Robot applications might be the key to understanding the human mind

Zygmunt Pizlo

Psychology, Purdue

Collaborators: Yunfeng Li, Tadamasa Sawada, Robert M. Steinman

The starting point for this discussion:

- A half century ago the Artificial Intelligence (AI) community moved away from Cognitive Psychology and went on to develop independently, so
- it should not come as a surprise that AI researchers actually know next to nothing about Human Intelligence (HI), unless you count anecdotes in our popular press or some introspections.
- In fact, we psychologists, don’t know much more about HI, either, because we also have not studied it. Human Intelligence is hard to study.

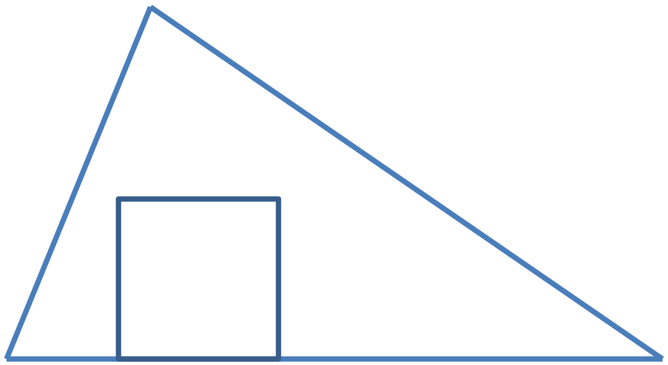
Studying HI is hard.

- HI is a difficult subject because humans are very smart:
  - If, for example, you want to study the creative thinking of such intellectual giants as Einstein, you may actually have to be as smart as he was. At the very least, you must really understand exactly what he accomplished.
  - Quite possibly, this requirement may explain why the best treatment, so far, of mathematical problem solving can be found in books written by George Polya, who was a mathematics, not a psychology, professor.
- HI is so hard that one can only wonder whether any robot can ever learn how to solve math problems or learn how to formulate new theories in physics as smart people often can?

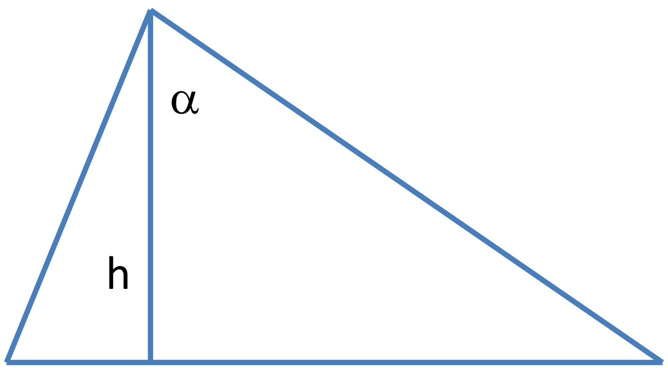
What kind of problems must be solved

- We still don’t know how math students actually solve problems in geometry. This shows that psychologists do not have a theory of HI (at least in this case).
- Existing robots, unlike math students, *cannot* solve such problems. This shows that AI researchers do not have a theory for solving such problems, either.
  - The current AI algorithms can perform fast searches, but they are not creative.
- If we had a hunch (hypothesis) about how students solve problems in geometry, we could test it by implementing this hypothesis in a robot…

Examples of problems in geometry:



1. Inscribe a square (find its position and size) inside a triangle.



2. Construct a triangle given its perimeter, one altitude and one angle.

These problems are not easy – I cannot see how IBM's Watson, or Deep Blue, or the algorithm for trading stocks could even begin to solve them, and note that 2D geometry is *much* easier than 3D!

Are we at a Dawn or near Doom?

- So, good math students are obviously very smart, but we have absolutely no idea about how their minds work, nor how to design a robot that could emulate these students' performance.
- So, given these facts, how can any AI researcher be prepared, or even qualified, to say when, if ever, robots will become smarter than human beings? They are far behind us now.
- Sadly, psychologists can only help a little because we have just barely scratched the surface of HI ourselves.

Solving a problem *vs*. finding one.

- Many problems we humans regularly solve seem to be impossible, *e.g.*, recovering a 3D shape from a single 2D image.
- How will a robot know which problems to attempt to solve, and which ones are insoluble?
- Also, once a robot does solve a problem, how will it know whether to keep trying to improve its solution or when to stop?
- At this point, the only way to answer such questions is to compare a robot's performance to a human being's.
- Note well that we must have a theory of HI to do this.

The role of computational and robotic modeling

- The human mind is a complex non-linear system. It follows that piecemeal studies, using simple, artificial inputs, may tell us almost nothing about how the mind handles complex, natural cases.
- Emulating the human mind in computational and robotic models is arguably the best, perhaps even the only way, to understand and explain the human mind.
  - This idea is neither new nor mine.

Miller, Galanter & Pribram (1960)

“The creation of a model is proof of the clarity of the vision. If you understand how a thing works well enough to build your own, then your understanding must be nearly perfect” (p. 46)

Richard Feynman’s (1988) conjecture:

“What I cannot create I don’t understand.”

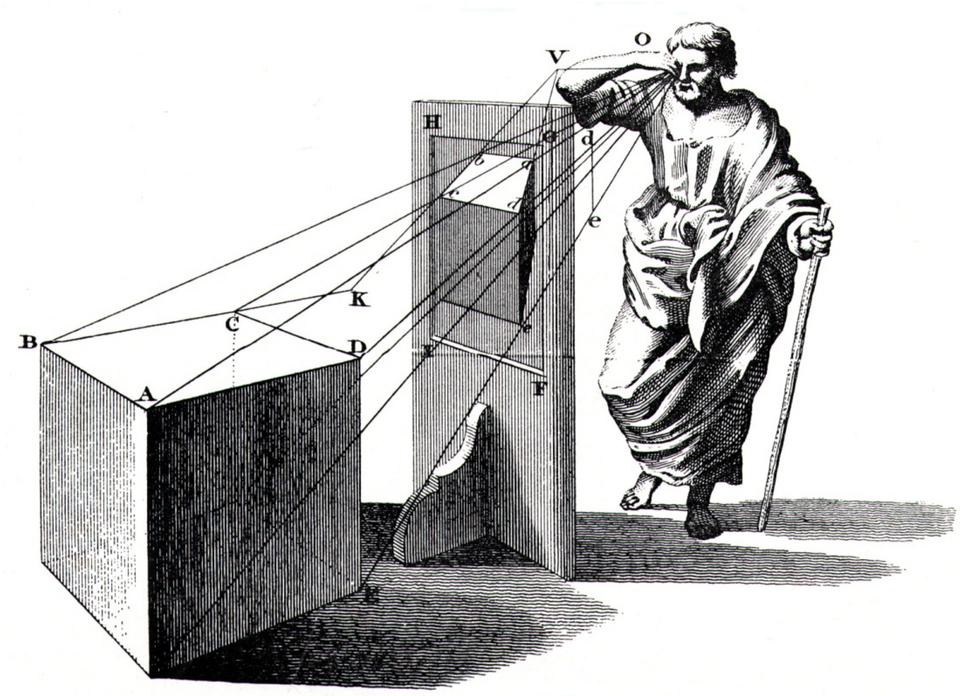
AI *vs.* HI

- This relation must be an active two-way street.
- The intelligence of robots produced by AI must be evaluated by comparing it to what we know about HI.
- But if we want to make progress in our understanding of HI, we need to emulate HI in a robot.

With this said straight out, I will now describe 2 problems humans solve by using their special kind of intelligence

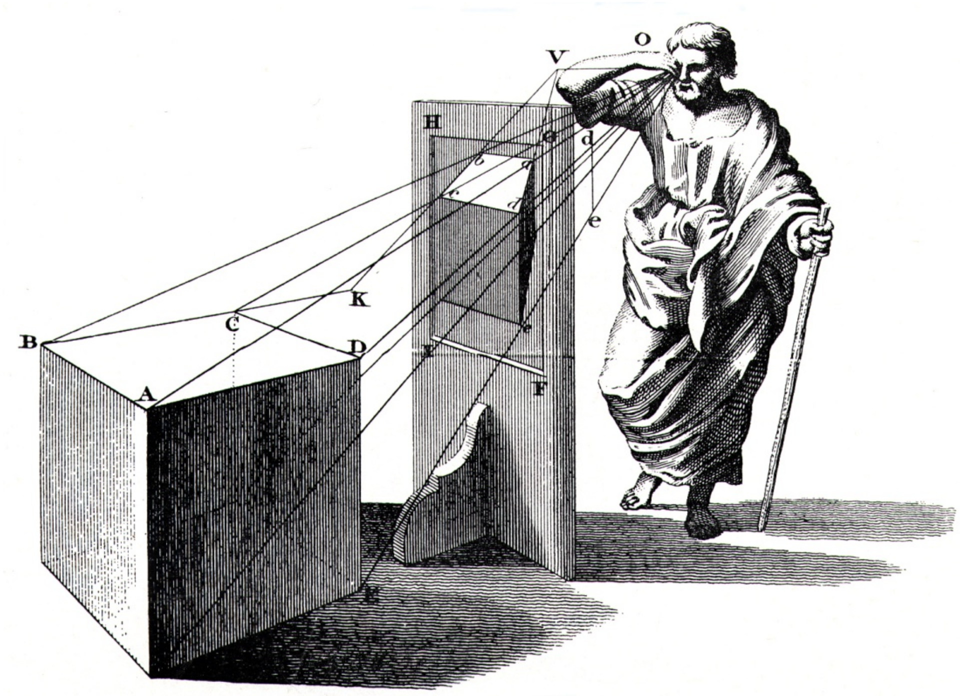
- Recovering the 3D shapes of objects from a single 2D image.
- Finding objects in images and recovering the Euclidean structure of 3D scenes.
- Note that computer vision has traditionally been considered the most advanced specialty in AI, so, if computer vision lags behind human vision, AI lags behind HI.

A 3D percept from a single 2D image



The world is 3D, the retinal image is 2D, and the percept is 3D. There are always infinitely many 3D interpretations that can produce any particular 2D image. How does the brain choose one 3D percept, not to mention how it manages to choose the correct one? Could it be familiarity?

A 3D percept from a single 2D image



The most symmetrical 3D interpretation of the 2D image of a cube, is a cube. So, with a cube, symmetry led to a veridical interpretation. Technically, we assumed that the visual system minimized a cost function that evaluated departures from 3D symmetry.

[Ames chair demo]

Can symmetry work more generally, *i.e.*, with objects other than a cube or a chair?

Symmetry is ubiquitous in our natural environment

- Animals are symmetrical because of the way they move.
- Plants are symmetrical because of the way they grow.
- Man-made objects are symmetrical because of the function they serve.
- It follows that:

Symmetry is a *natural* shape *prior* (*a priori constraint*, *or predilection*) because it is ubiquitous in nature and because it makes it possible to recover 3D shapes of objects from 2D images.

Symmetry is not the only prior used by the human visual system, but it is the most important one.

The Nature of Priors

- Note well that *a priori* constraints applied to a few *abstract* characteristics of 3D objects are much more effective than priors learned from experience with *concrete* objects.
  - How could you actually learn priors for *all* objects "out there"?
- Note also that abstract constraints can be applied in *real-time* and they can be applied *to unfamiliar objects*.
- Furthermore, *a priori* constraints, such as symmetry, are mathematical concepts: they need not be derived from experience and need not be updated. We are most likely to be born with them. Symmetry is prominent in everything that lives, *e.g*, your brain and DNA.

Three additional effective shape priors

- Maximal planarity of contours
- Maximal 3D compactness (arg max V^2/S^3)
- Excluding degenerate views

Again, note that *all* of these priors apply to abstract characteristics of 3D objects, not to the objects themselves.

These priors do not have to be learned or updated through experience.

Now, let's look at a demo illustrating how well our model, based on these four constraints, can recover 3D synthetic and real shapes from a single 2D orthographic image…

Recovering 3D shapes

Can symmetry be used to recover not-so-symmetrical shapes, too?

- Sure

- I just showed you how the human mind (its visual part) solves what seemed to be an insoluble problem, namely, recovering a 3D shape from a single 2D image.
- There is good reason to believe that the computer vision community would have never even tried to solve this “impossible” problem, if they did not have an “existence proof” of biological vision.

What do we know about 3D vision in a robot?

- The last 20 years of computer and robot vision assumed that *all* visual interpretations are only *2*-dimensional (2D) because camera images are 2D.
- The AI community knows that there is a 3D symmetry prior, but insists on *not* using it because they want their robots to be more general than we humans are. Robots should not be restricted in any way by what might prove to be “unnecessary” priors (constraints).
- But without 3D priors, there cannot be 3D vision as we know it.
- ***There simply isn’t a more intelligent way to see than the way that we humans do it.***

We made our robot emulate human 3D vision

A robot (Čapek), equipped with a stereoscopic camera, is tasked to recover the 3D scene in front of it, and then plan and execute navigation tasks within it.

The first step is to find 3D objects in 2D images

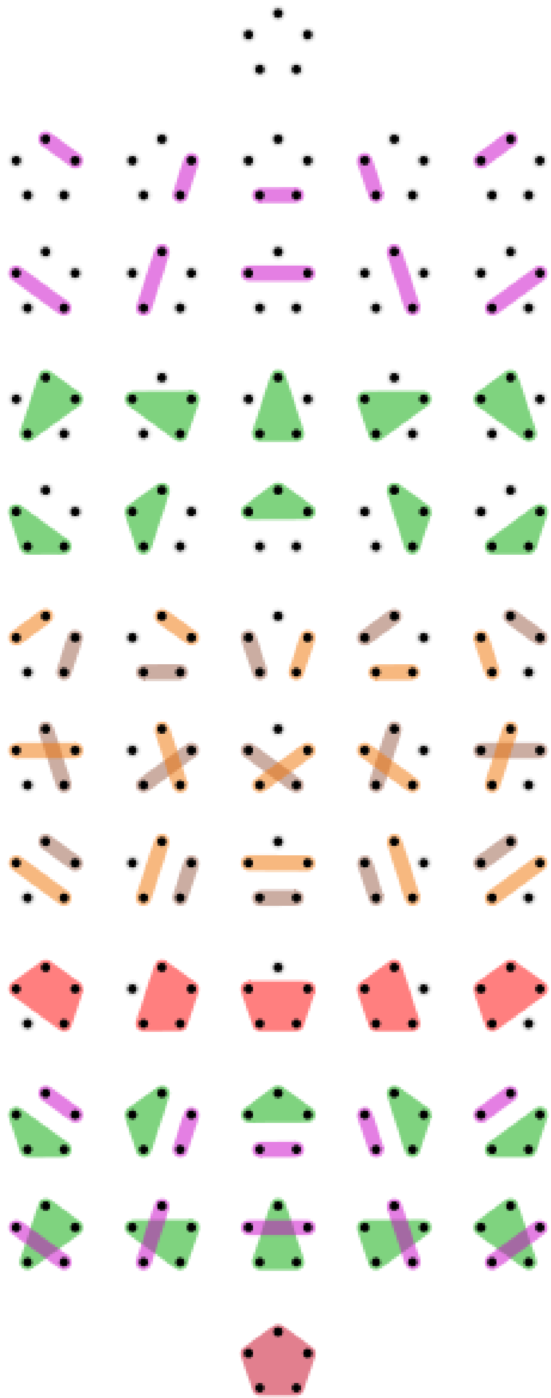
- The human retina has and can use 6 million cones to provide visual input under normal viewing conditions.
- This photosensitive surface is similar to what we use in contemporary camera images (2k by 3k pixels).
- Making sense of such an image boils down to analyzing *partitions* of the image, so interpreting an image is a combinatorial optimization problem.
- In Psychology, this task is called the Figure-Ground Organization (FGO) problem – an example from real life is shown on the next slide.

FGO in robot vision

- At this point, real images *cannot* be handled reliably by *any* of the existing robot vision systems.
- Typically, robots are first trained on thousands of 2D images, after which they try to detect what they have learned.
- The robot’s performance is always poor, very far from perfect, and it almost never generalizes well to unfamiliar categories of objects, if it generalizes at all.
- Besides, because the robot’s visual representation of the 3D world is 2D, the robot would surely claim that the Earth is flat if you let it speak its mind. This might seem a bit odd in the 21st Century. There is no reason to assume that these robots will become smarter than we are anytime soon.

FGO without any training

- If we exclude training on familiar objects and scenes, the visual system *must* use some abstract rules to choose the unique and correct organization of the retinal or a camera image.
- A brute force approach would call for examining all organizations (partitions) and choosing a single one based on some well-crafted criterion.
- But first, we must know how many partitions are there in order to evaluate the feasibility of this kind of approach. The next slide shows what it entails.



Take a set with only 5 elements.

There are 52 partitions (organizations) of this set.

The number of partitions of a set of n elements is called a Bell number (B_n).

The Bell number for cameras with more than 5 pixels

- The number of partitions of 10 pixels is about 100,000.
- The number of partitions of 100 pixels is larger than 10^{100}.
- How much vision can you have with 100 pixels (receptors)?
- There are *only* 100 receptors in the central $3' \times 3'$ region of your retina (the size of a finger nail viewed from a 10 meter distance).
  - The number of partitions within this small region is already larger than the number of atoms in the Universe.
  - You cannot evaluate even a tiny portion of these partitions (interpretations) within any reasonable amount of time.

So, can the FGO problem actually be solved?

- The human retina actually has 6,000,000, not 100, cones.
- So, the number of partitions of the retinal image is greater than $10^{6,000,000}$. If you evaluated one billion partitions each second and started at the Big Bang, as of today, you would have checked only 10^{26} partitions.
  - FGO is more difficult than playing chess.
  - Note that it would not help much if your computations were a billion times faster than what I assumed.
- So, the very first step towards understanding the sensory input to the visual system presents what looks like an *impossible* problem.
- Effective biological vision is the only proof we have that this problem can be solved. The next few slides will illustrate some important and unexpected characteristics of this remarkable solution.

Goal

- Detect the individual objects in a 3D scene, including their sizes and orientations.

- Identify the 2D regions in the image that represent individual objects.

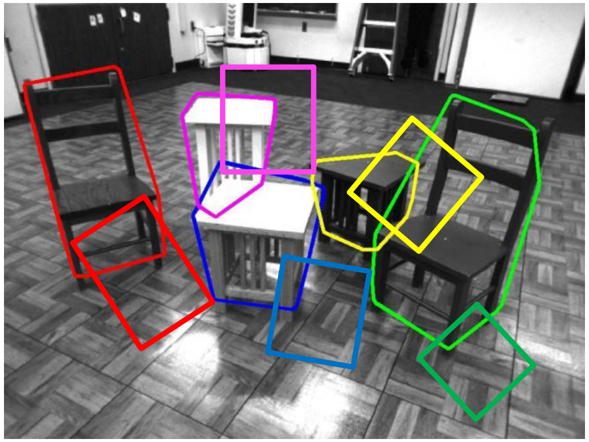
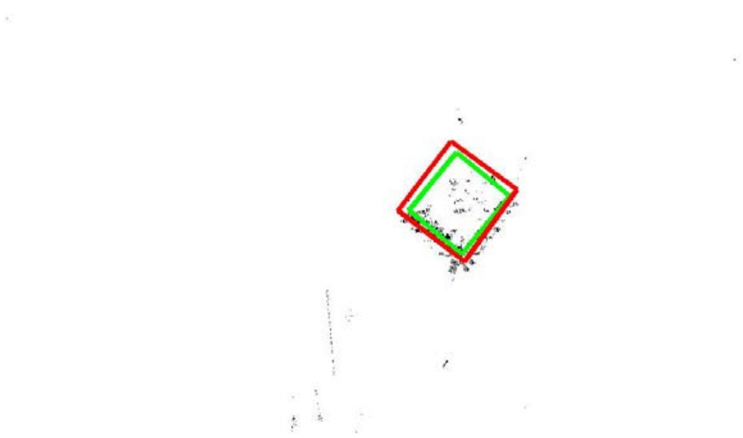
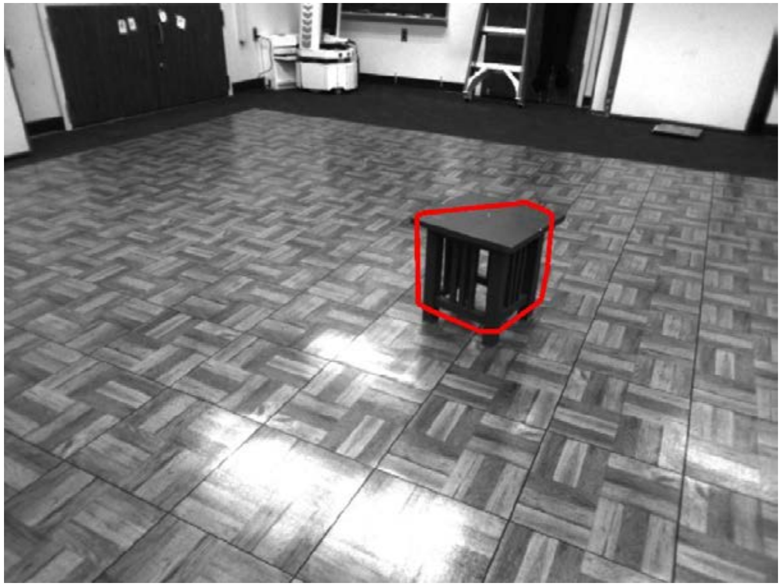
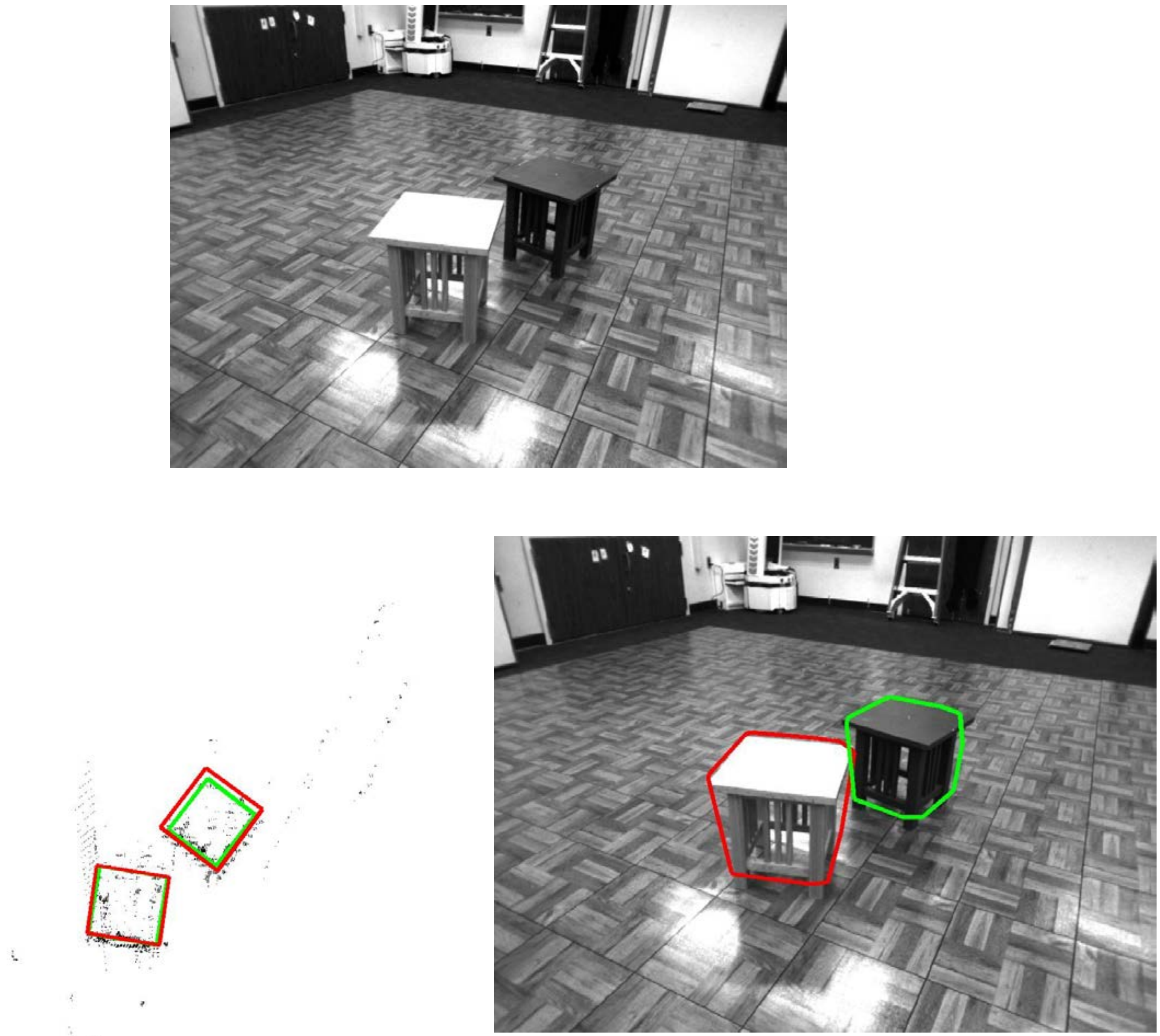


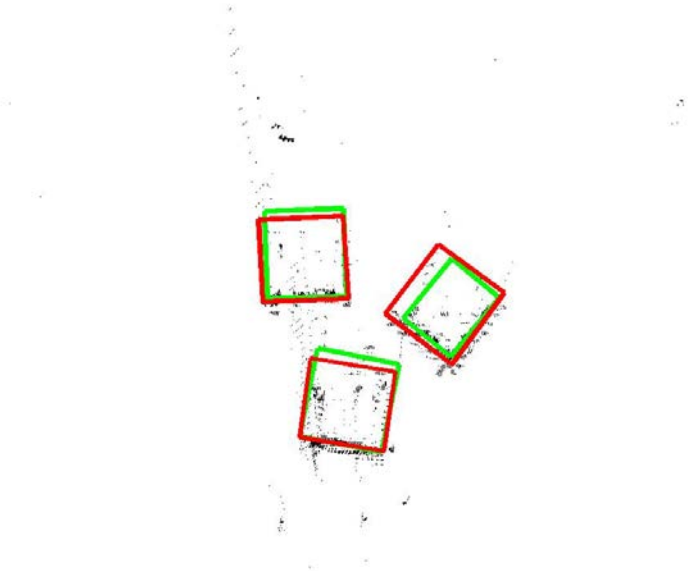
Figure-ground organization (FGO)

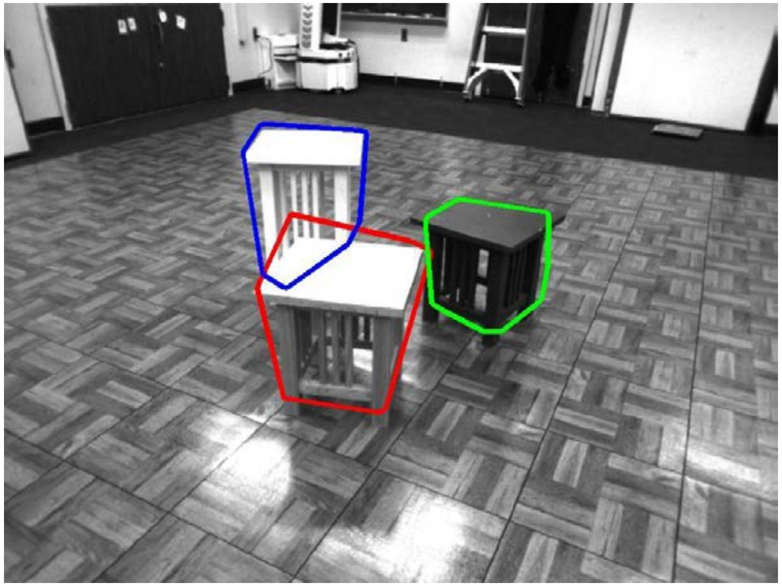
- Humans have no difficulty in finding unfamiliar objects in novel scenes.
- So, our goal is to have Čapek do the same – it will solve the FGO problem without using either familiarity or learning.
- It will do it by using only *intelligent* algorithms.





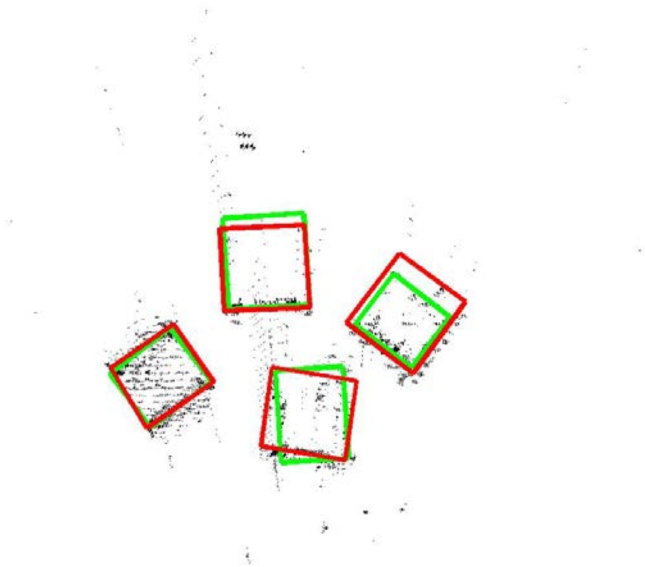


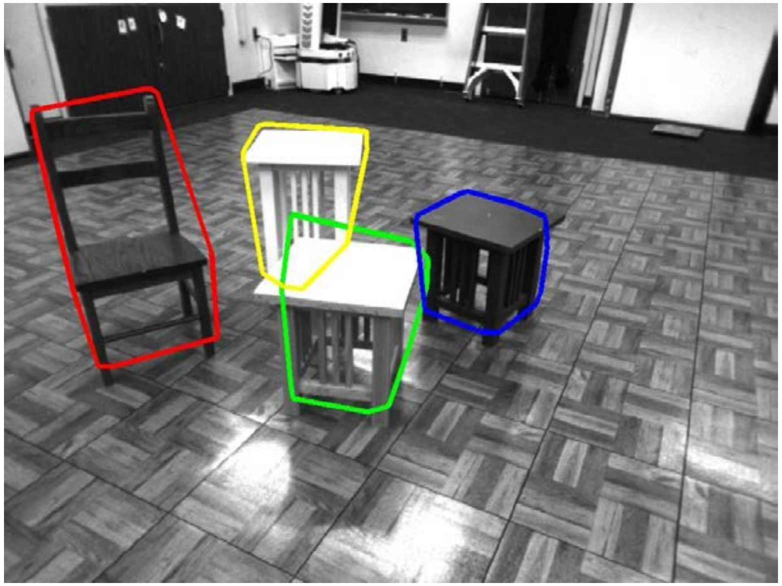


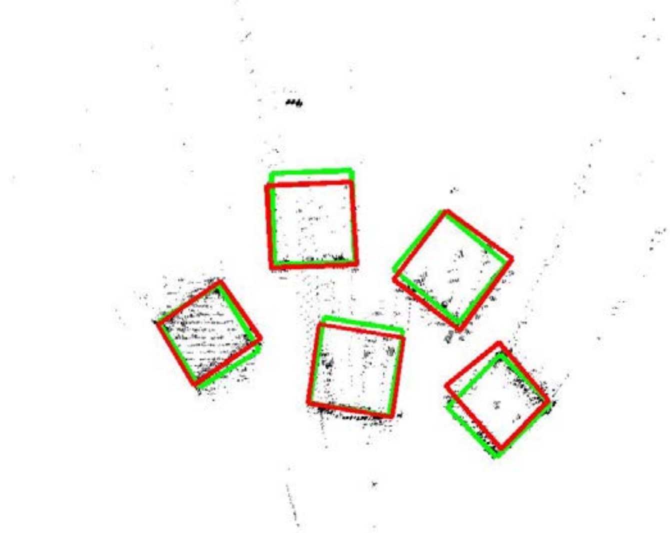


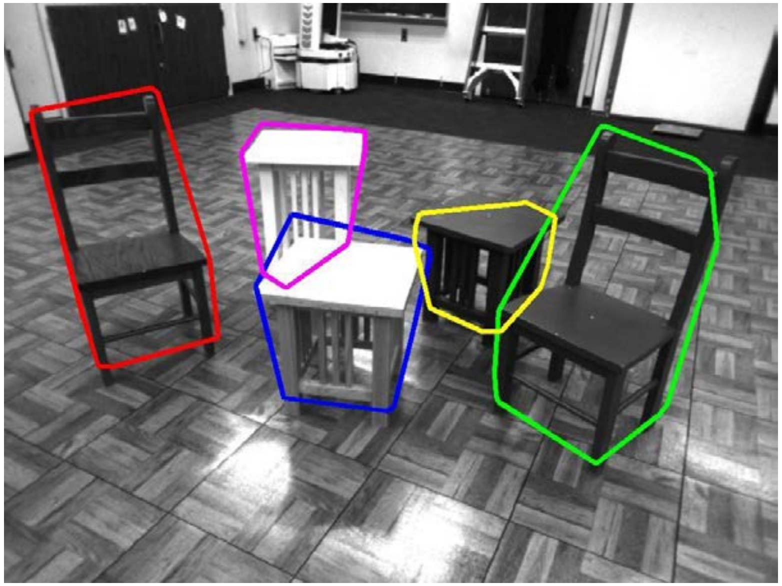
Seeing behind objects

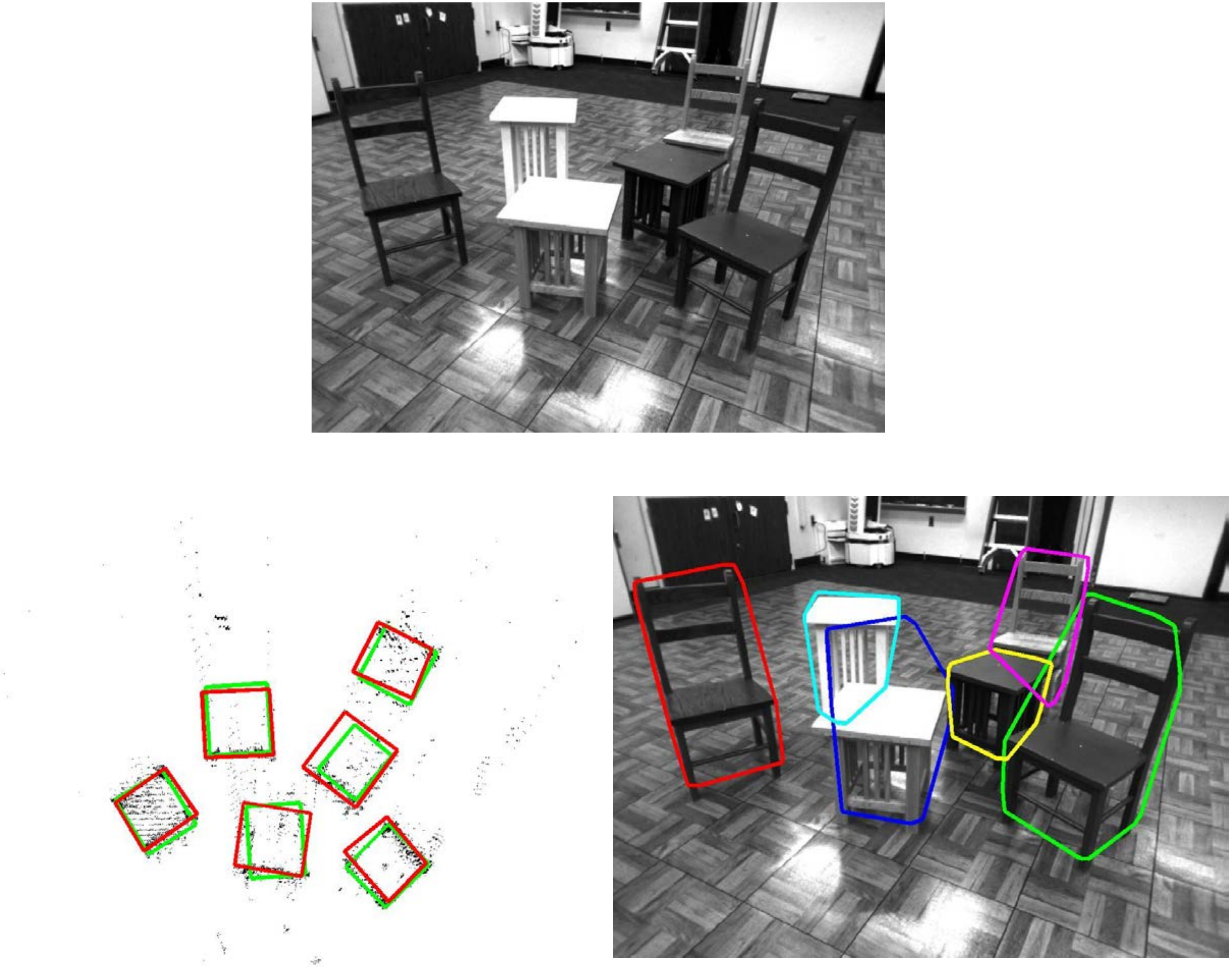
- Čapek, like us, can also see where objects end on the invisible side on their back.
- It can, like us, also “see” the invisible spaces behind objects: this capacity is essential for planning spatially-global navigations to perform actions in our environment, something we do very well.
- We do it all of the time.
- Can any super-intelligent robot do it?

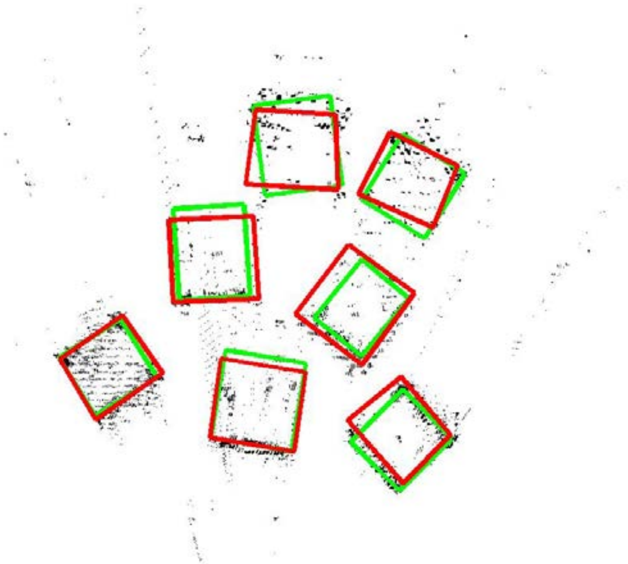


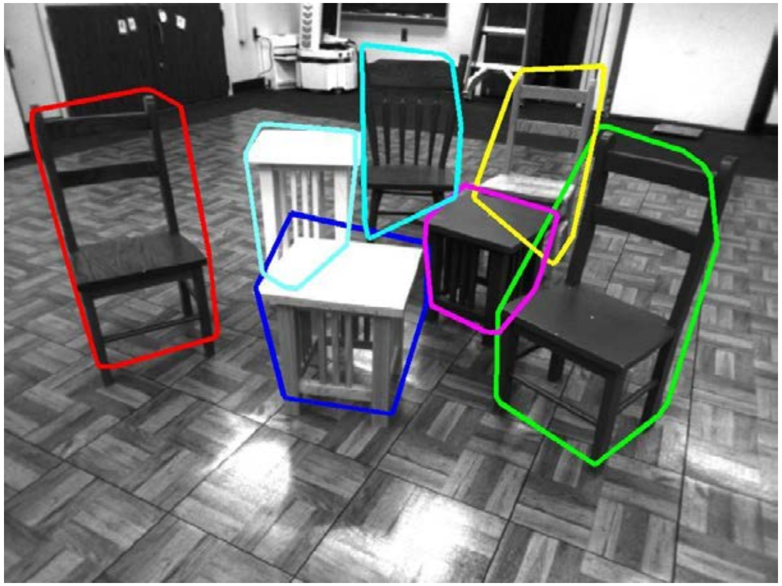


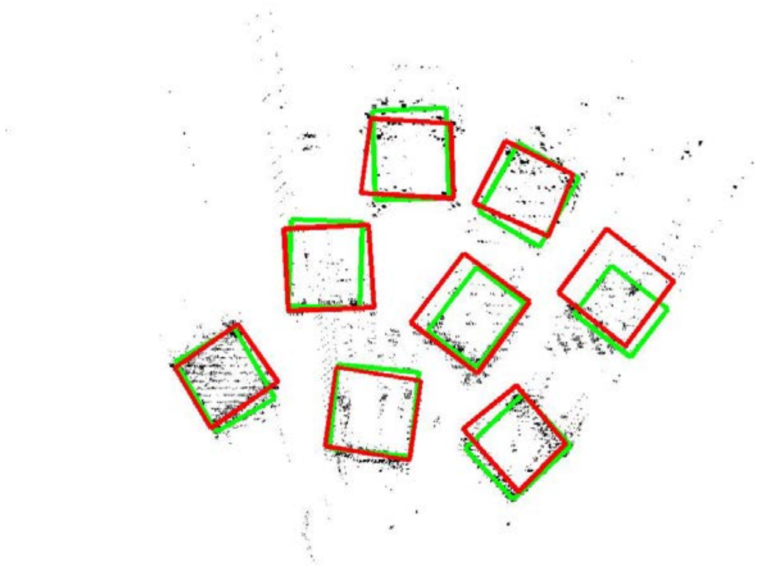


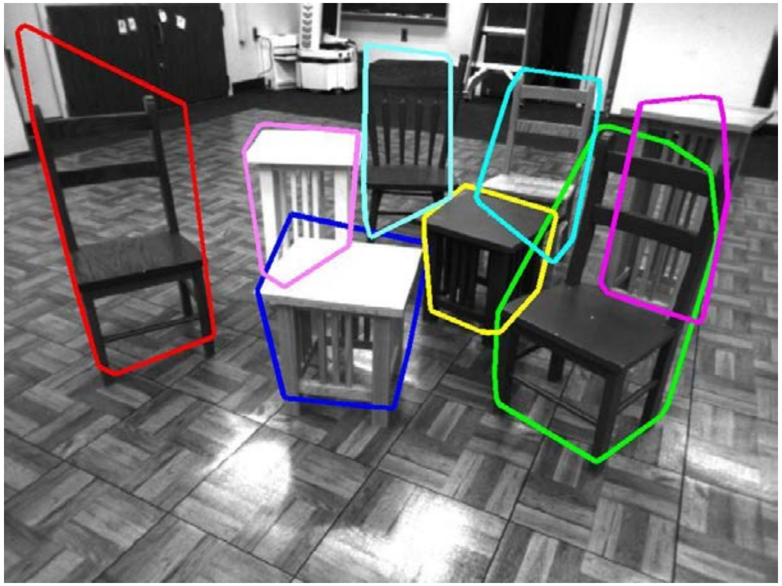


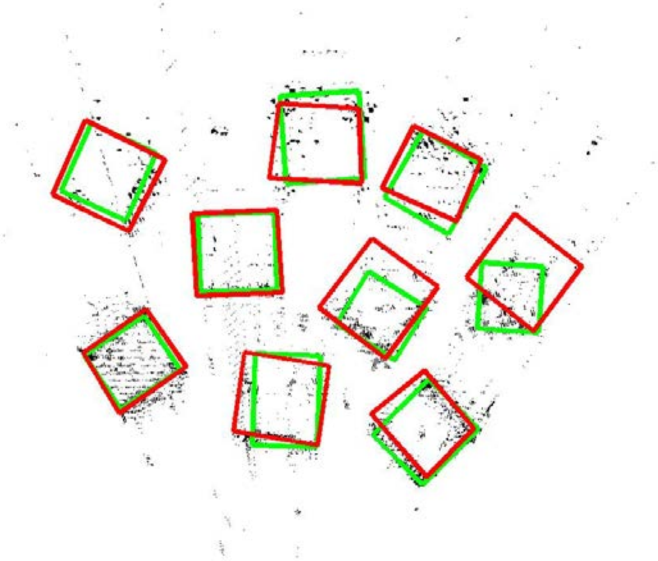


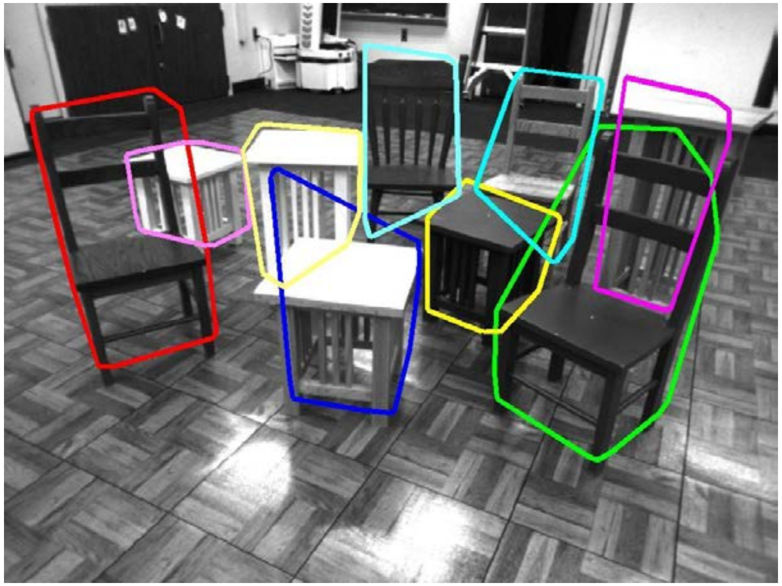


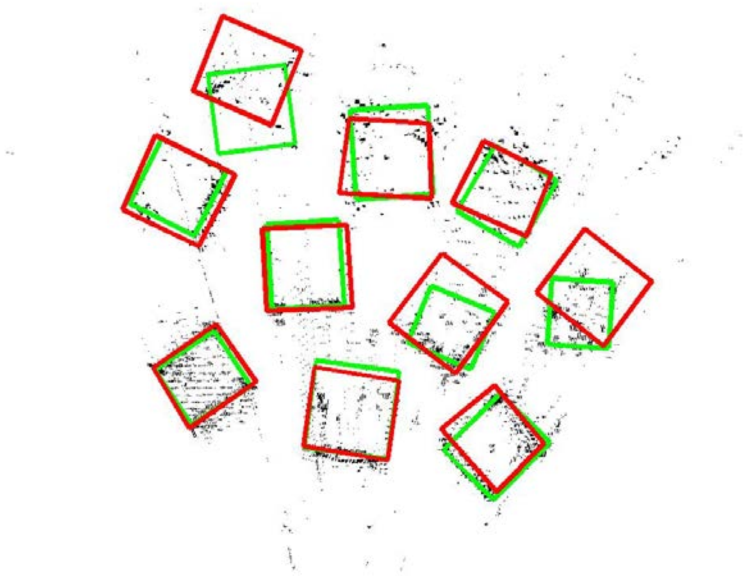


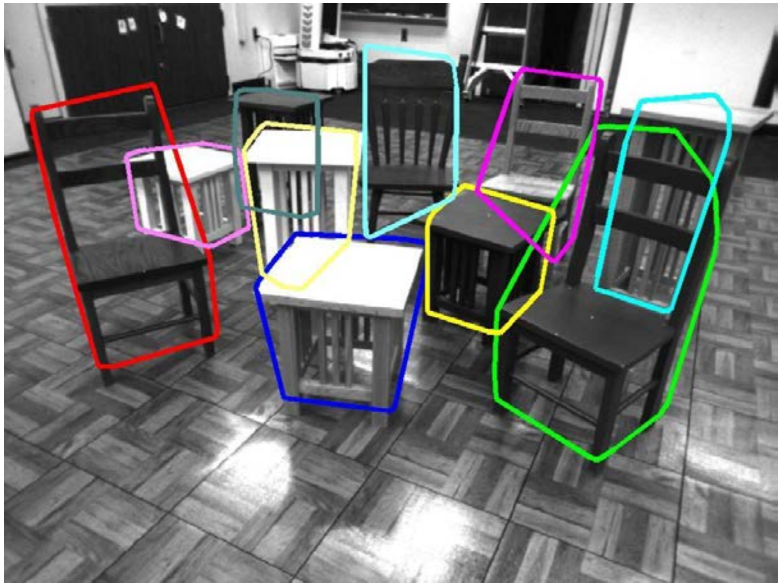


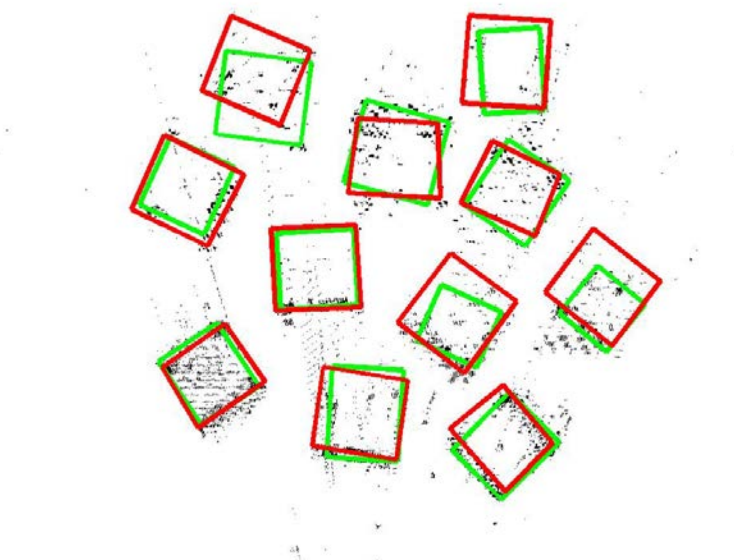


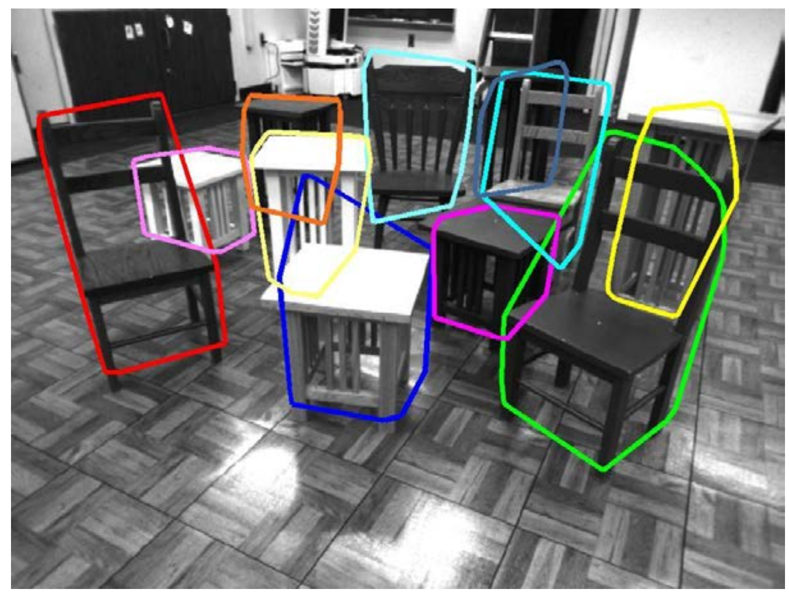












FGO within a dynamic scene

- [People wandering around]

Our model of 3D scene recovery

- What you have seen so far may be the very best way to solve the FGO problem and to perform 3D recoveries (recall these problems seemed insoluble when first brought up).
- If ours is the best way to solve them, there is no need to try to develop a robot with super-human seeing abilities.
- Note that making Čapek’s vision work like ours was *essential* for advancing our understanding how we humans see in 3D.

Summary

- Humans, among the animals, are said to be the most intelligent. Some of them can even be called “super-smart”.
- The only meaningful way to study HI is by emulating it.
- So far, we have barely scratched the surface in our attempt to understand the human mind and how it is used in the biological and physical world.
- Developing intelligent robots, which emulate us, can provide a powerful strategy and tool for achieving this goal scientifically.
- Let’s use it.

Conclusion:

- We should not be afraid of super-intelligent robots. *They are neither here nor even near*.
  - What is possible, and could be dangerous, is to build and use robots that are not-so-intelligent.
- We need to keep perfecting robots to bring their intelligence closer to ours. They are very far away now.
  - This may prove to be the best, perhaps even the only, way to understand and explain why the human mind works so well.


MAKING A
MACHINE
THAT SEES
LIKE US
ZYGMUNT PIZLO
YUNFENG LI
TADAMASA SAWADA
ROBERT M. STEINMAN
OXFORD