Science AMA Series: I'm Abe Davis, last week my research video about “Interactive Dynamic Video” (IDV) hit the front-page of Reddit, and a bunch of people expressed interest in learning more about it. So here I am, AMA!

This technology seems to have widespread potential applications, what are some of the areas you've envisioned it being utilized?

Is there any potential for a "real-time" implementation of the technique? Right now it appears you have to capture video of the object slightly moving and then extract movement during post-processing. How long does the processing currently take to isolate the object?

shiruken

There are a huge number of potential applications for this work - everything from entertainment, to structural engineering, health diagnostics, AR... It's a long list. However, some of these applications are closer to being realized than others.

Engineering: Our ability to recover vibration modes and frequencies may already be useful to engineers, who depend on this information to analyze structures but typically use lasers and accelerometers to obtain it. Cameras offer an inexpensive and easy-to-use alternative. The hope is that this will let us be more proactive about things like monitoring the health of decaying buildings and bridges.

Entertainment: The tech is already at a point where you could use it for cheap special effects. With better visual artists using the tool I'm sure they could do more impressive things than what I showed. The remaining barrier here is probably just user interface - I wrote all of the current code in C++/Matlab and made a bare-bones interface as a proof of concept.

AR: There is definitely potential here, which is what I was showing with the Pokemon video. However, getting this to work realtime on a handheld phone is really difficult. IDV is significant progress on one of the major challenges toward fully dynamic VR, but there are still others (like tracking/mapping/stabilization)

There are a lot of other applications too, but I'm sure I'll touch on them in other questions :-)

© et al. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.
Hi, your software seems to work with low frequency vibrations, 6/10/16 hz. Would there be a problem with objects vibrating at higher frequencies?

Your software uses footage from a normal camera, can you use other equipment (kinect, for example) to refine the end product?

Have you consider working with Face2Face (face reenactment) technologies? This could bring the cost of SFX in video production even lower.

Thank you, have a nice day.

wakimaniac

The frequencies we recover are limited to the nyquist frequency of the camera we use. If you use a 60fps cell phone, that will limit you to modes under 30Hz. This covers most visible motion. Though we have used high speed cameras to recover higher frequency modes (see our CVPR paper on visual vibrometry) and I've since simulated some of these with IDV (Higher frequency modes help if you want to simulate different physical properties by shifting mode frequencies down).

The video being referred to can be found here, and is worth the watch!

The ‘touching’ of objects in the description does not refer to tactile feedback.

Ax3m4n

Good point on the tactile feedback. It’s a subtlety that can be hard to explain. “Touching” something in real life causes two responses, the object responding to you and you responding to the object. We address the former, but not the latter (you essentially poke the object with a virtual stick). It would be cool to try and combine this with some sort of haptic system though...

My mind was blown watching your demo, and then doubly so when I discovered you’re still a pre doctoral student! Great work. Any tips for getting the most out of your phd?

NSFForceDistance

Thanks! I’m finally graduating in September after 6 years, and I’ll be starting a postdoc at Stanford in the fall :-) Picking the right problems is really important, and it requires the right mix of pragmatism and wide-eyed optimism. Those things can be hard to balance.

Is this a 2d effect or can this be translated into 3d somehow? (Or maybe it already is? Hopefully this question makes sense.)

SpookyGhostThing

A lot of its strength comes from not depending on 3D information - which means you don't need to know the geometry. If you have the geometry you can use the two together to greater effect. This is something I'm hoping to work on in my postdoc.

IDV alone doesn't give you 3D though. Part of why these results are so impressive is that we sidestep the problem of 3D reconstruction, which is very hard, and which people thought you would need in order to do what we do. In fact, when I first showed other researchers these results, a common
reaction was "Wow! How did you get such a good 3D model for that bush???

What problems in graphics still remain? Do you have an interest in something to solve next?

disco_arf

The graphics research community has proven extremely adaptable over the years. It used to be that graphics was mostly rendering and ray tracing, but faced with the prospect of becoming victims to their own success, the research community seems to have really embraced other ideas and applications. Visualization, human computer interaction, computational photography, vision, fabrication… today, SIGGRAPH (the top publication in graphics) seems to be more of a standard than a limited group of topics. This makes graphics a very exciting field. It’s wide open, so it’s a great community to try creative off-the-wall ideas (The Visual Microphone is a good example of something that isn’t "typical" graphics, but we published at SIGGRAPH)

This is pretty neat! I am curious as to how the program is able to identify objects, especially in a messy visual scene. According to my understanding of the video, it appears as though only a single object (e.g. the bush or the jungle gym) was targeted in each scene for IDV. However, in a natural environment and in your videos, I am sure these tiny vibrations occur in all objects in the scene (e.g. the planter or other parts of the jungle gym). Does your program distinguish between these objects and target a specific part of the scene, or does it compute the range of motion for all parts of the scene?

I am also curious about the scale of the forces in the video. Am I correct in assuming that an abnormally large force would make objects appear "rubbery"? For example, would a strong breeze make IDV difficult to create?

Thank you for your time and for the cool research!

Three_flower

Thanks! In some cases the object can be isolated by the frequencies at which it vibrates. While a lot of motion could be going on, individual objects tend to have something like a signature in the frequency domain. By only using certain frequencies we essentially filter out the objects that move at other frequencies. This doesn't always work though, and in some cases we just apply a mask to the image. The segmentation is easier if you use the spectral info, but applying a mask to one image is also pretty easy.

I should note that for objects that have a lot of damping, whether they are in the background or not, you often need a mask (because their frequency response is broad spectrum)

Does your method generate a kind of 3D model based on observed movements in the source video, then map the texture of the moving object to that 3D model? Or does it instead warp the texture in order to produce the movements? It looks a bit like the latter, but I couldn't be sure and wanted to ask.

This looks like it would be great for quick and easy displays like in the example video, but doesn't seem to scale well. That is to say that, past a certain point, it doesn't look like you can keep throwing more data at it to make it more realistic. What do you think are the biggest limitations of this tech?

super6plx

One of the technique's biggest strengths is that it doesn't rely on a 3D model. That's why I can walk outside my apartment and capture one of these things with just a cell phone and a tripod.
However, it is true that we deal with oscillations around a rest state, which limits the current technique to motion that objects will “bounce back” from.

This info could compliment other techniques very well though. For example, in physically-based animation of 3D characters vibration modes are used on top of articulated models to add physical behavior in response to articulated motion. In this sense, there is reason to believe IDV is ripe for use with other types of information. We also showed how other physical properties can be estimated when you have a prior on geometry in our CVPR paper about visual vibrometry.

I know you mentioned that this technology could be employed to create low-cost interactions between CGI and the real world. Have any companies approached you about this proposition and do you expect any to?

User0218

Honestly, I’ve gotten a lot of emails in the past two weeks and I’m a bit behind on reading/replying. Hopefully I’ll be able to catch up after I hand in my dissertation (soon) :-p

I’m sure this technique will make its way into some special effects/visual art software, though which and when remains to be seen

Can this be used to give inaccurate physics to certain objects? (such as a springy/wiggly Effiel Tower)

NoobRising2

Absolutely. In my implementation I have sliders that let me edit the stiffness of an object, and it’s damping (we make a guess as to what the real values are, but can change it for visual effect). This is great for entertainment applications. For example, I made the wireman at the beginning of the video much springier than he is in real life.

Oh boy. Do you see this DEFINITELY ending up being used (besides entertainment) for the porn industry? After all, VR wasn't immune.

certifiablenutcase

Oh man, after the top comment last week I thought this might come up.

About a year ago, while I was developing the technique, I had the idea that you could use it for health tracking/diagnostics. I figured you might be able tell something about a persons body fat and distribution that could be useful. So I put my camera on a tripod, took off my shirt, pressed record, and did a little truffle shuffle... for science...

For those of you who don’t know the reference, here is an approximate visual:

http://i.imgur.com/yaSat.gif

So then I tried looking at the vibration modes, and simulating my own body fat. And itkind ofworked. I mean, it wasn't perfect but it was compelling enough to see that it could work. I even included it in the supplemental material to our ACM TOG submission. I also had sliders that could change my material properties to make me more or less... jiggly.

So, honestly, I would be surprised if someone didn’t use this for... less publicly acceptable... purposes... but regardless of where you stand on that, using this technique to find physical properties of the human body could have real benefits in diagnostic health.
Hi Abe,

Was really impressed with the video you showed, what a wild piece of technology. So, pardon me if this question is naive: you showed that to create your model, you analyze the vibration modes of an object given a stationary camera. Is it possible to replicate the model given a moving camera and a stationary object?

Zeraphil

Thanks! If the object doesn't move then we won't be able to learn its vibration modes. With really good tracking we might be able to get the technique to work with a moving object and camera, but it would be harder. I've been able to compensate for some minor handshake, but not someone walking around with a camera.

Are you worried about a future where people can manipulate video just like we already do today with pictures via photoshop? I am unsure about a future where we won't be able to distinguish between what is authentic and what has been manipulated.

iPreferPeace

I think we will get to a point where humans have a hard time distinguishing, but clever algorithms will be able to.

Already there is a fair amount of research going on that looks at how we verify images and video. It's a bit of an arms race between the manipulators and the forensics. It's kinda fun to watch ;)

Are there any plans for making this research into a vfx kit for video-makers in the eventually future? Also how did you get into computer science as well as what is the programming language you use the most often?

LionOfMyth

I'll be doing a postdoc at Stanford next year, then hopefully applying for faculty positions. I might help someone else create a product, but it would be part time. I love research too much to quit and work full time on a company right now.

I got into computer science when I was in high school. My school didn't teach it at the time, but I got a book, and asked a prof. at Johns Hopkins if I could work on the computers in his lab (I'm from a Baltimore city public school that offered independent study). I used to be really into video games (I still enjoy them but don't have as much time to play) so I was really interested in graphics. I also worked at Firaxis Games for two summers in college (It's where Sid Meier is/ they make Civilization).

I mostly use C++ and Matlab, though my general philosophy is to pick up languages as I need them.

Do you see this technology being used for real time communications? In my own work, I've looked at things like FaceTime and Skype but informants always note that sensation is a big part of what is missing. Not only for interacting with a long distance romantic relationship. But also things like hugging their kids when they are away on a work trip. Or feeling the breeze when they are using tech to "hang out" with a friend at the beach. Seeing faces is something we're biologically primed to want but that
tactile piece is very important too.

Right now this technology is like a little window that connects two geographically disparate places. But the glass between cannot be broken. Could this technology be used to simulate a communication experience that feels more present? Where, for a moment, it feels like it isn't a window but that conversants are side by side?

**firedrops**

One of the exciting things about this work is that it offers cameras as a low cost, ubiquitous tool for learning about the physical properties of objects. This could help with what you are describing, but it would only address one of many challenges facing that kind of application. That kind of general haptics, in particular, is tricky.

A haptic system could use IDV for input, but I think that's just one piece of the puzzle.

Could this simulation be used to work out the force applied to an object through its movement?

**J_tt**

That's a good question, and something I've been thinking a lot about. There are fundamental ambiguities if all you have is video, but I think there are still interesting/exciting things that can be done. I'm not sure I understand the second part of your question, but certainly the first related to a very complete understanding of objects in video, and would require resolving some of the ambiguities in our current implementation.

What kind of weaknesses do you see with the current technique and how do you see them being improved upon in the future?

**TakingItCasual**

Because it works on a particular scale and type of motion, and it assumes a stationary camera (on a tripod) it's hard for this technique to work "by accident." What I mean is that if you download a random video from youtube, most of the time the technique won't work. Changing that would require addressing a huge number of challenges. But that's what research is for :-)

Hey thank you for doing this AMA.

Your work seems really interesting, I just finished my first year at University, and I would like to work on something similar (live video analysis etc.), once I learn enough. Would you recommend me some resources or what to focus on, while learning? Thank you.

**TheAzgra**

Math, computer science, maybe some physics? Linear algebra is super useful, and it's often not taught very well. It's worth learning and re-learning, because it ends up being so useful in so many different contexts. I re-taught myself linear algebra in grad school, despite taking it as an undergrad, and found the experience (of relearning it) very useful.

How do you actually simulate the movement of the object?
TheFirstOf28

We decompose the movement we see in the input video into independent modes. We then take an image of the object (e.g. a single frame from the input video) and warp it by new combinations of these modes, in different mixtures and amounts. Does that help at all?

We also have a paper about it here: http://www.interactivedynamicvideo.com/ISMB_Davis_2015.pdf

Everything about this screams villain movie.

Young brilliant scientists working at a cool sounding acronym making breakthrough discoveries that will help advance technology and society.

Your lab is broken into late at night and your tech is stolen by a villainous entity.

How are your devices like the IDV and visual microphone modified and used for unintended and evil purposes?

BobbinWilliams

Look to the internet and you shall find your answer. We published the visual microphone and people yelled "Big Brother!" We published IDV and people yelled "Boobs!"

Now, I wouldn't necessarily call either of these applications evil (it would depend on how they are used), but they aren't the first thing I'd list in an NSF proposal...

What do you think about implementing this kind of technology into virtual reality? It would make the experience much more real since you could have virtual objects interacting with the physical things around you. Also, where could the average college student find more information about how to do interactive dynamic video? This is a truly exciting project!

Kizzle

Thanks! I think there are a bunch of cool things you could do with it in VR. Those applications are probably a bit more down the road than SFX and engineering, but I'm sure we'll get there.

We have a publication about the work here: http://www.interactivedynamicvideo.com/ISMB_Davis_2015.pdf

It might be a bit hard to follow if you haven't studied the material before (I didn't understand a lot of this stuff in college). Math, computer science, physics, and signal processing all help a lot.

Hey Abe! Really great and impressive work! I do have one question however: how resource-hungry is this technique? Like, do I need a beefy computer and a lot of spare time to process 5 minutes of video or can it be calculated in real time on a smartphone?

browniebrian

It depends a lot on the video, and the mode frequencies of the object. Some videos process in a couple minutes on my laptop (like the wire man), but some ran on the server for a couple hours (we tried it on several minutes of HD video, which was slow...). This could all be made faster, as it's all unoptimized matlab code right now, but it would take some more engineering.
aside from the obvious pornographic implications, which industries do you foresee gaining the most from this emerging technology, be it recreational, film, games, or other medium?

MaxwellVador

I don’t know if they will gain the most in the long run, but I think special effects and engineering (e.g. structural health monitoring) will probably be the first to benefit. In its current state IDV may already be valuable in those areas. Other things, like dynamic AR, won’t benefit as much until other technical challenges are also addressed (things like tracking and mapping on a phone), though the eventual payoff could be great.

Do you have a peer-reviewed journal article on the IDV technique?

nbo10


The videos have newer examples, and there are a few subtle tweaks, but it's mostly what's described in the paper.

I wanted to change the name after it was accepted, but they wouldn't let me. I'm calling the chapter in my dissertation "interactive dynamic video" though, and it will have some of the newer examples in it.

What do you think about the computing overhead with real-time, virtual/augmented reality? Would reducing resolution through image conversion help?

calloustt

Reducing resolution helps with processing time, but here is also the issue of input data. The more input video you have, the more information, and the better the simulation. There are things to try that might bring us closer to realtime, but there are a lot of challenges and I think working on less input is maybe even harder than just processing time.

In your TED talk, you show how you got sound and touch senses from a silent video. Do you think it will be possible to use a similar technique to allow a viewer to pick up on smell or taste of a video?

invisiblesimmer

I'm not sure how it would work... In fact, I don't think we have any devices (even expensive, specialized ones) that are very good at this right now. It would be cool though... (gazes off in moment of deep speculation)

Hey, /u/abe-davis, maybe this is interesting to you:

[https://hci.rwth-aachen.de/dragon](https://hci.rwth-aachen.de/dragon) (direct-manipulation interaction technique for frame-accurate navigation in video scenes)

You click on an object, and you can play/reverse the video by dragging the object along its path.
There also is a demo video, and a demo executable (OSX only).

callosciurini

Cool, thanks! I'll check it out post-AMA

Quite amazing! I recall the potatochip mic from some months ago, but was totally not expecting the wire man! Very cool. Also very nice to see a fun shift from typica "surveillance" and all that dark stuff. I would love to play around with such an animation tool - perhaps not even made from actual video, but from computer generated animations.

hatyn

Thanks! Yea, and this time nobody seems to be asking me about JFK's assassination! At least not yet... until I post this comment... then I'm sure someone will find a connection...

I am headed into a PhD in CS and I was very impressed when I saw the video. I assume that this is result of tons of hard work, therefore, I'd love to know if you have ever struggled with staying motivated? How do you organise your time and balance work/hobbies? I wish you good fortune in the papers to come!

rndmusr12

Thanks!

It's a constant struggle, and I'm not the best at time management. I've been getting better though, and this work is fruit of some of that progress.

Grad school can be really hard. It can be isolating, and easy to get discouraged at times. I went through a tough period a few years in that coincided with some health issues, and it really shot my productivity. I sat down and decided that I needed to form better habits, so I made a list and started making changes. Eating better and exercising was a big one, and just being more disciplined with myself in general.

Another tip I'd say is that, if you can, try to work closely with someone more experienced early on. Miki Rubinstein, my second author on the visual microphone (he also has a TED talk) taught me a ton about how to do research, and helped me get enough momentum to be confident I could execute projects more independently (esp with ideas that seem crazy at first).

Good collaborators are really valuable in general, and once you find them you can keep working with them. It also just makes the experience more enjoyable. Bonus points if you can find people with skills that complement your own. Justin Chen (/u/JustinGChen) is another good example of all these qualities, which is why I keep working with him :-)

I don't know much about academia in CS but I was just curious about how difficult it is to secure a faculty position/ number of postdocs required before people even try for tenure track. In general, what is the typical path like? Do most people go into industry instead? How rare are faculty jobs etc?

lostthesis

In general it's pretty competitive, but I think it depends on the kind of job you want too. Top tier research professorships are super competitive, but jobs that focus more on teaching are less insane.
I love research and enjoy teaching, so I want to be a research faculty. I'm doing a postdoc at Stanford before applying for prof positions though. Postdocs don't seem to always be necessary in computer science, but for me I think it will be a good opportunity to smooth the transition from student to (hopefully) faculty.

I would consider industry research, and might consider starting a company down the road, but right now I'm interested in faculty.

Hey Abe,

I'm an undergrad thinking about doing a PhD in a field close to yours. Related to that, I have some questions.

What were your reasons for wanting to do a PhD? How competitive is a graphics PhD at a big school like MIT? Any advice for somebody starting one work ethic/expectations wise?

4dahalibut

I did a PhD because I really love research. Especially in graphics and vision, I love the mix of creative freedom and technical rigor.

PhD programs at places like MIT are very competitive, and the expectations are very high. That's part of their value though. If you really love research it can be worth it, but a lot of people find that's not the case and end up struggling with grad school. People around here are very smart and talented, but grad school is a marathon, and the more common challenge people seem to face is motivation.

A big tip for grad school is to work with good people. People you enjoy working with. The people you work with have a huge impact on the experience that you have, and that can be for better or worse. Fortunately, my experiences have been mostly for the better.

It's worth noting that while I've been pretty successful with publications for the past couple years, I try a ton of ideas that don't pan out. That can be discouraging for a lot of people - and its ok for it to be discouraging. If you don't enjoy it (for more than a passing slump) it might not be worth doing though. Especially in computer science, the opportunity cost of a PhD is huge, and academia is not for everyone.

All that being said, I love research, so I look forward to continuing it :-)

Good luck!

Hey Abe, I'm about to graduate with a bachelor's in Cognitive Science and I really want to work and develop artificial intelligence. Most places require previous work experience. Where would you suggest working for a preliminary AI job?

notesonblindness

Hm, I'm not sure. I pretty much stayed in academia since graduating, living off of a tiny stipend and the occasional internship so that I can work on whatever I want.

I suspect that starting out as a programmer, and being very proactive about moving toward more AI-oriented stuff is one strategy. It would probably help to learn material on the side, maybe by taking classes. For this you would want to pick a place that is actually doing some AI stuff though, and it might be good to be up front about your desire to move in that direction. I'm probably no the best person to ask about career advice in industry though. Good luck!
Hi, the results of your work are quite awesome! As a CS grad student, this looks like a very interesting technique. Is there a paper that goes into more detail of how the technique works?

hahawin


Will this software eventually be licensed out? This could be huge for hobbyist youtubers and the like to make far more convincing cgi videos.

Bald_Sasquach

Probably, but it could take a while. I'm planning to keep doing research, so hopefully someone else will license it and turn it into a product (or a tool in a product, e.g. Adobe premiere)

It is a lot of fun to play with. It's kind of a pain right now though, because I only wrote a very bare-bones UI to show it works. Personally, I've wanted to make a music with it for a while but I haven't had the time.

Firstly, where did you get your bachelors and where would you like to see this technology most utilised?

InsidiousIntentions

I did my bachelors at Stanford, and my masters and PhD (almost done) at MIT.

I'd like to see it used as a general imaging technology. Something like RGBD cameras, where people see it as a tool for building lots of other technologies.

I think it could have a lot of impact in nondestructive testing of structures. I personally enjoyed playing with it for little special effects things, though it was a bit tedious with my little home-grown UI.

Are there any plans to make use of this tech using VR?

Scarbane

No concrete plans right now, as I plan to stay in academia and keep doing research for the time being. Someone else might - possibly one of my collaborators or a company that licenses the tech from MIT.

What were some of the biggest challenges in developing the IDV system?

ImmaDuuck

In some ways the technique is surprisingly simple, but it draws on ideas from a lot of different areas, which I think might be why nobody had thought to do it before. The related work section in our paper alone covers computer vision, physically based animation, and a few different subfields of civil mechanical and structural engineering. I had expertise in some of these areas, but not others, which meant learning a lot of basic material from other communities. Justin (/u/JustinGChen) helped a lot with that, as he works in civil engineering here at MIT.
It's there a tactile feedback for the user?

7h476uy

No, not at the moment. It would be interesting to use this with some kind of haptic system to give tactile feedback though.