safe and trustworthy autonomous systems. Therefore, my future research will place significant emphasis on *validating and improving the theory and algorithms on real platforms* such as ground vehicles, manipulators, and drones. Real robots provide valuable datasets for us to refine the theoretical assumptions of our algorithms and to design data-driven methods to better predict the uncertainty of the real world. I hope the seamless combination of theory, computation, and system validation will unveil a future of safe and intelligent robots to the public.

2 Past Research: Certifiable Outlier-Robust Geometric Perception

Perception, at the front line of an autonomous system, consumes sensory information and builds an internal representation of the 3D environment. As mentioned in Section 1, a perception pipeline typically includes a *front-end* that extracts and matches relevant features from raw sensor data, and a *back-end* that estimates a geometric model given the feature matches, by solving a mathematical optimization (*cf.* Fig. 2). In practice, due to various sources of imperfections and uncertainties (*e.g.,* sensor noise, occlusions, misdetections, etc.), a large amount of *outliers* –measurements that tell no or little information about the underlying geometric models– can be generated by the front-end (*e.g.,* wrong matches shown by red lines in Fig. 2). Therefore, designing an *outlier-robust* estimation back-end that can tolerate a large number of outliers is of paramount importance for safe and trustworthy autonomy.

Unfortunately, from a theoretical standpoint, *safe perception is fundamentally intractable* because performing robust estimation by discerning *inliers* (*i.e.,* the correct measurements such as green lines in Fig. 2) from outliers, is known to be *NP-hard* and *inapproximable* [1]. Consequently, existing outlier-robust estimation algorithms are divided into *fast heuristics* (*e.g.,* RANSAC [8]) that run in real time but offer little performance guarantees, and *global solvers* (*e.g.,* branch-and-bound [9]) that guarantee optimality but run in worst-case exponential time.

My past research has led to the development of the *certifiable perception toolbox*, containing the first set of *tractable* outlier-robust geometric estimation algorithms with *provable guarantees*. This toolbox contains two *general-purpose* algorithms: a fast and robust *estimator*, and a trustworthy *certifier*.

Estimator: outlier-robust estimation with graduated non-convexity. Although outlier-robust estimation is theoretically intractable, I have shown that it is possible to design an estimator to solve, in polynomial time, *a majority of the problem instances* to global optimality. In [16] (ICRA 2020 best paper in robot vision, RAL 2020 best paper honorable mention), I developed a graduated non-convexity (GNC) scheme for robust estimation that can *tolerate up to* 70-80% *random outliers* in a broad range of geometric perception problems including point cloud and mesh registration, shape alignment, category-level object perception from 2D/3D landmarks [19, 13], and pose graph optimization in simultaneous localization and mapping (SLAM). Notably, in [13] (RSS 2021 best paper finalist), GNC, together with an imperfect keypoint detector, can accurately estimate the pose and shape of a wide range of vehicles in self-driving applications. Compared to existing robust estimation techniques that are either randomized or whose performance largely depends on a good initialization, GNC is both *deterministic* and *initialization-free*, which significantly increases the algorithmic repeatability and reliability, and decreases the amount of human design efforts (*e.g.,* parameter tuning). Recently, my collaborators and I have further made GNC almost parameter-free by leveraging insights from statistics [1], which makes GNC a plug-and-play algorithm. GNC *is implemented in both Matlab's navigation toolbox [11] and the popular GTSAM library for SLAM.* To boost the robustness of GNC against an extreme number of random outliers (*e.g.,* having 95% random outliers is common in point cloud registration), my collaborators and I have designed a general *graph-theoretic pruner*, called ROBIN [17, 14], that can filter out gross random outliers before executing GNC. Although ROBIN runs in worst-case exponential time, we leveraged parallel computing so that ROBIN exhibits real-time performance in practical applications. Combining ROBIN and GNC has led to TEASER [24], a point cloud registration library that can *tolerate over* 99% *random outliers and runs in milliseconds.* The open-source implementation of TEASER has over 800 stars on Github, and it is becoming a go-to approach for 3D rigid registration in robotics and computer vision.

Certifier: global optimality certification via semidefinite relaxation. GNC can solve common outlier-robust estimation to global optimality, but it comes with no optimality guarantees and it can fail without notice. Therefore, I proposed the notion of a "certifier" that aims to *certify the success* or *detect (and correct) the failure* of GNC. Designing an efficient certifier is grounded in rigorous theory and computation. First, I brought the machinery of polynomial optimization (POP) into robot perception and showed that many fundamental outlier-robust perception problems can be reformulated as POPs [21]. Second, although POPs are NP-hard in general, I, for the first time, empirically demonstrated that POPs arising from common geometric estimation problems are *polynomial-time solvable*, by designing *sparse* and *exact* convex semidefinite programming (SDP) relaxations based on Lasserre's theory of *moment and sum of squares relaxations* [21, 18, 20]. Solving the convex SDP can not only compute an optimal solution to the original POP, but also provide a *certificate of global optimality.* Third, in collaboration with world-renowned experts in convex optimization, I developed an SDP solver called STRIDE that solves exact semidefinite relaxations with unprecedented *scalability* and *accuracy* [23]. STRIDE is capable of certifying success or correcting failure of GNC in real robotic applications such as 3D reconstruction, satellite pose estimation, and vehicle pose and shape estimation [21]. The impact of STRIDE also goes beyond perception. POP is a universal modeling language for decision-making problems subject to complex constraints (*e.g.,* in robotic motion planning [6]), and STRIDE provides a general tool to solve them with optimality guarantees. Furthermore, solving large-scale degenerate SDPs has been a long-lasting challenge in mathematical optimization, and STRIDE is the first scalable algorithm that can solve large-scale exact SDP relaxations of POPs to high accuracy [23].

3 Future Research: Towards System-level Safe Autonomy

My past research has focused on designing a *certifiably optimal geometric estimation back-end* for visual perception (*cf.* Fig. 2). My future research aims to start from the certifiable perception back-end, and reach my long-term vision of system-level safe and trustworthy autonomy, by building an advanced toolbox combining theory, computation, and system validation (*cf.* Fig. 1). In the following, I detail two important steps that are crucial towards this goal.

Safe perception by integrating front-end and back-end (*cf.* Fig. 2). As shown in Fig. 2, a full perception pipeline contains not only geometric estimation, but also feature learning and matching. Therefore, optimality of the geometric estimation back-end may be vacuous without appropriate assumptions on the feature matching front-end (*e.g.,* if a robot is trying to localize an incoming vehicle, but all its predicted keypoints lie on a pedestrian, then the optimal pose estimation will not match the actual vehicle pose). To bridge the gap between optimality and safety, I plan to integrate the front-end and the back-end by investigating three topics. *First, I intend to establish "estimation contracts", which are problem-specific conditions on the input data (*e.g., *feature matches produced by the front-end) that ensure optimal solutions correspond to a correct understanding of the environment.* For example, in [24], my collaborators and I have developed such a contract for point cloud registration. I plan to delve deeper into the geometric foundations of robot perception to construct estimation contracts for other problems (*e.g.,* category-level perception). *Second, I aim to develop feature learning methods that are tailored to geometric estimation and its estimation contract.* Existing feature matching front-ends are typically trained separately from the estimation back-end and hence have two drawbacks. On one hand, the training process requires a large amount of data with labeled geometric models that may be expensive to acquire. On the other, it is challenging to ensure that the separately learned features will promote the correct estimation of the optimization back-end. Therefore, I propose to model the *joint* task of (supervised or unsupervised) feature learning and geometric estimation as a *bilevel programming* problem [4] and design efficient algorithms to solve it. In a preliminary work [22], I designed a general *self-supervised perception* framework that does not require labeled geometric models and achieves similar performance as compared to supervised oracles. Future work will extend [22] in seeking more efficient computational algorithms for solving the bilevel programming problem, and designing suitable loss functions to promote estimation contracts. *Third, I plan to verify the correctness of feature matching (in satisfying the estimation contract) by quantifying the uncertainty of neural networks.* Even though the learned front-end can output matches satisfying the estimation contract for the entire training dataset, real-world noise can perturb the output in unexpected ways. Consequently, it is important to bound the noisy perturbations such that estimation contract is satisfied during deployment. To this end, a promising direction is to leverage recent progress in neural network Lipschitz constant estimation [7]. Notably, a major technical tool for estimating the Lipschitz constant of a neural network is semidefinite programming, for which I have developed much expertise from my past research.

Safe autonomy via integrated perception and action.

While most existing research treat perception as a *passive* task (*i.e.,* to understand the environment given sensor measurements), perception is supposed to be *active* and *adaptive*, where the robot needs to move and act to effectively place the sensors, just as we humans constantly turn our heads and change our position to gather desired information. Therefore, my longer research vision is to enable a *tight integration* of perception and action, particularly aiming at scenarios where a passive perception alone would fail. Fig. 3 illustrates a *tracking* example where the robot needs to actively integrate perception and action. In Fig. 3(a), the blue car (say our robot) is tracking the yellow car (the target), *e.g.,* by estimating the pose of the yellow car. Since the yellow car is observable, the perception system can obtain a confident and safe observation. In Fig. 3(b), however, a red car changes lane and occludes the yellow car from the blue car. Without proper action, the blue car would lose track of the yellow car (or wait until the red car moves out of the way). However, if the blue car also switches lane (Fig. 3(c)), then it can actively re-observe the yellow car and regain confidence. Meanwhile, the lane-changing action must respect safety constraints –the blue car can only switch to the right lane because the left lane is occupied by another green car– and this is where safe control and planning comes into play [5]. I believe a two-way *feedback* mechanism is necessary between perception and action, such that the failure of one system can call for help from the other. I plan to investigate proper *mathematical formulations* for such problems and design efficient, general-purpose, and certifiable algorithms with applications on real robotic platforms. Because safe control and planning becomes a critical component of safe active perception, I will also seek collaborations with control and planning researchers to tackle this challenge together, especially on designing safe control and planning policies under uncertain or misleading state estimations.

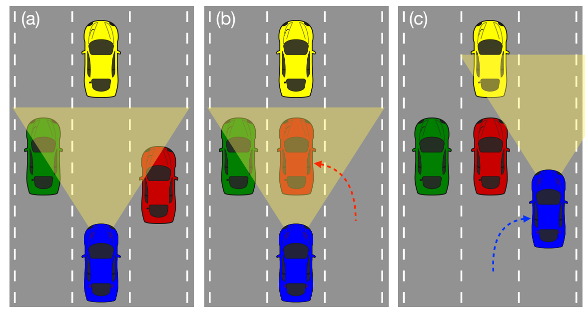

Figure 3: Safe *active* perception with a target-tracking example. (a) A blue car (*i.e.,* the robot) is tracking a yellow car with certified correctness. (b) A red car changes lane and occludes the yellow car, increasing the uncertainty in the perception system of the blue car. (c) By switching lane, while respecting safety constraints, the blue car obtains a certified observation of the yellow car.

Potential funding opportunities. My past research on certifiable perception has been funded by ARL, ONR, NSF, and Google. Given the wide interest in safe and trustworthy autonomy and its profound impact on society, I am interested in applying for both government and industrial funding, including DARPA (*e.g.,* DARPA Young Faculty Award is seeking applications for *Integrated Perception Learning and Control for Autonomous Robots*), NSF, ARL, ONR, Google, Amazon, etc. I also plan to establish a broad range of collaborations with roboticists, control and machine learning researchers, and applied mathematicians.

References

[1] Antonante, P., Tzoumas, V., **Yang, Heng**, Carlone, L.: Outlier-robust estimation: Hardness, minimally-tuned algorithms, and applications. IEEE Trans. Robotics (2021)
[2] Bajcsy, R.: Active perception. Proceedings of the IEEE **76**(8), 966–1005 (1988)
[3] Bandeira, A.S.: A note on probably certifiably correct algorithms. Comptes Rendus Mathematique **354**(3), 329–333 (2016)
[4] Bard, J.F.: Practical bilevel optimization: algorithms and applications, vol. 30. Springer Science & Business Media (2013)
[5] Brunke, L., Greeff, M., Hall, A.W., Yuan, Z., Zhou, S., Panerati, J., Schoellig, A.P.: Safe learning in robotics: From learning-based control to safe reinforcement learning. arXiv preprint arXiv:2108.06266 (2021)

[6] El Khadir, B., Lasserre, J.B., Sindhwani, V.: Piecewise-linear motion planning amidst static, moving, or morphing obstacles. In: IEEE Intl. Conf. on Robotics and Automation (ICRA), pp. 7802–7808. IEEE (2021)

[7] Fazlyab, M., Robey, A., Hassani, H., Morari, M., Pappas, G.J.: Efficient and accurate estimation of lipschitz constants for deep neural networks. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)

[8] Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)

[9] Hartley, R., Kahl, F.: Global optimization through rotation space search. Intl. J. of Computer Vision **82**(1), 64–79 (2009)

[10] IEEE Spectrum: Fatal tesla self-driving car crash reminds us that robots aren't perfect. `https://spectrum.ieee.org/cars-that-think/transportation/self-driving/fatal-tesla-autopilot-crash-reminds-us-that-robots-arent-perfect` (2017)

[11] Mathworks News and Stories: Trusting robots to navigate new spaces (2021). URL `https://www.mathworks.com/company/mathworks-stories/slam-mapping-algorithm-for-autonomous-vehicles.html`

[12] MIT News: Making self-driving cars safer through keener robot perception (2021). URL `https://news.mit.edu/2021/heng-yang-self-driving-cars-0916`. (Spotlight)

[13] Shi, J., **Yang, Heng**, Carlone, L.: Optimal Pose and Shape Estimation for Category-level 3D Object Perception. In: Robotics: Science and Systems (RSS) (2021). Best Paper Finalist.

[14] Shi, J., **Yang, Heng**, Carlone, L.: ROBIN: a Graph-Theoretic Approach to Reject Outliers in Robust Estimation using Invariants. In: IEEE Intl. Conf. on Robotics and Automation (ICRA) (2021)

[15] The Verge: Americans still don't trust self-driving cars (2020). URL `https://www.theverge.com/2020/5/19/21262576/self-driving-cars-poll-av-perception-trust-skepticism-pave`

[16] **Yang, Heng**, Antonante, P., Tzoumas, V., Carlone, L.: Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection. IEEE Robotics and Automation Letters **5**(2), 1127–1134 (2020). ICRA Best Paper in Robot Vision, RAL Best Paper Honorable Mention.

[17] **Yang, Heng**, Carlone, L.: A polynomial-time solution for robust registration with extreme outlier rates. In: Robotics: Science and Systems (RSS) (2019)

[18] **Yang, Heng**, Carlone, L.: A quaternion-based certifiably optimal solution to the Wahba problem with outliers. In: Intl. Conf. on Computer Vision (ICCV) (2019)

[19] **Yang, Heng**, Carlone, L.: In perfect shape: Certifiably optimal 3D shape reconstruction from 2D landmarks. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2020)

[20] **Yang, Heng**, Carlone, L.: One ring to rule them all: Certifiably robust geometric perception with outliers. In: Advances in Neural Information Processing Systems (NeurIPS) (2020)

[21] **Yang, Heng**, Carlone, L.: Certifiable outlier-robust geometric perception: Exact semidefinite relaxations and scalable global optimization. IEEE Trans. Pattern Anal. Machine Intell. (2021). Submitted. (arXiv)

[22] **Yang, Heng**, Dong, W., Carlone, L., Koltun, V.: Self-supervised Geometric Perception. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2021)

[23] **Yang, Heng**, Liang, L., Carlone, L., Toh, K.C.: An inexact projected gradient method with rounding and lifting by nonlinear programming for solving rank-one semidefinite relaxation of polynomial optimization. Mathematical Programming (2021). Submitted. (arXiv)

[24] **Yang, Heng**, Shi, J., Carlone, L.: TEASER: Fast and certifiable point cloud registration. IEEE Trans. Robotics (2020)