

<https://ieeetv.ieee.org/hidden/channels/ras/ras-interview-series-vincent-vanhoucke>

Vincent Vanhoucke

IEEE RAS Interviews

Q1: So if we could start with just a little bit of personal history. If you could tell us your full name, where you were born, if there are elements of your family story that you think are relevant to your career, feel free to share those. And then if you could just walk us through your academic education.

Vincent Vanhoucke: So my name is Vincent Vanhoucke. I was born in France and I have been, as a kid, I was based in Japanese cartoons of giant robots and was fascinated by the angle of technology, of trying to understand how to bring intelligence to the world of robotics. I have not always worked in robotics. I was initially working on different aspects of artificial intelligence. Started my career doing speech recognition, then switched to working on computer vision for a number of years. And that led to working on robotics problems, trying to fuse perception and actuation and understand how those two mesh together.

Q1: Could you tell us about your education and then also your work history?

Vincent Vanhoucke: Yeah, I was trained as an engineer back in France. I then went to do a master's degree at Stanford, where I started working on speech recognition, early days of AI. That was before the big AI revolution sort of changed things materially. I worked in a company doing speech recognition for a little while, worked on a startup that did computer vision, computer vision before it worked. So the company ended up being acquired by Google. And I've been at Google for a bit more than 16 years or alphabet at large, working on various aspects of AI at Google first. In the past year, I was part of Google DeepMind, and a lot of the work I did in robotics was done both at Google and Google DeepMind. Two months ago, I decided to make a career switch and I am now working at Waymo.

Q1: Congratulations. You've already answered one of my first questions, which is about first memories of robotics. And you mentioned Japanese popular culture. Is there anything specific? Is there a specific piece of culture or a specific robot that you remember as one of the first that you can tell us about?

Vincent Vanhoucke: Oh, the first robots, where there was all these cartoons that had giant robots that were saving the world, that also felt kind of threatening and dangerous at the same time as well. But used for good, so there was all the potential for something dangerous happening, but also the potential for good of humanity. The original classics. So in France, Grendizer was the name of the main cartoon that people really related to. I know that in the lore of Japanese giant robots, it's kind of an obscure cartoon that nobody really knows about. But for some reason, it resonated with the kids in France and became a huge popular phenomenon that I was born just at the right time to be caught into that phenomenon. And I think I'm a big fan from early on.

Q1: You mentioned that your training and your early work was in AI. And there's a kind of blurry boundary sometimes between AI and robotics. Perhaps you could talk a little bit about what that line means for you and what the first project that you thought of as a robotics project, as opposed to an AI project, might be.

Vincent Vanhoucke: Yeah, so I came to robotics through computer vision. And at the time, I want to say about eight years ago, we started to see computer vision getting to a level of performance that, we hadn't solved the problem, but it felt like within the next decade, computer vision would become more of a solved problem. And I started thinking really about, OK, what does it mean for the world? What are the consequences? Imagine you have a perfect computer vision system that can recognize all the objects in the environment and understand the context, understand the semantics of a scene, what could be really changed. And I started looking into robotics because it felt, as an outsider to the field at the time, it felt like the connections between artificial intelligence and robotics were very tenuous. There was not a lot of connections being made between the modern AI, deep learning, and the world of robotics. A lot of robotics seem to be developing along its own trajectory, focused on control theory, focused on a different way of approaching the problem that didn't have much to do with learning.

So the first sort of brush that I had with robotics is, we worked with some postdoc and professors at UC Berkeley. And we basically set up a robotics problem as if it were a computer vision problem. So, in computer vision, you take a lot of data and you train a neural network to recognize the objects in that data. And so we took, but it's very data intensive. It requires a lot of data collection. So we set up what we call the “Arm Farm” at the time, which was a collection of robots that would basically attempt to pick up objects in bins. And all that data was collected and used to train a big neural network that would learn over time how to grasp objects and how to pick them up and how to recognize them. At the time, the idea of directly connecting robots to the training of a neural network was not popular. There had been people working on it at one time or another, but it had kind of fallen out of favor as an approach to trying to do robotic control. But we got really good results. We got really interesting results, particularly when it comes to generalization.

So the idea that you could grasp any objects, including objects that you had not seen before, that you just perceived for the first time in front of you, that was not something that was even considered as a possibility. A lot of the grasping literature had really developed around, okay, let's assume you have a CAD model of the object. How do I grasp that object? Or let's assume you have the same object repeated multiple times, maybe in different configurations. How do you learn over time to grasp that one particular object? The idea that you could actually learn grasping, the concept of grasping in all generality and generalize over multiple objects was fairly new. And so that initial aha moment, if you will, made me, at the time I was leading a computer vision team at Google, and decided to completely pivot the entire effort towards robotics because it felt new, it felt fresh. It also felt like something that was neglected by the community and we could bring a different angle to the problem. And that was the birth of this robotics and AI team that we built at Google and we've been developing for the past eight years.

Q1: So it sounds like that early project connecting robots to neural networks set a research agenda and a development agenda for you for a while. Were there other kind of key research questions that you were interested in early in your career, or that have become important later?

Vincent Vanhoucke: Grasping was one. The other line of research that we started around that time is trying to understand if, instead of going, using traditional methods of physics-based control, we could learn our way through all of the control problems that exists in robotics. And we started, for example, with quadruped locomotion. A lot of that, at the time, was not using any form of machine learning or reinforcement learning and a lot of it was very analytical. And we did some of the first work bringing deep reinforcement learning to quadruped locomotion. All of that required us to build simulation software, to enable the robots to be simulated and collecting of data in simulation, figure out the sim-to-real transfer gap, how to take a simulated robot and match its characteristics well enough to a real world robot so that the policies that you learn in simulation transfer to the real world. So that ended up being another sort of deep research agenda that since a lot of work in the community has been done on learning for locomotion.

Q1: Yes. So as you have suggested, you are helping to set an agenda for the field that persists to this day. As you think back to your kind of early stages of this, what kinds of challenges did you face? Were they technical challenges? Were they challenges from the kind of institutions that may not have bought into this new way of thinking, funding or political or regulatory challenges? What were kind of early obstacles to your research?

Vincent Vanhoucke: I think the main obstacle was one of trying to get the two communities of the robotics world and machine learning AI world to speak a common language, and to have some degree of mutual understanding. In fact, in the very early days, it was very hard to get academic papers on machine learning and robotics even published at conferences, to the extent that we actually created our own conference around robotics and machine learning. It's called the Conference on Robot Learning, and it's been going on for eight years now and has been growing in a very healthy way. And as a community has formed around the themes of machine learning and robotics that was very fragmented before. So the community building aspect was one of the most rewarding part of that journey, bringing people together, giving them some common tooling, some common vocabulary to speak about the problem and also raising a little bit the awareness of the robotics problem in the machine learning community.

At the time, a lot of the computer vision agenda had been driven by the fantastic success on the Image Niche Challenges, on the co-detection data sets. A lot of the literature on computer vision ended up really focusing on the kind of problems that were shaped by those contests. And as a result, computer vision can lose track of the purpose of computer vision in the sense that it stopped at the pixel. Once you've annotated the pixel on an image, a lot of people consider the computer vision problem solved. When in robotics, when you're trying to do computer vision, it's for a purpose. You want to understand the world. You want to understand the 3D geometry of the things in front of you. You want to have a functional use of computer vision. So bringing back

into the ethos of what it is to do computer vision, this notion that you want to have a functional goal in mind and you're not just trying to annotate pixels for the purpose of finding cats in YouTube videos, but to actually use computer vision for a functional goal was one of the things that we were able to bring back. And a lot more of the research now is trying to figure out, how do we do 3D well? How do we use computer vision in the real world? And so both communities, I think, benefited from this kind of merger of goals and we were able to sort of shape a little bit the overall research agenda in very fruitful ways.

Q1: I mean, it's interesting hearing you talk about community building and institution building. I'm going to ask you a question that I think you've answered for yourself, but maybe you could think a little bit about that community, which is about funding. So you've been at Google, you've been in the corporate world, but maybe I'll ask you about other sources of funding for you. And then maybe if you could think about this new interstitial community. Does it have the funding structures that you think it ought to have?

Vincent Vanhoucke: I've been very lucky at Google and Alphabet in general that a lot of the founders of the company are big robotics fans. So both Sergey Brin and Larry Page have historically funded very big robotics efforts. We have Astro Teller at Google X, who has always been a supporter. Jeff Dean, the chief scientist of Google, is also a big fan of robotics. All that to say that funding has never really been a major issue for us within Alphabet. And we've been able to sort of really invest in ways that were unique and different. One of the advantages of Google has been that we have access to a lot of compute and a lot of software engineering talent that we were able to leverage into building a research agenda for robotics that was materially different from what you see in academia. We were really trying to not compete against academia in terms of the research teams we were approaching. We always had in mind, what can we bring to the table that nobody else can? And so it was a lot about scale, about doing experiments on lots of robots at the same time, using a lot of compute. And a lot of people looked at it and said, well, you're doing something that is very expensive and very impractical. And I always say, "yes, and that's the point." We are trying to explore the outer envelope of what's possible in a way that nobody else can, so that we can help inform maximally the community on what could be possible if we relax some of the constraints that we're under. I'm a lot less familiar with the funding world outside. As a result, we're in a little Silicon Valley bubble to some degree. And so we've never really had to really think too deeply about how the funding situation can impact the rest of the ecosystem.

Q1: My next question you've also answered aspects of, but maybe more holistically, as you think about how the field has changed over the course of your career, what do you think the most, are there significant changes that you haven't talked about that you see as—?

Vincent Vanhoucke: The big change is that, as I mentioned in the early years, we felt like the renegades bringing AI and machine learning to robotics, and there was kind of a pioneer sort of spirit to it. This is no longer the case. The community since has really embraced AI and machine

learning as one of the paths to success. Much of the conference this week has been about AI and robotics and a lot of people have really sort of raised to the challenge of how do we integrate AI and robotics with genuine interest. So that has changed materially and, not that AI has solved all robotics problems, not by any stretch. I think there is a very healthy skepticism of just throw AI at the problem and everything will solve itself. But at least the awareness and the interest to figure out all those thorny issues is there. I also think the level of ambition has risen. There is a lot more of a realization that the limits on what robots can do have somewhat been self-imposed by what people were studying and were focusing on. And now those limits are being expanded, and we can think about robots in the real world, robots in human-centered environment, interactive robots in a way that was very difficult to envision maybe even 10 years ago.

Q1: As you think about the future of the field, you talked about the synthesis of AI and robotics. Are there other kind of developments that you think are going to be increasingly significant in the future or that you hope will be research directions in the field?

Vincent Vanhoucke: One of the very interesting questions right now is, we have a number of technologies that are on the cusp, or are already being deployed in human-centered environments. I have recently moved to working at Waymo on self-driving cars. The self-driving problem, self-driving cars are kind of the first robots that are being deployed at scale in human environments. The new challenges there are really about scaling and how scaling robots in the real world really brings new problems to the table that we hadn't considered before and that are not necessarily the main subject of inquiry in academic circles. So safety, obviously, is the most important formal challenge, but also human-robot interaction. How do we want, as a society, to interact with those robots once they enter our human environment and how do we make the robots compatible with human society? A lot of those new challenges are becoming very measurable because, now we have actual robots being deployed and so we can study the problem, not in an artificial setting but in a live setting.

Q1: Given changes in both the conceptualization of the field but also these scaling issues, what do you think a student today entering the field as it is now, what would you recommend to them? How would they prepare? What kinds of strategies would they pursue to succeed in this new environment?

Vincent Vanhoucke: Things are getting a lot more interdisciplinary and I think it's a lot harder for an incoming student to just really specialize in a very narrow field. The breadth of things that you need to understand and be aware of has grown because essentially robotics is really a system level problem. All the interesting problems are in the integration of different things like perception and actuation. It's on the practical aspects of bringing robots to the real world. I think a lot of the problems that can be reasoned about in a lab and in a very specific field have been explored to death, and a lot of the interesting exploration now is how do we bring everything together and make it work. You can't just be a computer vision person. You have to be a computer vision person that understands enough of the actuation to be able to control a robot.

You can't just be a control person, because you have to understand how that interfaces with the perception aspects for instance.

Q1: And that kind of applies at the individual level. There's also the team, right, that would be competent to deal with safety issues, and HRI issues and regulation issues. In some ways, it seems like you're better prepared to take those on because of the environment you work in as opposed to say an academic department. How do you think the structure of teams or labs is going to change or needs to change in order to accommodate these scaling issues as these technologies move out into the world?

Vincent Vanhoucke: I think diversity is key. The challenge, I think, in academia is going to be to bring together people from different disciplines to solve the problems. And so, at the institution level, being able to have access to that diversity, having different labs working on different things but being able to be brought together to really sort of answer the questions in an interdisciplinary fashion is going to be key in the future. The academic setting is actually very well-suited to this in a number of ways. We have a lot to bring to the table in industry, but I think academia can bring different bent to those questions because there is a lot more potential for diverse disciplines to be brought together. You can have an entire lab that focuses on human-robot interaction for example. It's very rare that you have a very significant group of people interested in human-robot interaction in the same room in a non-academic setting. Bringing that kind of different sort of power and perspective to the problem in general I think is really a strength.

Q1: I'm going to ask this next question in a very particular way which may not be the right way to ask the question, but I'm going to ask you about generative AI, and the various ways in which you see that relevant to researching the field more broadly. I know that encompasses aspects of machine learning, but generative AI is a kind of language we're using, so what are your thoughts on that set of technology?

Vincent Vanhoucke: So at Google we had a sort of a early peek at the rise of generative AI before it entered the public sphere and, when we started seeing the glimmers of you know large language models and large multi-model models coming, we immediately started asking ourselves, what does it mean for robotics? And, in fact, much of our program for the past two or three years has been driven by the intersection of generative AI and robotics. So the first projects we did were to bring generative AI into the planning problem, and that gave us some early successes where you could take a very high-level task, like make me a coffee and use a language model to decompose it into smaller tasks that the robots could perform in a way that was compatible with the robots affordances and things like this. And that foundation gave us a lot of confidence to really pursue what happens when you take every component of the robots and replace it with some form of generative AI. So we've explored taking the perception system and replacing it with a vision language model, taking the actuation and replacing it with a language model that produces code that corresponds to the controller's code and, or that produces

trajectories that the robot can then execute. And bringing that together into a large sort of a robotics foundation model that can take, as input, images, vision, can take, as an input, goals that are expressed in natural language, and execute that on the robots at scale in a way that is compatible with being human-centered, being safe, being competent, has been really a fantastic journey. And I think there is a lot more that we can do with generative AI as a kind of a new substrate on which to build robot intelligence. So I've been really advocating for a lot more research to be taking place in this place.

Q1: As you think about the overall trajectory of your career, what would you say are your most significant contributions to the field? And a related question, not the same question: What are you most proud of?

Vincent Vanhoucke: I'm proud of having had a small hand in bringing machine learning and robotics on the map, putting them on the map. I think there was really a gap there that was, that needed, not only sort of a technical push to sort of validate that machine learning centric approaches to robotics were relevant and valid and had interesting new things to bring to the table, but also a social component, which was getting to bring roboticists and machine learning people in the same room, and getting them to talk and share and actually sort of understand each other better. I'm most proud of that because I think this community that we've created around the conference on robot learning has been just fantastic. It's been a very, very stimulating environment and bringing people with very different perspectives together, and the amount of new work that has come out of that community has been fascinating. In fact, the conference is, if you look at its impact, it's only second to ECRA at this point amongst all the robotics conferences, and it's been growing in in size and impact for the past few years.

Q1: It's been interesting how often community building has come up as the aspect of people's work that they are most proud of, and think most significant in the long term. As we near the end of our time together, are there topics that we haven't talked about yet that you would like to talk about, or aspects of your career that we have not yet covered?

Vincent Vanhoucke: The pitch I gave this week to the community was to really start thinking about robotics and value alignment with humans. I know it sounds maybe like a very philosophical question, but what has been fascinating to me over the last few months and years is that this is becoming a topic we can study and, because all the pieces are now available for connecting general human values to actual outcomes in robots, actually doing things differently if you give them a position on what is the right thing to do, and what is the wrong thing to do. So the, this idea that we can actually sort of think about values and human preferences and human compatibility in a way that is something we can study, something we can measure, something we can evaluate, and have grounded discussions about, I think is something that I would love to raise more awareness in the field about. Because I think, you know, our robots are soon to make it really out of the factory floor in a big way, and we need to have a perspective on what it means

to live in a human-centered environment that has autonomous systems like robots evolving in it, and what we want this future to look like for humanity.