AMNON SHASHUA: So unlike most of the talks that you have been given, I'm not going to teach you anything today. So it's not kind going to be teaching type of talk. It will be more towards let's look at the crystal ball and try to see how the future will unfold and future where computer vision is a major agent in this transformative future.

So I'll start with transportation. This is the field where Mobileye is active. And then I'll move towards wearable computing, the field where OrCam is active. These are two companies that I co-founded. One, Mobileye in 1999 and OrCam in 2010.

So before I just-- a few words about the computer vision. I'm assuming that you all know about computer vision. It's the science of making computers see and extract meaning out of images, out of video. This is a field that in the past 20 years, through machine learning, has made a big jump. And in the past four years, through deep learning, has made another jump where there are certain narrow areas in computer vision and perception where computers reach human level perception and even surpass it. Like facial recognition is one of those areas.

And the belief is that in many narrow areas in computer vision within the next five years we'll be able to reach human level perception. So it's a major branch of AI. It goes together with machine learning and as said, a major progress. And one very important thing, which is relevant to the industrial impact of computer vision is that cameras are the lowest cost sensor that you can imagine. A camera sensor costs a few dollars. A lens costs a few dollars. All the rest is computing. And every sensor needs computing.

So if you can reach human level perception with a camera you have a sensor that the cost is so low that it can be everywhere. And that this is very-- this is very important. So I'll show you where things are standing in terms of avoiding a collision.

So avoiding collision, you have a camera. Behind the windscreen looking, facing forward, analyzing the video coming from the camera. And the purpose of this analysis is to avoid collisions. So what does it mean to avoid collisions? The software needs to detect vehicles, it need to detect pedestrians, need to detect traffic lines, traffic signs, need to detect traffic lights, detect lanes, to know where the car is positioned relative to the lanes. And then send a signal to the car control systems to avoid an accident.

So let's look under the hood what this means. So I'll let this run a bit until all the information appears. So if we stop here, what do we see? So the bounding boxes around cars means that the system has detect cars. Red means that this vehicle is in our path. The green line here is the detection of the lane. This is a traffic-- this is a no entry traffic sign. This is a traffic light being detected here. These are the pedestrians and cyclists. Even a pedestrian standing here is being detected. Let's no-- let this run a bit further. All right. So these pedestrians crossing the street.

So this is running at about 36 frames per second. So now imagine also the amount of computations that are being running here. Again, this is the traffic sign, traffic light, pedestrians, pedestrians here. So this is-- this is what the system does today, detect objects, detect lane marks, measure distances to the objects. And in case you are about to hit an object, the car would engage. Engage, at first it will give warnings. Then later it will apply automatic autonomous braking in order to avoid the accident.

And here is a list of many, many functions that the camera does in terms of detecting objects and detecting-- trying to interpret the visual field at details that are increasing over the years. Now computer vision is also creating a disruption. So if you would ask an engineer say, 15 years ago, what is a camera good for in this space? The engineer would say the camera is good for detecting lanes. Because there's no other sensor that can, you know, find the lane marks, not a radar, not a laser scanner. And it may be good for helping the radar infusion-- radar camera fusion to compensate for shortcomings of the radar. Traffic signs, OK it, will be good for traffic signs. But that's it.

But what happened over the years is that the camera slowly started taking territory from the radar. Until today, the camera is really the primary sensor for active safety. Active safety is all this area for avoiding accidents. And you can see this through this chart. So in 2007, now we launched the first camera radar fusion. So there's no disruption there. This is what normally people would think a camera is good for, combining with a radar.

2008, camera is also doing traffic sign recognition. No disruption there. 2010, camera's doing pedestrian detection. No disruption there because there's no other sensor that can reliably detect pedestrians. Because they emit radar very, very weakly. And pedestrians mostly stationary object. And radars are not good at detecting stationary objects.

But then in 2011, there's the first camera only forward collision warning. And that was the beginning of a disruption. So forward collision warning is detect a vehicle in front and provide a warning if you are about to collide with a vehicle. And this was a function that typically was in the territory of radars. So a radar sensor is very good at detecting vehicles, very good at ranging, very, very accurately can get the range of a vehicle, say 100 meters away up to an accuracy of a few centimeters. No camera can reach those accuracies.

So nobody believed that one day a camera will take over the radar and do this function. And this is what happened in 2011. And why did this happen? This happened because of a commercial constraint. The regulator, the American regulator, it's National Highway Safety Transportation Agency, NHSTA, decided that by 2011 all cars need to have as an option two functions, forward collision warning and lane departure warning. Now this creates a problem because forward collision warning requires a radar. Lane departure warning requires a camera. So now put two sensors in the car, it's expensive. If you can do it with one sensor, like with a camera, then you save a lot of money.

So this pushed the car industry to adopt the idea that the camera can do forward collision warning. And like all disruptions, once you start small you kind of grow very, very fast. So in 2013 the camera is not only providing warning, but also safe distance keeping to the car in front. It's called adaptive cruise control. Then 2013 also provides emergency braking. So the camera not only decides that you're about to collide with a vehicle, it will also apply the brakes for you. So 2013 was only partial braking. So to avoid the accident up to 30 kilometers per hour.

And then in 2015, this was few months ago, the camera is now involved in full braking. It's one G-Force of braking avoiding an accident of about 70 - 80 kilometers per hour. And mitigating an accident up to 220 kilometers per hour, just the camera. So the camera is taking over and becoming the primary sensor in this area of active safety.

Now, why is that? As I said, these are the milestones of the camera disruption, is first the camera has the highest density of information as a sensor. You know, laser scanner or radar, the amount of pixels per angle, per degree is much, much smaller. It's orders of magnitude smaller than a camera. So you have a lot, a lot of information from the camera. It's the lowest cost sensor. And also the cameras are getting better in terms of performance under low light.

So with a camera today you can do much more, not only because computing has progressed, not only because algorithms are now better, but also because the physics of the camera are progressing over time, especially the light sensitivity of the camera. So we also came to the conclusion that we need to build our own hardware and our own chip. And these are very, very advanced microprocessors that they can per silicon area are about 10 times more efficient than any general purpose chip. And I'll not spend more time on this.

And so this field has two major trends. One, on the left hand side, is the active safety, which is driven by regulators. So the regulators see that there is a sensor that is very low cost and saves lives. So what does the regular do? They incentivize this kind of function to the car industry by coupling it to star ratings. So if you want to get your four star or five stars the NCAP stars on the car, you have to have this kind of technology as a standard fit in the car.

So this pushes the industry by mandates. It pushes the industry to have active safety installed in every car. So by 2018 every new car would have such a system. The left hand side is the trend to the future, which is autonomous driving. Now autonomous driving has two facets. One is bringing the probability of an accident to infinitesimally small probability. So zero-- zero accidents. Because the more you delegate the driving experience to a robotic system, the less the chance of an accident.

So it brings us to an era where there will be no accidents. But not less importantly, it has the potential to transform the entire transportation business. How we own cars, how we build cars, the number of cars that would be produced. And I'll spend a bit more time about that as I go forward.

Now, in terms of the left hand side, the regulation, this is an example. So you see here a Nissan Qashquai 2014 has five stars. And to know how to get the five stars what you see here, these are all the tests. These are autonomous emergency braking tests. The car needs to detect the car in front, the target car, and apply the brakes before the collision. And the car is being tested. And without that they'll not get the five stars.

You can see this also in the number of chips that have been shipped. So every car has a chip. So this chip, the microprocessor, is getting the information from the camera and all the algorithms are on this microprocessor. So we started launching this in 2007. So in the first five years there were one million chips, so one million cars with the technology. And then in 2013 alone, 1.3 million. Then you see here, 2014, 2.7. This year is going to be about five million. So you see this doubling. And this is really the effect of the regulation.

So in many industries regulation is an impediment. In this industry, regulation is something good. It pushes the industry to install these kinds of systems, you know, standard. OK. So another example of how this is moving, there's also an increasing awareness. So this is a commercial from 2014 Super Bowl by Hyundai. So Hyundai is showcasing their new vehicle called Genesis. Now, there are many things that you can show when you want to showcase a new vehicle. You can talk about the design of the vehicle. You could talk about the engine, the infotainment. But they chose to talk about the active safety.

So I'll show you.

[VIDEO PLAYBACK]

- Remember when only Dad could save the day? Auto emergency braking on the all new Genesis from Hyundai.

[END PLAYBACK]

AMNON SHASHUA: OK. So this is the camera behind the windscreen, detecting the car in front, or a pedestrian, and will break before the collision. Now to show you what this is about-- so that was the commercial. So in a commercial you can show anything you like.

So now I'll show you something really from the field. So in 2010 Volvo introduced the first pedestrian detection. So the same thing, detect a pedestrian and if you are about to collide with a pedestrian the car would brake, apply the brakes automatically. So in 2010 they had about 5,000 journalistic events, where they put a reporter behind the steering wheel, tell the reporter to drive towards a mannequin, toward a doll, and low and behold the car would brake just before-- a fraction of a second before you hit the doll.

But then when you buy the car you can do your own testing. So I downloaded from the internet, its a bunch of Polish guys. And it's a bit funny but you'll actually get a good feeling of what this system does by looking at this clip.

OK? So this is automatic emergency braking. Today it works, it avoids accidents about 70 kilometers per hour. OK? So now you have a better idea of what I'm talking about. So now let's go into the future. So this was just setting the baseline. OK. What is active safety? Where's computer vision inside this?

So now let's look in the next four years. And the idea is to evolve this kind of technology to a point where you can delegate the driving experience to a robotic system. And then the question is, what needs to be done. And this slide shows that there are two paradigms. And the reality is somewhere in between these two paradigms. The right hand side is where we are today. You are based only on sensing. You have camera. Maybe you have also a radar, or laser scanner for redundancy. You get the information from the sensors. You have algorithms that try to interpret the visual field and take action in case of an accident, or control the vehicle.

On the left hand side is the extreme case, is the Google approach, where there is little sensing involved. It's a lot of recording. So you prerecord your drive. Once you have prerecorded the drive all what you need to do is to match your sensing to the prerecorded drive. Once you've found the match, you know your position exactly. So you don't need to know to detect lanes. You know all the moving objects. Because the recording contains only stationary objects. So all the moving objects pop out. So that the load on the sensing is much, much smaller than in the case where you didn't do it pre-drive and you didn't record.

This recording, the problem with the recording, is that we are talking about tons of data. It's a 360 degree, 3D recording, at several frames per second. So the amount of data is huge. So there's issues of how you manage this, how you record it, and how you update this over time. Because you have to continuously update this kind of data. And reality is going to be somewhere in between.

So the first leap that is undergoing and happening in the next five years is to reach human level perception. Now it sounds very, very ambitious. But there's lots of indications that it is not science fiction. It is really-- there is a very high probability that one can reach this. So in certain areas, like face recognition, certain categorization tasks-- now if you look at the academic achievements, they have surpassed human level perception. I'll spend a few slides on this later.

So going from adaptive-- going from driver assist to human level perception, first, we need to extend the list of objects. Not only vehicles and pedestrians, but vehicles at any angles, know about 1,000 different object categories in the scene, know about how to predict a path using context, which today is not being used. Detailed road interpretation, knowing about curbs and barriers and guards, it's all the stuff that when we look at the road we naturally interpret it very,

very easily. These are the things that needs to be done in order to reach human level perception.

And the tool to do this is the deep layered networks, which I'll spend a few slides about this in a moment. And the need for context, so these are examples, for example, path planning. You want to fuse all the information available from the image, not only to look for the lanes, because in many situations you look at an image you don't see lanes. But a human observer would very easily know where the path is just from looking at the context. In modeling data environment, ultimately every pixel give you a category. Tell me where this pixel is coming from, a pedestrian, from a vehicle, from inside of a vehicle, barrier, curb, guardrail, lamp post, so forth and so forth.

3D modeling of a vehicle. So put a 3D bounding box around the vehicle so that we can know which side of a vehicle I'm looking at, whether it's the front, or rear, left side, right side, what is the angle. Know everything about vehicles as moving objects and do a lot of scene recognition. I'll give some examples about that later.

So just deep networks, I know that you have-- you all know about deep networks. I'll just spend a few slides just to state what is the impact there, not the impact from the point of view of a scientist, but the impact from the point of view of a technologist. Because there isn't much science behind this.

So the real turning point was 2012. 2012, you know, the AlexNet, they show they built a convolutional net that was able to work on the ImageNet data set and reach a performance level, which was more or less double the performance level of what was done before. This is another network by Fergus. Very, very similar concept of convolution pooling. Convolution pooling two-three dense layers and you get the output. This is the ImageNet data set. You have about 1,000 categories over one million images. And these categories are very challenging. You know, you look at the images of a sailing vessel or images of a husky, the variation is huge. It's a really very difficult task.

2011 the top five-- so the task is that you need to give a short list of five categories. And if the correct categories is among the top five then you succeeded. And the performance was about 26% error. And this AlexNet reached 16%. So it's almost double the performance. So this caught the attention of the community. It's a big leap from 26% to 16%. Now if you look what happened since then, so 2012 for this

ImageNet competition there was one out of six competitors use deep networks. A year later 17 out of 24 competitors used deep networks. A year later 31 out of 32 using deep networks. So deep networks took over basically. If you look in terms of the performance, of the human performance is about 5%. And right now we are at 6%, 5%, by the latest 2015 competitors.

People start cheating. So I think this is more or less Baidu was caught cheating on this test. So I think 5% is more or less where things are going. And this is human level perception. Another-- another big success was the face recognition. So this is a data set called face recognition in the wild, which contains pictures of celebrities where every celebrity you have pictures, you know, along a spectrum of many, many years. You can see the actor when he was 20 years old and then when he's 70-80 years old. Even for humans, this task is quite challenging, knowing whether two pictures are from the same person or not. And the human level performance is 97.5% correct.

Now if you look at techniques not using deep networks, they reached 91.4%. And in 2014 a group by Facebook and Lior Wolf from Tel Aviv University, they built a deep network to do face recognition and reached 97.3%, which is very, very close to human perception. And since then people have reach 99% on this database. And again, human level perception is 97.5. So this is another area where these deep networks, also in speech. This is a recent paper by Baidu headed by Andrew Ng. They surpassed, just doing an end to end network which learns also the structured prediction, better performance than Siri, Cortana, Google Now. OK?

So the impact for automotive is that networks are very good at multi-class So the more categories you have, the better the performance of the network would be. Very good at using context, imagining or planning a path. So taking an image as an input and output would be the path. And you're cutting short of all the processes of looking for lanes and this kind of algorithms. Network will be ideal for pixel level labeling. For every pixel give me a category. And you can use the networks for sensor integration, for determining the control of the vehicle by fusing a lot, a lot of information coming from various cameras.

So the challenge of using deep networks is that deep networks are very, very large. They're not designed for real time. The networks that you find in academic papers and the success are for easy problems. The problems that I've shown right now, the ImageNet, the face recognition, are relatively considered easy problems in the context of interpreting the image for autonomous driving.

So let me show you what are the things that one can do. Let's start with the path planning. So this clip that I'll show you here the purpose of the network is to determine the path. So this is the green line. Now these clips are from scenes where it will be impossible to detect lanes. Because there's simply no-- simply no lanes. If you look at this, any lane detection system would find nothing in this kind of scene. Yet when you look at this image, you have no problem in determining where the path is. Because you're looking at the entire image context. And this is what the network is doing. It's being fed the input layer is the image, the output layer is this green line.

Or for example, if you at this urban setting. There are no lanes in an urban setting. Yet the system can predict where the path is by fusing information from the entire context. These are roads in California where they have these reflectors called Botts dots. It's almost impossible to reliably, you know, fit lanes to these kinds of information. Yet if you look at this holistic path planning it reliably can tell you where it is.

Let's look at free space. So free space, the idea of when you want to do autonomous driving you need to know where not to drive. Right? You don't want to drive towards the curb. It's not only that you don't want to hit other moving objects. That's the easy part. You don't want to hit a barrier or a guardrail. So you want to know where the free space is. So you can think of a network that for every pixel will give you a label. And let's now focus only on the label of road versus not road.

So all the pixel green are road. Everything else is not road. So you can see that the green is not going over the curb, which is-- which is nice. But let's have it run a bit more. And then I'll stop it at the place where you'll see where the power of context. Says let's assume I stop it here. Now look at the sidewalk there. The color of the sidewalk and the color of the road is identical. The height of the curb is about one centimeter. So it's not that the height here, the geometry-- it's basically the context.

The network figured out because there is a parked car there. That part is not part of the road. So in order to make this judgment correctly, one needs to not just look at a small area around the pixel and decide whether it's road or not road. One needs to collect information from the entire image. This is the power of context. And this is something that the network can do. You can see here, where the blue and red lines. Red means it's on a vehicle. Blue that it's on a physical barrier. So if I run this back here-- and this is done frame by frame. So it's a single frame thing.

Same thing here, this height is one or two centimeters. The color of the sidewalk and the color of the road it's identical. So being able to make the correct judgment here is very, very challenging. And this is where a network can succeed. Here the network also predicts that this is a code for being a curb. The red is the side of a vehicle or front of the vehicle. And the next one it predicts that this is part of a guardrail, the coding of this is part of a guardrail.

So the system has about 15 categories, guardrail, curb, barrier, and so forth. Let's keep the questions for later. And so forth. So this is one area we call semantic free space. So for every pixel in the scene tell me what it is. Of course, I'm interested-- first and foremost I'm interested to know where the road is. And then at the edges of where the road ends to know what is the label. Is it a side of a vehicle, front of a vehicle, rear of a vehicle. Is it a curb barrier guardrail and so forth.

And this, again, is done by deep network. I'll skip this one. And then you can apply this from cameras from any angle. So this is a camera looking at a corner, looking at the 45 degrees on the right. So the system can know where the free space is. This is a camera from the side, with a fish eye. Again, using the same kind of technology the system can know where the free space is. Same thing here. Here as well, day night.

3D modeling, 3D modeling is to be able to put a bounding box, a 3D bounding box, around the vehicle. And the color here is that the green is front, red is rear, blue is right hand side, and yellow is left hand side. If you let this run-- all right. Now the importance of putting a 3D bounding box around the vehicle is that now you can place a camera at any angle. So it's not only camera looking forward, but the camera at any angle, because the way a vehicle is defined is invariant to the camera position, so this is kind of a preparation for putting cameras all around the vehicle at the 360-- 360 degree.

Scene recognition, for example, being-- to know that this is a bump, is also being done by a network that takes an image and outputs where the bumps are. The same thing-- same thing here. More complicated than that is knowing where this top line is. So when you go and detect traffic lights-- so detecting traffic lights is the easy problem. A more complicated problem is to know the relevancy of the traffic lights, which traffic light is relevant to what direction. The third one, the most difficult problem, is to detect the stop line.

The problem with stop line is that when you see the stop line it's a bit too late. You see this stop line 20-30 meters away. So it's too late to start stopping and have a smooth stopping.

You want to predict where the stop line is 60-70 meters away. So here, you want your algorithm, or your network, to understand that you are approaching a junction and start estimating where the stop line should be so they can start slowly reducing your speed, such that by the time you see where the stop line is you already reduced your speed considerably.

I'll skip this. Knowing lane assignment. Knowing how many lanes are and which lane you are is also done by a network. So the network will give a probability whether that this is a lane, this is a lane. For example, it knows that this is not a lane. It has here red, zero probability. So as you can see here-- I'll skip this one. So these networks, so for every task there is a network. And these networks are quite sophisticated in accessing, integrating a context at traffic light. I'll skip this with traffic light.

So multiple cameras, this is how it is-- it looks like. You have the red ones are three cameras behind the windscreen. One is about 180 degrees. The other one is about 50. The third one is about 25 degrees. And then there are another five cameras around the car that give you all 360 degrees. And this kind of configuration, first launch of it, in a series produced car, is going to be 2016. So I'm not talking about science fiction. These are how images look from some of these cameras.

So let me show you a first clip of automated driving. This is kind of a funny-- funny clip. This is an actor who played a major role in Star Trek. So I'll not say his name. Let's see whether you can identify him yourself. And he has a program called-- program for kids called Reading Rainbow. So this program is 20 years old. And he came to Israel and he wanted to drive the autonomous vehicle that we have for his kids program.

So he was driving my car. So my car is autonomous. I can drive from Tel Aviv to Jerusalem without touching the steering wheel. It's-- I do that all the time. So he was driving it. And it's a bit funny. So let's-- but you'll get a feeling of what this is. So let's run this. It's two minutes.

[VIDEO PLAYBACK]

- Yes. They can. That's because technology companies, like Mobileye here in Israel, are about to introduce self-driving technologies to the world.

AMNON SHASHUA: You know who he is?

- In the not too distant future, just like in a science fiction movie. A driver will be able to hop in a car, tell it where you want it to go, and voila, the car will do the rest. So right now I'm driving like everybody does. My hands are on the steering wheel and my foot is on the brake, or the pedal, as required. And I'm in control of the car. But when I take my foot off the pedal and do this, now the car is driving itself. Wow. This really is amazing.

I feel really safe with the car doing all of the driving. OK. Now watch this. And this is something that no one should ever do in a regular car, ever. Wow. That was freaky.

[END PLAYBACK]

AMNON SHASHUA: OK? So anyone from the young people know who he is? So this is Jordy, from *Star Trek.* He had this visor. He was blind. He had a visor. OK. So let's spend a few minutes to talk about what is the impact of autonomous driving and how it's going to unfold.

So this is far from science fiction. It's actually unfolding as we speak. The first hands free driving on highways is coming out now. The first one is Tesla. They have already launched-- they made this public a week or two ago. Their first beta drivers are driving with the system. And I presume within a month it will be also installed to all other drivers. And this is-- you can do hands free when driving on a highway, unlimited speed. So you can drive at highway speeds, let go of the steering wheel, and the car will drive.

GM already announced that middle of 2016 they have the super cruise, more or less the same kind of functionality. Audi also announced 2016. And these are just the first comers. We are working with about 13 car manufacturers that within the next three years, three to four years, having this kind of capability. So this will be in the mainstream.

Now what I put there in red is that the driver still has primary responsibility and has to be alert. That means that the technology is not perfect. It could make mistakes. Therefore, the driver has to be-- is still the primary-- has the primary responsibility. So at this stage there's no disruption here. It's just a nice feature to have. For the car industry, this is the first step to start practicing towards reaching autonomous driving.

The second step starts 2016, and this is with the eight cameras that I showed you slide before. Here, the car can drive autonomously from highway to highway. So on ramp, off ramps, are done autonomously. So you-- you with Google Maps or whatever navigation program you chart your route, and until the car reaches city boundaries it will go all-- go autonomously.

From highway to highway it will switch from highway to highway and do that autonomously. Still, the driver has primary responsibility, and is alert. So there's no-- nothing here is transformative. It's a nice feature. Again, it's part of a phased approach of the car industry to start practicing.

Starting from 2018, would come the first small disruption. The first small disruption is that technology would reach a level in which driver is responsible. The driver must be there but not necessarily alert. So it means that the driver is an attendant. The driver is monitoring just like a pilot sitting in an airplane while the plane is in auto-pilot. The driver needs to be there in case there is a problem. The system will give a grace period of time until the driver needs to take back control. So it's not taking control in instant-- immediately. And so this transition from primary responsibility to monitoring, like in aviation, will be the first disruption, the beginnings of a disruption.

So let's try to imagine what kind of disruption this is. So let's take Uber as an example. So today you have free time. So you own a car. You have a free time, say between 3:00 PM to 5:00 PM So you take your car and apply Uber and take passengers and earn some money. That's Uber today. Now let's look at 2018 - 2019. You have zero skills and you don't have a car. All what you have, you have a driver license. So you are willing to be an attendant. So you say, OK, now I have free time. An Uber car would come with an attendant. You switch places with the attendant. You sit behind the steering wheel and you do nothing. You don't control the car. You don't control the passengers who are coming who are being taken by the car. You simply sit there. Zero skills, therefore your payment is very, very small.

So now these cars can drive 24/7 because attendant can be replaced every hour or so. So here we have another business model which makes this public transportation, the Uber type of public transportation, now much more powerful than it is today. So this is kind of the beginning of disruption.

What will be the next step? The next step 2020-2022, imagine that a driverless car can drive without passengers. So this is one step before you can allow a car to drive autonomously. So without passengers means that all what you need to prove is that the car, your car, would not hit other cars or pedestrians. But if it hits an infrastructure nobody gets killed because there are no passengers in the car. No passengers meaning nobody in the car. Now this is already a major disruption because what it means, it means that the household does not need to own multiple cars. One car is enough. I drive to work with the car. I send the car back home. Takes my wife, take her to work, comes back home. You get the picture.

So this is kind of a beginning of a major disruption. Then about 2025 - 2030, sufficient experience with mapping data, car to car communication, one can imagine how these cars would be completely autonomously. And that is where the major disruption happens. OK? So this is autonomous driving. Let me go to the second part about wearable computing. And then we can take questions. So this will be much shorter.

So again, computer vision, but now the camera is not beside us, like in the car. The camera is on us. Now if the camera is on us, the first question that you would ask, who needs a camera to be on you? Right? So the first market segment for something like this are the blind and visually impaired. So the way to imagine this. you are visually impaired or a blind person so you don't see well or you don't see at all, or you don't see well. So it's very, very difficult for you to negotiate the visual world. You cannot read anything unless it's few centimeters from your eye. You cannot recognize people unless they start talking to you. So you can recognize their voice. You cannot cross the street because you don't see the traffic light. You cannot go on a bus because you don't think what the bus number is. So basically you are very, very constrained, very limited.

Now let's assume that you have a helper standing beside you. Now this helper is relatively intelligent and has correct eyesight. Now the helper looks at you, sees where you are pointing your hands, for example, or pointing your gaze, looks at the scene, understands what kind of information you want to know, and whispers to your ear the information. So say you want to catch a bus. You know that the bus is coming because you hear the bus, maybe you see a silhouette. So you look at that direction. The helper looks at the bus. It sees that there is a bus. Tells you what the bus number is.

You want to cross the street. You know the traffic light is more or less there. But you cannot-- you don't know what the color of the traffic light is. So the helper looks at your gaze, sees that there's a traffic light there. Tell you it's a green light. You're opening a newspaper. You point someone on the newspaper, the helper would read you the article. Or there is a street name. You point towards the street name. The helper would look at the scene, understand that there is a text in the wild, and simply read you the street name. A familiar face appears, the helper will whisper, you know, Joe has now-- is now in front of you. And so forth.

So if you have now replaced this helper with computer vision you can imagine how this could help someone who is visually impaired. So let me show you-- so first of all, the number of visually impaired is quite big. So the number of blind people in the US is about 1.5 million. That's not big. The number of visually impaired, and it's people that their ailment cannot be corrected through lenses, is about 26 million. So this is a sizable number. World wide is above 400 million people who are visually impaired. And they don't have much technology to help them.

So this is what OrCam is doing. It's a camera which clips on eyeglasses. And there is a computing device, which you put in your pocket. And the way you interact with the device is with your hand, with your finger. Because the camera is on you it could see also your hand. Once you point, the camera starts to extract information from the scene and talks to you through an earpiece. So let's look at the clip.

[VIDEO PLAYBACK]

- Hi. I'm Liette and I'm visually impaired. I want to show you today how this device changed my life.

- Massaryk.

- Great. Let's go there.

- Red light. Green light. 50 shekel.

- 50 shekel. Let's buy some coffee.

- Breakfast. Bagel plus coffee with cream cheese [INAUDIBLE].

[END PLAYBACK]

AMNON SHASHUA: OK? So you get the idea. So we started 2010. 2013 we had already a prototype working. And we had a visitor, John Mark from the New York Times, and he came and he wrote a very nice article about what the company is doing. And we thought that at that time it would be good to launch the website of the company and try to get a number, say, 100 first customers, so that we can start experimenting, do field studies with a prototype device.

So we launched the web site. We wrote that the device cost $2,500. That was June 2013. And the first 100 people who would purchase the device will receive the device in September. So within an hour those 100 devices were sold. And then we kept a waiting list, which today is about 30,000. And we started shipping the devices about a month ago.

So in the last year this device was with about 200 people. And we got a lot of feedback from real users and improved. And let me show you some real users. So this is Marcia from Brazil. The device at the moment only works in English. Later we'll put more languages.

And so she's being trained to use a device. And this is a short clip of about two minutes. And, you know, watch her body language. And also she explains how she copes with her disability, especially how she distinguishes between different money notes. They're all green. So how do you distinguish between them? So let's have a look at this.

So the device is reading the newspaper for her.

[VIDEO PLAYBACK]

- [INAUDIBLE]

- $50.

- $50. Cincuenta dollars.

- Cincuenta. Let's see if [INAUDIBLE].

- It green. All green and I put mark color, yellow, green, orange. Different note. [INAUDIBLE]

- $20. [? Genia ?]

- [INAUDIBLE]

[END PLAYBACK]

AMNON SHASHUA: OK. Here's a recent-- from CNN. It was aired a month ago. It also gives a bit more information about the device. Let's run this. It's again two minutes.

[VIDEO PLAYBACK]

- Two weekends ago I sat down and read *The New York Times.* I haven't done that in maybe 30 years. My wife came down. I had a cup of coffee. I'm reading *The New York Times* and she was crying.

- Just being able to read again is emotional for Howard Turman. He started losing his vision as a child. His new glasses don't fix his eyes but they do the next best thing.

- Put on my glasses, it recognizes the finger, snaps the picture. Now it just reads.

- The glasses have a camera that recognizes text and can read the world to him.

- Pull here.

- The technology is called OrCam and Turman says it gives him a sense of normalcy.

- Even finding out that Dunkin' Donuts has a donut I never tried was exciting.

- Dunkin' Donuts.

- It's a clip on camera. So a camera that you can clip onto any eyeglasses. And you have here a computing device, which you can put in your pocket. And the way it interacts, it's with a hand gesture. For example, it's written there, rental and tours.

- Rentals and tours.

- It's not perfect though. It uses a pretty bulky cable and sometimes it needs a few tries to get things right.

- It doesn't read script because everybody's handwriting is different. So it doesn't do cursive very well at all.

- OrCam has a harder time in bright light, or in tougher situations, like signs on windows.

- [INAUDIBLE] U donuts hours of operation. Low PM. Pound's PM. 9:00 PM. How was your service today?

- Shashua says improvements are on the way. Where do you see this technology going over the long term?

- Reading, recognizing faces, recognizing products, is only the beginning. Where we want to get is complete visual understanding at the level of human perception, such that if you are disoriented you can start understanding what's around you. For example, where's the door? The door is there. Where is the window? Where is an opening in the space around me?

OK? This is face recognition. So again, one of the first 100.

- Teach OrCam to recognize anybody?

- Yep.

- Who does it know?

- Libby, my mother.

- You want to show me?

- Yep. OK.

- All right. [INAUDIBLE] Let's see. [INAUDIBLE]

- Libby.

[END PLAYBACK]

AMNON SHASHUA: OK? So that's also face recognition. Last two slides. We started also providing the device to research groups. And this is one of-- this is a paper in ARVO where they took eight visually impaired and gave them the device for one month. And then measured the change in quality of life. And how they measure the change of quality of life, they interview them. And seven out of the eight reported significant change in quality of life. Now they sent us some of the interviews.

So on the next-- here, I'm showing you part of the interview. And what's interesting about this interview is that there is a trick question. The interviewer, after she tells him how the device is, you know, lifesaving and so forth, he tells her, well the device is very expensive. It's a few thousands of dollars. Is it worth it? So it's one thing to get something for free and say, it's very, very good. Another thing's is it's going to cost you thousands of dollars, is it worth it? And let's hear her answer, which is very nice.

[VIDEO PLAYBACK]

- In the first few days I had the OrCam I was in total awe of it because for the first time I was able to open mail and read it, instead of having my husband read my mail. And I was able to go to a restaurant and actually read the menu and order myself with the waitress. And that was exciting. When you can't do something for such a long period of time, the OrCam was incredible.

- Believe is what the estimate is. Do you think such a high price would be something people would be willing to pay for a device like this? Do you think it's marginally worth it right now?

- I think you're going to find that that's going to be on a case by case basis. You know, people who have money there's certainly no problem $2,000. I don't have money. I am low income. But I would save my money, scrape it together in order to get it at $2,000.

[END PLAYBACK]

AMNON SHASHUA: So that's interesting. Where is it going? So there are two lines of progress. One, is when this existing niche is to make the camera understand the visual field at higher levels of detail. So one of the things that we are now working on is, we call this chatting mode. So it's like-- it's the image and notation type of experiment, or the ImageNet together with natural language processing. Say you are visually impaired or blind and you're disoriented. You don't know where you are. So you would like the device to tell you every second what it sees. I see here Tommy. I see here chairs. I see here another person. I see here a wall, an opening, a painting, blah, blah, blah, blah, blah, until you get back your sense of orientation.

So you want the device to be able to have say, several thousands of categories, like in ImageNet, together with image annotation capability. The kind of stuff that people are now writing articles about. And being able to do this at the frame rate of say, once per second. So wherever I'm looking at tell me what-- what you see. This is one-- another thing is to have natural language processing, NLP ability. For example, if you are looking at an electricity bill. The system would know that you're looking at an electricity bill and give you just the short-- what is the amount due, for example. The system will tell you are looking at an electricity bill. The amount due is such and such.

So having more and more intelligence into the system. So this is one area. Another area is to go for a wearable device for people with normal sight. So here we're talking about, you know, real wearable computing. So this Apple Watch is wearable computing. But doesn't do much computing. Right? It displays, you know my text messages, emails, you know, measures certain biometrics. But that's not, you know, the holy grail of wearable computing. The holy grail of wearable computing is assume that you had Siri with eyes and ears. So you had a camera on you that is observing the scene all the time and providing you real time information whenever you need the information, like the people that you meet. What were the recent tweets of those people that you met? What is common between you and them based on

Facebook and LinkedIn and so forth?

So knowing more about the people you meet. Knowing more about the stuff that you are doing. And creating an archive of all what you are doing throughout the day. And this is a device like this. This is how it looks like. We call it Cassie. So this is a real device. It works continuously for about 13 hours. So you have a camera working continuously for 13 hours. And the purpose of this camera-- so the way-- you put it like this. OK? So the purpose of the camera is not to take pictures. It doesn't store any picture. The purpose of the camera is to be a sensor, is to interpret the visual world and provide information in real time. And if everything goes well, we'll start launching this within six months from now.

So this is the next big thing, to go into wearable computing, to go into a domain in which a camera is on you and processing information all the time. Unlike now, a camera on your smartphone in which you take a picture on demand. It's not working for you all the time. Here it's working for me all the time. All the time it's viewing the visual field. Whenever it finds something interesting it will send it to my smartphone, like people that I meet and other activities that I do. So this will be the beginning of real wearable computing. So wearable computing with sensing, with the ability to hear and listen, to hear and see, and process information in real time. This is the next thing that-- the next big challenge that we are working on.