**KURT ANDERSEN**

Our next speaker is - he's done an extraordinary thing in a very short period of time. He started off, only a few years ago at the beginning of this century, as a Princeton Physics BA and Applied Math PhD. He has applied computer power in various interesting ways to neuroscience and history, among other fields. Three years ago, he founded a company called Seadragon to develop software to deal with massive amount of visual information. Microsoft bought his company in 2006 and now employs him and his team in their sort of futuristic lab.

He was the architect for, what I believe, was the first great product of Seadragon in his work, which is called Photosynth. It's the kind of software, and this just doesn't do it justice you have to see it, "Takes a large collection of photos of a place or an object, analyzes them for similarities, and then displays those photos in a reconstituted three dimensional space. It's extraordinary. It is an unquestionable leap into the next - I think of that funny a cappella song in the 2020 sequence about our wacky Jetson's future. Well, this fellow makes you think it's closer than you think. I present to you, Blaise Agüera y Arcas.

**BLAISE AGÜERA Y ARCAS**

Thank you very much for that very kind introduction. It's a little bit sobering to follow Alex Steffen. The things that we're doing, that my team and I are doing, I think, are really not very important compared to those issues. But maybe there is some relevance in dematerializing that we can bring to there.

What I'd like to talk about today is, it's a little more than just Photosynth the tech preview, which probably some of you have seen since it's been on the web now, for some months. A lot of us grew up reading science fiction. And in particular, if you're my age, you might have grown up reading a lot of Cyberpunk in the '80s and then the early '90s. And there was this idea in the air that when computers began to connect together and network in a very real sense, we would start to see cyberspace. We would start to see a kind of mirror of reality behind the screen, and then when the web came along, it was sort of a surprise. You know it was so simple.

It's just text documents with some other lined words and hyperlinks. And it was both very simple and very beautiful, and very within reach. It was surprisingly within reach and that's why it changed everything so quickly. I think more quickly than anybody really realized was going to happen because it was so much simpler and because it made use of things that we already had and we already knew how to deal with. And as the years passed those kinds of ideas about, you know, of virtual reality visors, and goggles, and all sorts of things like this, kind of faded away. And cyberspace, these days, has come to mean mostly just being online, I suppose.

But a lot of what has motivated us and a lot of what we do is about bringing back certain aspects of that original vision in thinking about what's missing, what pieces are missing and what we can do to bring some of that original vision back, I suppose, in Ballard, on the

waterfront in Seattle. And that's where we are now. This is the Smith Tower. When it was being built, I think in 1914, I think it was the tallest building outside of Manhattan at the time. So we're not on the main campus in Redmond. We were actually going to move here originally before Microsoft bought us and part of the deal was we'd still get to move in here and keep a certain amount of independence. And it's worked out really beautifully so far.

But having said that, what we came to Microsoft with is actually much less than what we have now. And a lot of that much less is so because of collaborations that we've been making with Microsoft Research and with the University of Washington, which Microsoft Research works with very closely. This is a story - well, let me not jump ahead of myself.

Let's have a quick look first at what Seadragon did, and I don't want to spend too much time of this one, but this is the Seadragon engine, and some of you have probably seen this before in demos online. But the idea is that you have, in this case, quite a large number of, not a very large number, just a few hundred, photos, and these are in the six to eight-mega pixel range for the most part. Some of them are larger panoramas. We're accessing them all very fluidly. It almost doesn't matter whether you're looking at these locally or remotely. That is, whether the contents is on the hard drive or whether it's somewhere out there in the cloud on the web.

And the reason for that is that the way we think about interacting with documents on computers nowadays is, I think, a little bit broken in the sense that we think about an application that opens a document that means reading the whole thing into memory. And then if it's too big, well, it takes a long time. And if it's too big to show on the screen all at once then we have all of these scroll bars and other sorts of affordances that let you look at pieces of it. But you're not exploiting the fact that either you can see many objects at low resolution, or one object at fairly low resolution, or you can see a part of an object at very high resolution.

But you can never see the entire object simultaneously at very high resolution. The amount of information you can see at any given moment is constrained by the number of pixels on your screen. So why not think about documents as being sources of information that you interact with and this is something we've seen in mapping software. For example, we've seen it in online mapping software in the past few years. But we really propose to bring that kind of capability in a very generic way to all sorts of documents and all sorts of interactions.

This is one that I always enjoy showing off. It's "Bleak House" by Dickens. Every column is a chapter. And I'm not really saying that this is a good way to read an e-book, but it does show that the architecture is very generic. You can use these kinds of techniques to interact with documents whether there're images, or text, or more complex things. This is an issue of The Guardian that shows off some of the advertising possibilities for things like this. I think if we took this kind of approach we could do away with things like popup ads that are irritating and that are there because of the idea of limited screen real estate, which is really a glass ceiling, of course.

So in this case, we've taken a real issue of The Guardian, and it's nice to read with this kind of experience because things like newspapers and magazines really are inherently multi-scaled experiences. You have headlines, and small type, and so on. And in this case we've added a fictional ad that has some even smaller type than you can fit into the print publication, and even smaller type, and even smaller type, and we can see what the $CO_2$ emissions are of this car.

The reason that I think this kind of approach is really interesting for thinking about rich content on the web, is that there's a real barrier to clicking on a banner ad. For example, if you're on a webpage and you have some dancing figurine, you're not going to click on it because that's going to take you out of the context of that webpage or that environment. It's going to force you to go somewhere else, or it's going to pop something up. There isn't the sense of continuity about the way we interact with information typically in computers that we get in real life. If we can just look more closely and do it in a continuous and natural way, and back out anytime you want.

All right. So this was just toy mapping application, which we, I think, won't spend time on. But one thing that I want to point out, it's a really interesting fact, that the amount of processing power in our typical computer, even a laptop like this one, a fairly modest laptop, it has a hidden resource that we don't often think about, which is the graphics processing unit, the GPU. Almost all computers, nowadays, have one and by many measures it has a lot more power than the CPU does. It's almost like the tail that is about to wag the dog. And yet, aside from games, we don't really see very much use of the GPU, yet. And I think this something that is really going to change quite dramatically in the near future.

So what I'm showing you here is just that we can do all the same kind of Seadragon stuff in 3D and that was a capability that we just got for free because we're using the graphic's hardware that way, with the same kind of richness of content and so on. Now, this was just a toy. Back when it was Seadragon we didn't have a good use for the 3D. And then we saw this. This was about, I don't know, about a month of two months after the acquisition of Seadragon, and when I saw it for the first time it just blew my socks off. It was a project by a graduate student, Noah Snavely. He was at the University of Washington. He was co-advised by Steve Seitz at UW and Rick Szeliski at Microsoft Research.

Maybe it's worth actually just saying parenthetically, because people outside of computer science research don't really know this, I think. Microsoft Research is really a phenomenal place. It's about 15 years old and it's really become like the Bell Labs of computer science, nowadays. It's very different from what one usually thinks about when you think about Microsoft. At the last SIGGRAPH Conference the big graphics conference that happens every year somewhere between a quarter and a third of all the papers had a Microsoft Research coauthor which is staggering, I think a staggering number.

Anyway, so at the MSR technology fair in February 2006 we saw Photo Tourism and this is the idea behind Photo Tourism. Noah took collections of photos. And in this case, this is a collection of a few hundred photos of Notre Dame Cathedral that he mined from Flicker and

he made a three-dimensional reconstruction of Notre Dame Cathedral based entirely on those photos. I'll show this live in a moment. But that cloud of points that you're seeing in the background is Notre Dame in 3D reconstructed entirely from the photos and all of these little orange cones are where all the pictures were taken from. Those are reconstructions of the positions of the cameras when all of those shots were taken.

What you're seeing over here is the frustum or the viewing cone of one of those cameras. Photo Tourism had two different kinds of reactions, one from insiders in computer vision and one from everybody else. And from everybody else it was, holy shit you can do that? It was just nobody knew that it was possible to do this. And within the computer vision reaction was a little different. They also thought it was brilliant, but they thought it was brilliant for a different reason, because the idea of reconstructing 3D from images is actually one that has quite of a lot of history. Just like the web, I think, just like the text documents, and the idea of hyperlinks, and so on, it has an academic history. It's been going on for many years.

The first steps along that path happened in the early '80s, and so in that sense, Noah was really using off the shelf techniques to do his 3D reconstructions. The clever thing is that instead of trying to go for the what has always been thought of as the holy grail in reconstruction of 3D from images, which is to do something like make a game level, make a Quake like game level out of nothing but photos and video, instead, he and his advisors realized that there's a lot of value in the photos. And in fact, maybe the photos are more valuable than a Doom like game level that has Notre Dame Cathedral in it.

That the photos are actually what matters, and that putting the photos into context in relating them to each other and connecting them, and building a user interface that's all about relating the photos to each other, you actually have something a lot more powerful than just a 3D reconstruction. And it has all sorts of interesting implications.

So I don't want to spend too much time, I see I only have 12 minutes and 29 seconds left, so I don't want to spend too much time on the Algorithm, but I do want to give you a bit of appreciation for what's actually going on here. So what you do with images, to start with is you first do something called feature extraction, and then you match those features to each other and that incorporates images, and then you do 3D reconstruction based on the feature matches by David Lowe, at the University of British Columbia. He first showed it off, I think, in 1999. Again, it was the evolution of a lot of work, starting with more back in the '80s.

These are pictures from a paper of his of two objects, a plastic frog and a plastic truck. And he learned the features, or he had his computer learn a set of features in these two objects, and then he put them into this cluttered environment, and they're partially obscured, and they're upside down, and so on, and the algorithm finds the objects. This is really the key to being able to do reconstruction of scenes form objects. You find features in the photos and then we find where those features reoccur in other photos even if they're tilted or at different scales or different angles. This is just another example of the same thing.

And in this case, these photos that he used to learn features were taken from different points of view from this photo. But all of the occurrences of those features in the big photo are found. This is the same thing. When we do it those are, I don't know how visible it is, but that's lots of little boxes, each representing a feature in this image of the space shuttle. We've done some collaborations in Photosynth with NASA around doing reconstruction of the space shuttle and they've been really interested in using it for doing things like tracking the way the heat tiles get damaged over time and so on.

So you take little features out of the images and each feature, each of those little squares that you are seeing, gets what's called a descriptor. And the descriptor is a series of numbers that describes what that feature is in a way that you hope is unique to that feature but that's robust to different points of view in different cameras. In other words, these numbers that correspond to what's in this box, and this, by the way, in David Lowe's original, was just a series of 128 numbers. The hope is that those 128 numbers are going to be very similar for different points of view on this same feature from taken from different cameras and by different users and so on. Okay. We'll skip the geeky stuff.

You take a bunch of pictures, you get all the features in each picture, and then you match, and then you do a 3D reconstruction. So matching is just the process of taking all of the features that occur in all these images and finding their closest neighbors in other images, and therefore, figuring out when you have pairs of images that seem like they're looking at the same thing. And now, because you have features and you know where those features occurred in all of those images, you can solve a big system of equations that says well, where did those features have to be in three-dimensional space in order to land at the spots in the imaging plane where they landed on all those photos.

Okay. So think about it this way, what you're solving for, what you're trying to figure out, is two things simultaneously. For each feature where was it in 3D space? We know where it was in 2D space in each one of the photos, but where is it in 3D space consistent with having landed at that 2D spot on all the photos, and therefore, where would all the cameras have to have been in order for them to end up there? And that's what it is.

Well, when we first saw this, we immediately put our Catalan designer, who designed our cool Seadragon logo, on the job of figuring out what the Photosynth logo was going to be and came up with that. The reason that this seemed like such an interesting marriage was because all of this is about taking large corpuses of images and figuring out their relationships to each other.

And in case the connection isn't totally obvious by now, of course, for dealing with large corpuses of images remotely, if they're very high resolution, and if there are many of them, and if you want to access them instantly, Seadragon is a very nice technology for doing this, and we were just waiting for the right way to do things like spatially arrange photos relative to each other. And this was just the prefect marriage.

All right. So let me stop with all the PowerPoint stuff and pull up - now, this is live from the web. So any of you who have your laptops open, the wireless in here is pretty crap, so please close them so I get all the bandwidth I can.

(AUDIENCE LAUGHTER)

This is one of my favorite environments that's already up right now on the web. It's Gary Faigin's studio. He's the NPR art commentator in Seattle. When I click this button on this environment, we get something that looks a lot like a Seadragon view of all these images. And, if the network's willing, we see things refined in the Seadragon way. Some of you don't have your laptops closed yet. And then we can navigate these guys in 3D.

Now, in addition to taking a bunch of snapshots around Gary Faigin's studio, what we also did was to take really high-resolution scans of the artworks and throw in (INAUDIBLE). And that means in some of these images we can, in fact, in all of the images on the walls, we can go down and look at the stitches in the canvas. So theses are 80-mega pixel images, at least. And that's true of all of these charcoal sketches on the walls and so on.

Let's go have a look at that one.

**AUDIENCE MEMBER**

(INAUDIBLE QUESTION).

**BLAISE AGÜERA Y ARCAS**

Sorry.

**AUDIENCE MEMBER**

(INAUDIBLE).

**BLAISE AGÜERA Y ARCAS**

Yeah. Okay. You're asking about the cloud of points, right? So that's right. If we zoom out here, and we look at what's going on, this cloud of points that's been reconstructed is, in fact, all of those features that have been solved for in 3D based on that collection of photos. So that's the 3D model. You can see that Gary's hair over here really got quite a number of features. I suppose because it's go so many objects in it.

Okay. So this shows you, I think, why it's interesting, why there are interesting sorts of vertical applications for this kind of technology. You can imagine all sorts of fun things happening with it. For example, let's see, let's pull up, oops, it doesn't work for me? Okay. Let's pull up this guy. This is an interesting commercial application, I think. It shouldn't really be installing at the moment of the talk, but it's okay. This is a small shop in downtown Seattle that sells kitchen and bath things. And what we did here is to take a bunch of photos inside the shop and hook it up to this web sidebar.

So as you move through the environment you get things like ecommerce showing up on the left-hand side. We imagine that this is going to be a really interesting way to do things like take stores that have invested a lot of work in design in the physical sense, but that, maybe, don't have the resources to go and hire web design firms and so on to go and design custom sites for themselves, but still want to have an online presence and exploit all of the physicality of their assets of what they have. It allows the authoring to be done entirely with digital camera. So I think that's a very interesting application. And there are a bunch of other really interesting applications along these lines.

But the real power of this stuff comes from what we get to do when lots of people take photos and start to synth them together. And when people are allowed to synth their own photos and also, take their own synths and start to combine them and match them up from synths from other people.

Now, I'm just pulling up Noah's original demo. I'm sorry, I should really be wrapping up, but I want to show this last thing, this last thing working. This is Noah's original data set. Sorry, it's very early build of Photosynth. We had this working in that same summer and we didn't have the user interface worked out yet. So you can see this craziness of white boxes. But anyway, that's all the photos of Notre Dame Cathedral culled from Flicker. And it included a wide range of photos. Everything from extreme close-ups of little gargoyles in the archways, all the way out to, let's see, keep going, this is all the way across the river.

And there are some interesting things in here I can't resist showing. One of the things that happens when you start to match photos with each other is that you get - let's see, yeah, the UI really hadn't been worked out, at this point, very well yet. Here we go. This is a poster of Notre Dame and, in fact, it's recurrent because it's a poster of Notre Dame on Notre Dame. But we connect from the poster to the cathedral, itself.

And that's really interesting when you can think about sort of hyperlinks that you can make physically by taking pictures of environments, and putting them up on walls and other environments, and that becomes like a kind of like a wormhole or tunnel. So the kind of authoring possibilities and the sorts of things that you can make with this are really pretty interesting.

For me, for my money, this is really the first time that I've seen something that I think is convincingly a platform that we can really build the Metaverse on and it has that same sort of deceptive quality that the web did. It uses off the shelf parts it uses nothing but the corpus of digital photos and other kinds of visual assets that we've been so busy collecting and putting online.

You know, the materials are all there already and, in some sense, all we have to do is start to put it all together. And suddenly very large parts of the earth's surface, and in particular, the most interesting parts of it, for us as creations of culture, will start to knit together and start to connect, and it'll be something emergent and it will come from many peoples images. So that's coming.

Photosynth community is going to be released sometime next year, in the first half of next year, I think. So we're really looking forward to seeing what happens around it. What people do with it. Thank you very much. I'm not sure if there's time.

(APPLAUSE)

I know I was supposed to leave five minutes, but I left five seconds instead.

**KURT ANDERSEN**

Sure. You've got three seconds now. No. Beam me up. That is - every time I see it, it makes me feel stoned. So um, have designers or artists, yet had access to the authoring sufficient to be able to figure out applications that they might do with this?

**BLAISE AGÜERA Y ARCAS**

Well, I think probably not enough yet. But we're still refining. We did a preview of the front end of the viewer, of course, but we only had a few canned environments to put up. And so we haven't yet released the tools that actually let you knit things together. Developing those and making that really work robustly has been a lot of what the past year has been about, so soon, soon.

**KURT ANDERSEN**

I mean, it just seems like it is ready, as amazing as a set of tools as it is, it's ready for 100 flowers to bloom to discover its power in ways that I certainly can't predict.

**BLAISE AGÜERA Y ARCAS**

We hope so. And I don't think we can predict them either.

**KURT ANDERSEN**

From Microsoft's point of view, I mean, you made the Bell Labs comparison; there's all kinds of things Bell Labs did that they had no idea how they were going to become important five, ten, twenty years later. Do you have a sense beyond the going into a store and having the ecommerce available on it, what kind of big practical applications there might be here?

**BLAISE AGÜERA Y ARCAS**

When I make these web comparisons, I guess, one of the things that I'm trying to say, is that it's a little bit like, you know, if you were a VC and I were at a startup company called Mosaic, or something like this, and I wanted to make a pitch to you well, what's the practical application. You know, of hyperlink text documents in 1989/1990.

**KURT ANDERSEN**

Right.

**BLAISE AGÜERA Y ARCAS**

You know, you can come up with a few things but it would look a little bit Jules Verne-ish. And you'd come up with some things and they'd be true, but they'd also be so much less than the

reality of what that's become. It's created an entire eco system. And I think in a similar way, this is about creating an eco system in which all sorts of things can happen, some of which we can predict and some of not. You know, when you start to hyperlink photos together and you start to create something that looks like the web, but in a geographic or geospatial form, you know, 1,000 flowers will bloom. I think that's a -

**KURT ANDERSEN**

Yeah. And the Notre Dame example is so amazing when it becomes this easily assessable global network of thousands of millions of images. When you mentioned the Photosynth community going online early next year, is that what that will enable?

**BLAISE AGÜERA Y ARCAS**

Yes.

**KURT ANDERSEN**

Uh huh. And the only sad and disappointing part of this, I'm sure, for most of the people in this audience, is that it's only available on Windows.

(AUDIENCE BOOING)

**BLAISE AGÜERA Y ARCAS**

Well, actually -

**KURT ANDERSEN**

Will that change?

**BLAISE AGÜERA Y ARCAS**

I know that this is a really sore point, especially with the design audience.

**KURT ANDERSEN**

Yeah.

**BLAISE AGÜERA Y ARCAS**

And you should - well, I can't make announcements right now about exactly what we're doing. But -

**KURT ANDERSEN**

There's reason for hope?

**BLAISE AGÜERA Y ARCAS**

There's absolutely reason for hope. I will say this, we're really interested.

(AUDIENCE CHEERING)

**KURT ANDERSEN**

Is this, in developing this and making it better and perfecting it, the focus of your work, or is there some next thing that in 2009 you will be able to show and blow everyone's socks off in a new fashion?

**BLAISE AGÜERA Y ARCAS**

Well, I hope, I hope so. Yeah. I have a few different projects that I'm overseeing right now. Live Labs has a bunch of things in the pipe that are very, very interesting. So I'm hoping that me and other people from Live Labs are going to have quite a few interesting things to show over the next few years.

**KURT ANDERSEN**

And I could be here another half-hour asking about what you're up to, but I am interested in, and especially for this audience, a thing I read about that you did outside of your work with this company, in which you researched Gutenberg's technology and essentially have a hypothesis that rewrites the history of printing.

**BLAISE AGÜERA Y ARCAS**

Well, a little bit of that. At least the first couple of decades I suppose.

**KURT ANDERSEN**

And if you could just explain that in -

**BLAISE AGÜERA Y ARCAS**

In 30 seconds or less?

**KURT ANDERSEN**

Yeah, exactly.

**BLAISE AGÜERA Y ARCAS**

That's tough. That's a whole other talk. But okay, I'll try and make it very short. Gutenberg didn't invent printing. There was woodblock printing before that, there was printing in China, but the thing that Gutenberg was assumed to have invented was the mass production of type, which it the font. That's what we call a font. You take a steel punch, you punch it into a copper matrix, and you cast thousands of identical lead types from that.

Now, when my collaborator, Paul Needham, and I went and started analyzing very high-resolution images of Gutenberg printing we found that the letterforms, in these earliest pieces of Gutenberg printing, are all distinct. You can actually look at every lowercase A and track it as it gets reused throughout the document and it's distinct, it's subtly distinct from every other lowercase A in ways that are consistent with wear or damage to the lead type.

**KURT ANDERSEN**

So typical forensics, basically.

**BLAISE AGÜERA Y ARCAS**

Right. And the outcome of all of that was that we postulated some mechanisms by which he produced these things in a way that was much more manual and much less mass production. In essence, every type, more or less was unique. And printing was a very evolutionary process. Inventing it was very evolutionary over its first 50 years where we don't have any documentation. It was all guild secrets, so there's no primary documentation of how it worked. And this assumption, that we used to have that the way Gutenberg did it in 1450 was exactly the same way it was still being done in 1850, was exposed for the ridiculous idea that I think it was all along.

**KURT ANDERSEN**

It's good to be reminded that it's all a work in progress.

**BLAISE AGÜERA Y ARCAS**

Yeah.

**KURT ANDERSEN**

Blaise, I want to thank you very much.

**BLAISE AGÜERA Y ARCAS**

Thank you. Thanks so much.