Berlin Robotics Summit 2011

Speakers' Corner Presentation: The Singularity

Andrew Davison, Imperial College London

1 Introduction

The aim of this presentation was to look even farther forward than in most of the other sessions, and test the Summit attendees' views with regard to the 'Technological Singularity' idea, recently much discussed by 'futurists', technologists, science fiction writers and scientists but still on the fringes of mainstream scientific debate.

I started the presentation by asking for a show of hands on the following questions:

1. Who knows what The Singularity is?
2. Who thinks it might happen during our lifetimes?

Probably around half of the attendees had at least heard of the idea; and at the start of the discussion I don't think any hands went up in answer to 'who thinks it might happen during our lifetimes?'. I then explained the Singularity concept, briefly presented the main arguments for it, and we had some discussion. At the end I revisited the second question, and there were three people (including me) out of about eighteen agreeing that it 'might' happen during our lifetimes.

A question I had wanted to get to by the end of the session was given that it might happen, what should we do about it? Should foresight of this possibility change the way that we do robotics research? However it was not appropriate to move onto this question when the majority of those present were very sceptical about the whole concept. I must admit that overall I was surprised that such a big concept seemed to be much less on the radar of the world's top roboticists than I might have expected.

2 The Singularity

The Singularity is usually defined as an event that could happen in the future when the accelerating progress of technology becomes so rapid that human life is irrevocably changed (for the better or worse). Perhaps the most significant event associated with the Singularity, and certainly the one most relevant to roboticists, is the arrival of human-level artificial intelligence, and then very soon afterwards vastly super-human AI. It seems clear to me at least than human life would not continue in anything like the current way once super-human AI exists. A super-human AI would be able to develop concepts and technologies we cannot understand; would continually self-improve; and for me by definition it would be impossible to control. Various futurists have predicted what might happen to humanity if this were to happen. Some of the possible scenarios are catastrophic of course. The most optimistic scenarios, predicted and hoped for by Kurzweil and others, forsee humans gradually merging completely with their technology, on a path where there might first be great advances in medicine with aspects like longevity expansion, body repair by nanorobots, and brain-computer interfaces. The final destination of this predicted path is 'uploading', where expanded human minds are finally transferred from biological brains to another computing substrate.

This, of course, is far out thinking... and as a normal cynical scientist I can understand reluctance to take it seriously. But having thought quite hard about this idea for several years I have not yet found a strong counter-argument which persuades me that it is not possible that something like this will happen in the next few decades. In many ways I would like to hear one! My belief in the idea that the rate of change in technology is accelerating is informed not just by the views of futurists, but strongly by my observation of what is happening in my own field of real-time computer vision, where the new computing power, algorithms and devices we can take advantage of are year on year leading to increasingly staggering capabilities. If we might really be heading towards a Singularity, foresight of this should surely shape everything about the way we are doing robotics research. Let us look at the arguments.

2.1 Accelerating Change

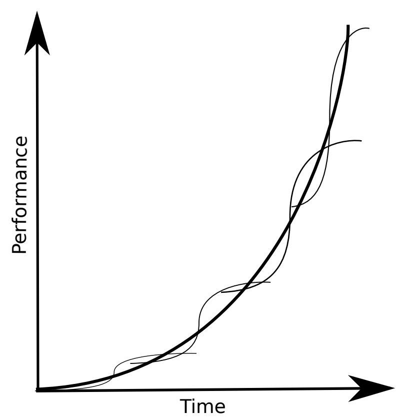

Figure 1: A sequence of overlapping S-curved paradigms, each with steeper progress than the last, leading to overall exponential progress.

The main argument put forward in favour of the Singularity is that the progress of technology has historically followed a law of accelerating change, well described by an exponential curve as a function of time, and that his shows no sign of slowing down. The obvious modern example of this is 'Moore's Law', originating in a specific observation over 40 years ago about transistor density by Intel founder Gordon Moore, but commonly used to describe the continuing 'doubling every 18 months' exponential performance improvement in computer processors and related technology. It is easy to forget how staggering it is that we have moved comfortably, even in my own experience, through talking about computers in the 'kilo', 'mega', 'giga' and now 'tera' eras.

The idea of the Singularity relies on similar laws describing not just computing performance, but the general progress of technology. The key reason for the rate of progress to keep increasing is that each new generation of technology has the benefit of the last, best so far generation which can be used to develop it. Different areas of science and the whole technological economy continually feed back on each other so that every advance makes progress easier in other fields. We are familiar for instance with the idea that scientific supercomputing has revolutionalised many areas of science and engineering through advanced simulation; or that the internet, online research resources and code repositories make it much easier to learn or put together projects in previously inaccessible domains; advances in those areas in turn feed back into the better concepts, designs, methods, factories and robots needed to make the next generation of processors. While some areas of technology do not currently seem to be experiencing exponential progress (e.g. transport speeds?), it can be argued that as more and more industries become information technology and software-dominated they can experience the full benefit of accelerating change — for instance this might happen in manufacturing if 3D printing or nano-technology based self-assembly come to full fruition and an object is just another software file.

Accelerating change is not always smooth, but has been described as taking place via a set of overlapping 'S-curves', each corresponding to a *paradigm shift* (Figure 1). A new technology appears; takes some time to break through and overtake the previous model, but then goes through rapid take-up, investment and improvement before eventually reaching the end of its usefulness as some physical limit is approached and slows down; at this point the next technology waiting in the wings, and probably inspired and motivated by observation and desire to beat the previous one, takes over. Even within the narrow domain of digital computer processors, we can see this happening; the increases in performance of single-core CPUs are slowing down, but massively parallel processors such as GPUs, now easily programmable with tools like CUDA, have taken over the exponential progress in terms of the most processing capacity obtainable for commodity prices and are leaving CPUs far behind.

Singularity advocates argue that paradigm shifts have been occuring at an accelerating rate not just in the recent past but right back through human history and beyond into evolutional times. Consider the follow sequence of advances: origin of life; cells; reptiles; primates; upright primates; homo-sapiens; art; agriculture; city-states; writing; printing; industry; electricity; computer; internet; smartphone... The 'time to next event' vs. 'time' plot of these events on a log/log scale is approximately linear. Could we argue that each of these paradigm shifts has similar importance, and just that culture and technology have overtaken biological evolution as the main agents for change in the world (because they happen so much faster)?

2.2 Super-Human AI

If we accept the 'Strong AI Hypothesis', that general intelligence of the type humans display is achieved by an algorithmic process which can in principle be simulated by an artificial digital processor of sufficient speed and memory, then a continuing exponential increase in computer technology implies that we will at some point all have computers on our desks (and not many years later in our pockets) which are as powerful in raw processing and storage terms as the human brain. Rejecting the Strong AI Hypothesis implies either a spiritual/mystical view that the human mind is more then the brain, or a Penrosian type belief that deep in the brain there might be important processes which are not described by current physics. I think that most modern roboticists are inclined to accept that the brain can in principle be simulated by a Turing machine-like digital computer.

So if we do accept that, then the question of course is how powerful would that computer have to be that could simulate everything a human brain does with enough fidelity to produce the same kind of 'generally intelligent' behaviour, and how soon will we have it? I guess that depends on how fiddly you think things are, and how closely you would have to model what each element of the brain does — at the molecular or chemical level, or just at the gross level of unit connections and dynamics. I don't think anyone knows the answer to that precisely yet, but my guess is the latter, that it is the general pattern of connections and signal types that is important. And so if you believe that you have to figure out what level of processing you'd need to simulate that. I've heard recent estimates like 10^{15} calculations per second. Interestingly in his 1950 paper 'Computing Machinery and Intelligence' at the dawn of AI, Turing considered that memory capacity, rather than processing speed, was the main limiting factor for AI, and that estimated that 10^{15} bits would be the memory capacity needed to model a human brain. These figures of 10^{15}, whether for processing operations per second or storage capacity, must have seemed stupendous to Turing but for us in 2011 these are 'around the corner' figures for desktop computing, coming in 10–20 years surely without much doubt by projecting current curves. But a very significant point is that even if these figures are very wrong, like a million times off, then Moore's Law means this this only makes a difference of a few years. I agree that if it's more like 10^{100} and that you'd need to simulate at a much finer level then we are still miles away!

I don't think that even the idea that we will fairly soon have computers available with equivalent computational capacity to the human brain was controversial at the summit. The doubts were naturally about whether we will have any suitable software to run on these devices to achieve human-like 'Artificial General Intelligence' (AGI). Where is this fiendishly complicated software going to come from? It might come from a continuation of current AI research and lots of components of the type I develop in computer vision for instance, all joined up with machine learning, some kind of embodiment and training environment, and that something amazing and open-ended happens when you reach the right scale. We have already seen with phenomena like Wikipedia how quickly a vast amount of knowledge can quickly be assembled by relatively uncoordinated means and made easily accessible. Of course this information is not yet in AI-understandable form, though there are efforts in this direction. But we see how current communications and storage capacity might form the way for a set of self-learning and communicating AI agents/robots to build up and share a vast store of 'general knowledge'.

The time, probably near in the future, when commonly available computers achieve the capacity to 'in principle' simulate the human brain is a particularly important event in my opinion. Although we still see many weaknesses in current AI and robotics systems, we have never yet run those systems on computers with the processing and memory capacity that we estimate the human brain has; but we will be able to do so soon. Some 'magic', emergent, behaviour may happen with the current techniques we have for machine learning, vision, etc. when applied at this scale; or maybe not. It may be that AI still needs serious re-invention, and that we are missing vital algorthmic pieces in the software which might be needed for very general reasoning. But surely we should not dismiss candidate algorithms for AGI (Artificial General Intelligence) until we have tested them at this scale.

Another completely different approach to the software problem might be much more like reverse-engineered biology, where continual improvements in brain scanning give us eventually the ability to see how a whole brain is wired together at the neuronal level and therefore to simulate it functionally on a computer of sufficient capacity. This architecture could then possibly be reverse engineered, improved and expanded.

An important point is that even if your software is not very good, perhaps vastly inefficient compared to what the brain is doing, exponential growth in computing might mean that this just means the need to wait a few more years for the right processor to run that software on.

Another point in favour of the singularity that I find convincing is that the human brain, with its capabilities which are dramatically more advanced than those of any other animal, evolved really extremely rapidly to separate us from our common ancestors with the other apes. Surely whatever the human brain does that is special is one particular evolutionary trick (extra piece of code) just massively scaled up and repeated. Personally I am with others in thinking that what we can do that other animals can't is all about a massive, well-organised memory store which we are able to use constantly for prediction.

3 Reactions and Discussion

This is clearly a controversial topic, and as I said in the Introduction, after explaining and discussing these ideas with the group, I took a final show of hands where there were three of us with the opinion that the Singularity 'might' happen within our lifetimes. To repeat, natural cynic though I am, I find it hard to find any strong argument that the concept doesn't at least merit very serious thought.

There is a tendency I think for scientists working deep in a discipline to see all the local difficulties of that area and to extrapolate those into the future in a kind of 'scientific pessimism'. For instance, much of the discussion in this summit was about robot manipulation problems, with the perception and planning challenges they present, which is where I think we were mostly agreed that the main thrust of exciting robotics research will be focused over the next few years. There were views expressed that the very high dimensionality of the reasoning needed for manipulation or learning about manipulation makes it doubtful about whether we are making any progress at all on these problems at the moment, and also doubtful whether we would be likely to in the 'seemingly short' next 20–30 years.

But history teaches us that our 'intuitive linear' view of progress is usually overtaken by historical exponential growth. When a new technology appears we often overestimate the effect that is will have in the near future; but then dramatically underestimate the long-term effect. If half-way through the time allotted to an information processing project we have only completed 1% of the planned goals, this might actually be right on schedule for full completion under an accelerating change regime (the human genome sequencing project was an example of one about which there was much pessimisim based on early slow progress, but then was suddenly and surprisingly completed ahead of schedule).

Let's remember that we have only had computers *at all* for 60–70 years; and only had them at home for around 30; and look at how much has been achieved. People say that AI hasn't come anywhere, but AI is just the moving target of things computers can't yet do; there are plenty of problems that computers couldn't do once where it seemed like 'real intelligence' was needed (Being a travel agent? Vacuuming a floor? Route planning and even autonomous driving? Face recognition?) which are now just considered computation. Of course AI has achieved a lot already.