COVID Information Commons (CIC) Research Lightning Talk

Transcript of a Presentation by George K. Thiruvathukal (Loyola University Chicago), September 22, 2021

Title: RAPID: Leveraging New Data Sources to Analyze the Risk of COVID-19 in Crowded Locations

NSF Award #: 2027524

YouTube Recording with Slides

September 2021 CIC Webinar Information

Transcript Editor: Macy Moujabber

Transcript

Lauren Close

First let me turn it over to George of Loyola University of Chicago who's gonna give us this first presentation of the afternoon.

Florence Hudson

Wonderful.

Lauren Close

I'm gonna stop- George and I'll let you share your screen.

George K. Thiruvathukal

Okay, thank you so much.

Florence Hudson

And while he's bringing that up we just like to encourage people if you do have questions for the speakers feel free to put them in the chat. They can answer them asynchronously or we'll have an open Q&A session at the end. Go ahead, George.

George K. Thiruvathukal

Okay, if I can just get confirmation that you're able to see my slides only.

Florence Hudson

Looks great.

George K. Thiruvathukal

Slide 1

Alright thank you so much. First of all, when you get introduced as a fantastic speaker, it's kind of a tall order to live up to but thank you for that very kind introduction to all of us, Lauren. And I'm going to just dive right into this talk in the spirit of lightning talks. So, I'm going to speak about “Observing Human Mobility During COVID-19.”

Slide 2

Alright and our work is very much about being able to take advantage of what we call public cameras, public network cameras, and analyze this large- this vast amount of visual data captured from these cameras to look at how people are following social distancing during various stages of the lockdowns in various locations. We actually looked in this pilot study at just five countries and three states to analyze the effectiveness of these lockdown policies to see, you know, if we could- if this can be a tool in the toolbox to you know help reduce the spread of COVID-19. And the main challenge of the study is that we have millions of images- in fact, as of today we have probably nearly 7 million images. And then what we want to do is we want to be able to take those images, and we want to be able to look at where these images are located and basically look at the policies that are in those locations and be able to see if there are any patterns in the data. So, this is very much a big data project and I want to say that I usually tell people: do not try this on your laptop at home. It will probably not work, and you may overheat your computer within a few minutes of the study.

Slide 3

So, without further ado. So basically, this slide here is showing just some samples of the kinds of images we have available. These are all again coming from you know public cameras that that are available for-you know, they're on the net, you know, basically, typically, like governments and individuals or other organizations like national parks have these cameras and they embed them on their web pages. They want to make them available to others, not necessarily for research, but it's public- they're public cameras. So, unlike the mass network of closed network cameras that exist in many municipalities, for example, these are sources that can you know legitimately be used for study without raising significant privacy and security concerns. Alright, so in our overarching research project, which is the Cam Squared project, which I had mentioned on the title slide but didn't say too much about it in the interest of keeping on time, we actually have a process for discovering these cameras. We have, you know, papers on that topic. And the study we've done is basically where we capture data continuously from all of the cameras in our camera network so about 36,000 cameras that we have discovered through this automatic process that we created and what- we collected data between April 2020 and March 2021. Like many people working on COVID, you know, we thought at first, maybe it's going to be a short you know episode and then we're, like, well this is seeming to go on so maybe we should start some supercomputing jobs to just collect data from our entire camera network and just see what happens. Maybe we'll get around to analyzing it at some point and that's what we did. So, this is using a large supercomputer at Argonne National Laboratory where I have a visiting appointment, and we have access to some incredible resources there, but our storage network is capable of storing petabytes of data. It's just been upgraded to 200 petabytes. That's more than most people can even imagine what to do with, but we were able to, you know, collect about 70 terabytes of data just for the cameras that we were looking at here. Okay, again, kind of difficult to put that on your laptop. So, then what we did is we basically had a job that ran every day, usually about five to six times a day, and then we take a snapshot of all these cameras simultaneously using a cluster computing job. It takes about 30 minutes to collect data simultaneously from all of these cameras- nearly simultaneously is a better way of putting it. But that's a pretty extraordinary thing in its own right. It's just having demonstrated how we can collect so much visual data on super computers. Then what we did is we actually looked at the numbers of humans and vehicles over time because we wanted to look at human mobility more generally so not just pedestrians, but also people getting into cars again, people not being in cars and other vehicles. And then we also looked at how this correlates with the Oxford Stringency Index, which is an index that measures what's going on with policy in various locales. Our- the punch line of our study is that we think visual data, especially in future pandemics, will be a method that will be used because it is already being shown to be effective for being able to understand the policies which I'll show you in some charts toward the end.

Slide 4

Okay, so this is just a little bit about the workflow we use for this. I already mentioned the camera discovery process, and then what we did is we applied a couple of filters just to make sure we're actually focusing on cameras that are actually going to really aid in the study because there are many cameras that we have access to that are kind of uninteresting. This is an example of an uninteresting camera. We had found one camera that actually is used to monitor like a road closure sign to make sure that the lights are still flashing. That's not going to tell us too much about human mobility. And then so we looked- we did some pre-processing just to find out which of these cameras are giving us any kind of data related to mobility, so this could be- we see humans there, we see vehicles there, etc. Alright, and then after finding the cameras that actually meet these requirements, then we actually go through and do the analysis here. Zoom is actually blocking my last part of the diagram but I'm- the key parts are to first of all find the relevant models. We have two different models that we're using. One of them is called Pedestron, which is a time-tested model for being able to look at pedestrian-type traffic and other human traffic, you know, maybe not always a pedestrian walking like on a crosswalk, it's like in various settings. And then we have like YOLOv3 which is a general object detector. We use that for- and we've trained that for actually being able to look at vehicles. And then there's some intelligence built into aggregating the data by location. We basically use a number of, you know, geolocation type of services to be able to do that, and then that's how we were able to actually come up with these like analyses by either country or by state. Okay, of course the end result is we want to make some pretty charts, and I hope they're pretty enough for this presentation, we're still finalizing that.

Slide 5

Okay anyway so the one thing we do is, you know, we have cameras in many locations. So, in many cases we're able to take the geo coordinates we have for these cameras and then we're able to corroborate that with what Google Street View is telling us and there is some human involvement in this process. So, we had to actually you know for our representative cameras that we looked at we did actually you know correlate them with Google Street View and as you can see here like one of these cameras is showing- yeah is showing ours okay and another is showing Google's. Okay, and we see some of the same sign- they're not the same words on the signs but we definitely see the same placement of signs there and we have pretty high confidence that that camera is where we think it is.

Slide 6

The other thing that- to do this kind of work requires that you set up a validation data set that is going to be used to help us do a little bit better analysis with both of those models I mentioned to you. So, what we did is we actually took some of our, you know, a subset of our images and just used them to provide a little bit better labeling of you know this is a pedestrian, this is a car, and so on. And one of the challenges that we face in our data set is that because these are public cameras they're like at all like random angles they're like placed everywhere like they're- some of them are placed far away, some are close up. And so, it actually requires that you look at different scenes and do some, you know, training there to make sure that what when you actually look at any image in any of our data set, you have a higher chance of actually being able to classify accurately.

Slide 7

One thing, I think- I don't know what happened here. I just sort of didn't mention one thing but these that I mentioned the selected scenes. But I just didn't get a chance to mention briefly that you can see here in the lower the set of images. These are the ones where we're applying our object detectors to them and we basically- these object detectors are typically going to give us- the green is the false- is the positives and then we have blue which is showing some that are false positives, and red are the negatives. Because you can see that like these object detectors do really well at classifying. This is just showing pedestrians here, but they do a really good job of identifying pedestrians, even like distant objects, you know, managed to be classified fairly accurately as pedestrians okay or people.

Slide 8

Alright, so anyway, so what are the key findings? This is what I want to spend a little bit of time talking about okay. I'm going to- I just dropped something on the floor. Okay, well, the key findings are that this does really well. So, I wanted to show you a little bit about like how these charts are organized here. So basically, we have an opening okay like this is when like there's an opening happening. This is when there's a lockdown so unfortunately there's some pixelation happening here in when I went from .pdf to .png but this is an opening in France so O of France, lockdown in France, and then another opening in France. And it's clear to see that, and especially in places where there were consistent policies about opening and closing, we definitely see you know the uptick in activity that happens both vehicular and pedestrian traffic and the cameras that we have in those locations, okay, really ticks up. Of course, as bad news starts to tick, you know, come in you know we start seeing people even pulling back in advance of the lockdown that takes place, for example, in France. Okay we all know the U.S. is kind of an interesting story and one of the things that's quite fascinating and all of our you know with- especially our charts for places like Georgia, we can see that like because the stringency index is, you know, is pretty flat here, we can see that the patterns of you know corresponding to- okay, I'm sorry. I've lost a little bit of my yeah- so you can see like there's a lockdown and an opening in Georgia and then that opening just sort of continues forever, and of course, the activity you know is actually- we do see like especially like vehicular activity is continuing to increase as the, probably, as the economy is opening up, but you know you see it otherwise kind of an inconclusive story there. But definitely when it comes to places with consistent policies, and of course, Europe has actually been pretty good in this regard, we see that the patterns of openings, lockdowns, and if there's like a second opening that happened, okay we did not have a reopening in time for Germany you know at the time of this study. Yeah, we get all of this from the what's published about the, you know, the opening- the lockdowns especially. But yeah, you see the European countries are all doing pretty well. We do see that Australia, you know, and then the U.S. states are a little bit less consistent. Hawaii we actually are still trying to understand more about what's going on there, but one thing we know is that in Hawaii there were a number of people traveling to Hawaii and they were basically following different rules that most of the people who were actually living in Hawaii. So, they're like you know- the pattern looks a lot like Georgia but maybe actually a little bit better than Georgia in terms of what's actually going on. So, the takeaway though is that you know at least when we do see you know consistent policy, we are able to see a pretty strong connection between you know when analyzing the visual data of what's actually going on both with pedestrian and vehicular traffic. Certainly, the hope is that in the future there will be a little more consistency that you know when it comes to COVID-19 policy and- this is one of the reasons why we undertook our study in the first place is we did want to have a sense of how people are actually responding to whatever policies are present. And of course, I know I'm going to get one question about whether we've analyzed other states. Yes, we have, but we also wanted to make sure that we had enough data, especially visual data, for the locations we were analyzing because some states, for example, in the US we do not have as many cameras as others.

Slide 9

So anyway, I just want to say a little bit about that we have a very large research team. This is a team that comprises probably- we've had a variable size. We have about 15 undergraduates, we've had, you know, three graduate students involved, and of course we have four faculty involved you know myself and Yung-Hsiang Lu were, you know, the faculty leads for this work. And we have David Shoham, who has also collaborated with us from the Department of Public Health. He was at Loyola and has moved to East Tennessee State University, and Wei Zakharov from Information Science. So, it just wouldn't be possible to even do a study like this without these highly-talented students and faculty so I just want to make sure I acknowledge them.