Mission-Practical Biometrics

featuring Satya Venneti as Interviewed by Will Hayes

Will Hayes: Welcome to the SEI Podcast Series, a production of Carnegie Mellon University Software Engineering Institute. The SEI is a federally funded research and development center sponsored by the U.S. Department of Defense and operated by Carnegie Mellon University. A transcript of today's podcast is posted on the SEI website at sei.cmu.edu/podcasts.

My name is William Hayes, and I am principal engineer at the Software Engineering Institute. Today, I am joined by my colleague from the Emerging Technology Center, Satya Venneti, who is here to talk with us about her latest research in real-time biometric data acquisition. Could we start with a little bit of background on where you have been in your career and what brought you to this location?

Satya Venneti: Thank you. I have been in industry for 18-plus years. I was always interested in research. About one year and four months ago, I joined the Software Engineering Institute. It has been really great because I feel that I am a liaison between the university and the government and industry. That is where I always wanted to be: to take something that has been already developed in the research side and find practical applications for it, especially in the government sector.

Hayes: With the research we are here to talk about today, you are really very much on the leading edge. Could you just [provide] a thumbnail sketch of the technology we are here to talk about?

Venneti: Sure. I want to talk very briefly about biometrics in general. Biometrics is the science of measuring and analyzing biological data. Really when we talk about biometrics, there are two distinct phases: active and passive biometrics.

We talk about active biometrics. It requires when we actually collect the data and it's invasive and we need to have contact with the subject. And nowadays, we're talking about passive biometrics where it's completely non-invasive and we don't need any contact with the subject. And the analysis is done in real-time. So there are all kinds of applications, like machine emotional intelligence and behavior prediction and sentiment analysis in addition to all the traditional things, like authentication and identification, which active biometrics is used for. We are really looking at passive biometrics, and we are looking at how we can use that to increase machine emotional intelligence. By that, I mean how machines can detect, understand, and respond to their user's emotions in real-time

Dr. Andy Moore, who is the Dean of the School of Computer Science at CMU, predicted that 2016 is the watershed year for machine emotional intelligence. We are really on the cutting edge of technology here, and that is what I do my research on.

Hayes: If you could, for the layperson, give us some detail on what the technology is and how it comes to be.

Venneti: Sure. What is out there today that is the state of the art deals with practical lab settings? That is, it doesn't deal with motion. What we wanted to do is we wanted to deal with faces in motion with expressions and changing illuminations.

What we did for that is we took a technique called facial landmarking, which takes a face and marks 68 points on that face as features or landmarks on that face. As the face moves from frame to frame in a video, [facial landmarking] tracks the face. It can deal with plus-minus 90 degrees of motion along the pitch, yawl, and roll directions. It can also deal with occlusions, which is covering up the face, either intentionally with my hand or facial hair. It can also deal with changing illumination.

What it does is it gives x and y coordinates for each of these points along with confidence numbers. What is out there that is the face landmarking, which is the state of the art, is slow. It's robust, but it's slow. What we wanted to do is get it fast enough to happen in real time.
So, we implemented it on graphical processing units--that is, GPUs--and were able to bring it from about 20 seconds per frame to .2 seconds per frame.

Now we have open source facial landmarking out there that's even faster than 0.02 seconds-per-frame, but that's not as robust because it can't deal with as much emotion or occlusions, but it doesn't give us confidence numbers. What we did is we got the state of the art fast enough that we can now use a hybrid of open source and the state-of-the-art facial landmarking. So, we run them both in parallel and we use the open source landmarking, but every so often we do a sanity check with the state-of-the-art but slightly slower, more robust, landmarking system. That is what we did. We did a hybrid approach of the open source which is much faster but we got state of the art fast enough by putting it in a GPU so that we can combine these two. That is how we make it happen in real time. We get the best of both worlds. That is, we get robustness, but we make it happen fast in real time.

Hayes: The reliance on GPUs allows you to lower the technological barrier so that this can be done without a very expensive lab setting. You can do this in a much more reachable…

Venneti: Correct. Smartphones right now are being equipped with GPUs. The iPhone 6 and I think the Android—it's called the HTC Desire or something like that. Both these phones have GPUs on them. So GPUs are becoming more and more embedded and widely available technology.

Hayes: If we think about the potential for adoption of your research results, this lowering the technological barrier really speeds the ability for the industries that might find benefit from this technology to adopt and run with it.

Venneti: Yes, I would think so.

Hayes: So this is very much about how humans interact with computers and maybe opening new channels for input.

Venneti: Absolutely, yes.

Hayes: And more dynamic interaction.

Venneti: Yes. It's like teaching the machine emotional intelligence and body language and things like that which is how we, as humans, interact with each other.

Hayes: Maybe there are even applications for this in amusement park rides. We detect that the motion of the ride has a particular effect on the riders.

Venneti: Right. We can think about things like games and training and things like that. If you are doing military training, you can see how the person is responding to something in real time looking at their heart rate, for instance, seeing how it increases or decreases and things like that. You can ramp up or ramp down the game corresponding to how the user is reacting.

Hayes I remember some time ago going through an airport in Asia where there was an outbreak of flu. They were monitoring the temperature of people coming through to offer assistance to people who might be ill. So, there are safety related applications that go beyond that crude measure that we saw there. Could you talk a little bit about that?

Venneti: I actually think the medical applications are amazing. I am from India. If you look at all the developing countries, more people have access to smartphones than they have to healthcare. You can imagine how something like this, if you implement it on smartphones, would really revolutionize something. We are looking at actually putting what I am doing on smartphones. There are all kinds of medical applications.

There are also things like if you are looking at the rate of flow of blood in the face, and you detect that it is flowing at different rates and this half of the face versus this half, that might be an impending stroke. There are all kinds of medical applications. Also, I recall the Ebola virus when it came, people were looking at body temperatures, and now you could just get that from video.

Hayes: In the application relating to strokes, as I understand it—limited as my understanding is—the subtleties of diagnosing very minor strokes, this would be a way to break a new frontier in how much clarity we can get there.

Venneti: Absolutely, because you are looking at different regions of the face and seeing how the heart rate or the blood flow differs in those regions.

Hayes: I watched a video that is referenced in your blog post about baby monitors. I remember being a new parent, and looking in on the child to make sure he was still with us, and being nervous. It was fascinating to see how the algorithms allow us to magnify subtle movements and detect, *Indeed my child is breathing and moving in subtle ways*. That seems a more passive application where we're monitoring. Could you speak a little of those kinds of applications?

Venneti: Sure. So, first of all, our work is inspired by work that came out of MIT CSAIL, which stands for Computer Science and Artificial Intelligence lab. They came up with something called Eulerian Video Magnification. They say that it makes the invisible visible. What it does is it takes latent frequencies that are already there in video, and it magnifies them so that we can see that with the human eye. One of the things they did is they went to the neonatal unit, and they [recorded] babies. You don't really want to strap anything on them because they are so small and tiny and fragile, and they were monitoring their heart rate through video. So, that is one application.

You can also think of things like when you have sensitive meetings, say in embassies and consulates, where you are looking and trying to see how a person reacts to questions or something like that. Also, at all these border patrols and security checkpoints, you can be monitoring people for safety issues.

Hayes: It seems the pace at which we can digest and interpret the data is the frontier where you are really focusing your work that making this a real-time application is a big challenge. Could you talk about that?

Venneti: Right. So, there are actually two big challenges. There is making it real-time, but there is also making it practical. If you look at the state of the art right now, it works very well with lab settings where the person is extremely still, no facial expressions, and the lighting is exactly how we want it to be. But that is not how it is in real life, right? People are emotional all the time. Their facial expressions and the lighting is changing.

What we are trying to do is make the recently possible practical. We are taking things where we want to do this in natural settings where people are moving, having different facial expressions, and even occlusions, like I might do this or I might do that [covers forehead with hands] and you no longer can see this region. That is what we are trying to do. We are trying to take something that is already well established in the research field and make it happen in practical settings.

Hayes: Very much in keeping with the theme of emerging technologies, which is the center where you work. So, the practical applications you see in the near term, the customer set who you see adopting and running with this. Could you talk a little bit about that?

Venneti: Right. I think one of the things I am excited about is machine emotional intelligence. In our center, we deal with autonomy. When you look at human-machine interaction, really, we talk about how machines and humans can live together almost symbiotically. There is an example where I might have a robotic chef that is helping me in my kitchen. The chef might actually pick up a knife to do something, and I might get a little scared because I don't know why it is doing it. If the chef can see or if the robot can see that my heart rate has escalated, it can either explain why it is doing it or move back a little bit. That is how humans and machines can work together in tandem. Our work actually helps machines and humans work better together.

Hayes: You just made me think a bit about the black boxes we have in aircraft or other transportation mechanisms. We want to be able to reconstruct the final seconds before some disastrous event. It would seem there may be applications here where we can understand what the pilot was experiencing at the time and help us put things together.

Venneti: Absolutely.

Hayes: Is there any development along those lines that you could talk about?

Venneti: I think where we would like to take this research is to answer things like how can we transition it to the real world—maybe put it on different platforms—and then deal with also distance, for instance. So, how far the camera is from the person's face. Doing multiple subjects at the same time. We talk about scalability where you want to increase that and portability and then usability.

Also, things like right now, we have certain configuration parameters that we have to tweak depending on the subject. *My heart rate can be faster than yours* or depending on the speed of the camera or even depending upon the application. For instance, you can think of polygraphs as an application. *How big should my sampling window be? Do I want to detect peaks in five-minute intervals and five-second intervals? How fast am I asking you the question?* We want to automatically adapt to the human subject, to the surroundings, to the camera, to the application. We want to do research in how maybe we can do some learning and figure this out, maybe deep learning and machine learning and then *How do we also transition it to practical applications?*

Hayes: Is there also an integration with multiple modes of data collection that there may be other streams of data to be integrated with?

Venneti: That's a great point. If you look at things like audio, for instance, right? I think there is research which shows that the picture of the voice can show the emotion of the person.

If we look at facial expressions and gaze and posture, and we also look at lexical analysis, for instance. So, I am talking and maybe the words I am using and the heart rate I am looking at. If you can combine all these together that would be great. That would be like a multimodal emotion analysis platform. We would really be able to teach the machine how to do body language and how to emotionally understand its user.

Hayes: When we think of multi-factor authentication for people having legitimate access to systems, the possibilities here are fairly profound I imagine.

Venneti: Right. I think that there is research that has shown that every individual has a unique cardiac signature. Imagine if we could do something like identify you just by your heartbeat or look at your ECG and completely non-invasive.

Hayes: And the door opens as you approach it.

Venneti: Yes, that would be fantastic.

Hayes: Fascinating.

Venneti: There is not much you can do to fake that.

Hayes: I saw in the reading that we have machines that can analyze the gait, the pace; your posture even can speak about who you are. That is really fascinating that we can go this far in real-time, and that is quite an ambitious frontier.

Near term, what is your research targeting? What can we look to see from you coming soon?

Venneti: I think, as I said, I would like to make this more automatically adaptable to, as I said, the actual subject and the surroundings and the application. I would also like to look at engineering challenges: how to transition it to practical settings on different platforms. We are really looking at things like smartphones. The way we have made this faster is to make the algorithms run on GPUs or graphical processing units. Nowadays, more and more phones are equipped with GPUs. We have already made some initial experiments on actually porting it to iPhones and the Android phone.

I am trying to see if we can combine multiple feeds, like thermal feeds, because then you can use it at night and see if we can do work with facial landmarking on a thermal feed. I am really excited, but I think one of the most exciting things for me, as you said, is to combine different input from different sensors and combine all that together and have a machine that really understands the human.

Hayes: I understand you have brought a video that demonstrates this technology. Could we have a look at that?

Venneti: Sure. Absolutely.

Hayes: Oh, great.

Venneti: [Plays Video] This is a little demonstration we did of my colleague Chase Midler who has worked on this. You can see he is in a very naturalistic setting. He is drinking coffee. He is talking to us. You can see his heart rate is around 64.8 beats-per-minute. That is a resting heart rate. We have a few graphs that I will talk to you about.

Chase is going to get up now and do 50 jumping jacks. You can see him doing the jumping jacks here. He actually has a FitBit and he was talking. We cut the audio but he figured out that this detects his heart rate, his increased heart, rate faster than the FitBit. So the FitBit took a while to catch up, but you can see that his heart rate now is 79.2 beats per minute. There are three graphs. The first one shows the raw pixel intensities over time. The second one shows the trend of the mean heart rate over time. The last one is the peak. It is now slowing down, and that is all we had.

Hayes: You have worked with a number of collaborators. I think you've mentioned at least one so far. Others that you would want to talk a little bit about the work you're doing.

Venneti: Right. So I have worked with Professor Mario Savvides, who is the director of biometrics at Cylab. We also coordinated with Dr. Kris Kitani, who is a professor of computer vision at the CMU Robotics Institute. He really helped us kind of get some of the computer vision algorithms.

Hayes: So, bringing multiple streams of research together, another testament to the Emerging Technologies Center. That really is a nice opportunity.

Venneti: Yes. One of the applications that I am also excited about, that I forgot to mention, is we can use this to actually detect doctoring of videos. As the technology develops, there are also bad people who use this for their own purposes. When I went to a conference earlier this year, there was a group from Stanford and The Max Planck Center for Visual Computing and Communication and they showed, I think it's called Face2Face, where they can in real time put their expressions onto a video.

Now, if we look at the original video and the doctored video and we look for heart rate in different regions of the face, it is very easy for us to find that this video has been doctored because you can see that you get completely different heart rates at different points as compared to the original video.

Hayes: So you can defeat intentional deception. As we look at the way communication occurs today, there is more and more face time. There's, you know, video-based. My children's use of Snapchat with all the different features. You can defeat attempts to doctor that. That is fascinating. Satya, thank you very much for joining us today. This has been very interesting.

Venneti: Thank you, Will.

Hayes: I hope you will come back and give us an update as your research progresses.

Venneti: I hope so too. Thank you.

Hayes: Satya's also written a blog post, which is available on the SEI's website. We encourage you all to explore that. We will include resources relating to publications from Satya and her colleagues with the post of this podcast.

This podcast is available on the SEI's website at www.sei.cmu.edu/podcasts. It is also available on Carnegie Mellon University's iTunes website as well as the SEI's YouTube channel. As always, if you have any questions, please don't hesitate to reach out. Thank you.