

[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: Julie Meunier

---

Transcript

Lauren Close

Tout d'abord, je vais passer la parole à George de l'Université Loyola de Chicago, qui va nous faire cette première présentation de l'après-midi.

Florence Hudson

Génial.

Lauren Close

Je vais m'arrêter - George et je vais te laisser partager ton écran.

George K. Thiruvathukal

D'accord, merci beaucoup.

Florence Hudson

Et pendant qu'il le fait, nous aimerions simplement encourager les gens à poser des questions aux intervenants. N'hésitez pas à les mettre dans le chat. Ils peuvent y répondre de manière asynchrone, ou nous aurons une session de questions-réponses ouverte à la fin. Allez-y, George.

George K. Thiruvathukal

D'accord, si je peux simplement obtenir confirmation que vous pouvez voir uniquement mes diapositives.

Florence Hudson

Ca semble bien.

George K. Thiruvathukal

### *Slide 1*

D'accord, merci beaucoup. Tout d'abord, quand on vous présente comme un orateur fantastique, c'est un peu difficile à suivre, mais merci pour cette introduction très aimable à nous tous, Lauren. Et je vais plonger directement dans cette présentation dans l'esprit des exposés éclair. Ainsi, je vais parler de "l'observation de la mobilité humaine pendant la COVID-19".

### *Slide 2*

Notre travail consiste essentiellement à tirer parti de ce que nous appelons les caméras publiques, les caméras de réseau public, et à analyser cette grande quantité de données visuelles capturées par ces caméras pour observer comment les gens respectent la distanciation sociale lors des différentes étapes des confinements dans divers endroits. Nous avons en fait étudié dans cette étude pilote seulement cinq pays et trois États pour analyser l'efficacité de ces politiques de confinement afin de voir si cela peut être un outil dans la boîte à outils pour aider à réduire la propagation de la COVID-19. Le principal défi de l'étude est que nous avons des millions d'images, en fait, aujourd'hui, nous avons probablement près de 7 millions d'images. Et ce que nous voulons faire, c'est pouvoir prendre ces images, regarder où elles sont situées et essentiellement examiner les politiques qui sont dans ces endroits et voir s'il y a des motifs dans les données. C'est donc vraiment un projet de big data, et je tiens à dire que je dis habituellement aux gens : n'essayez pas cela sur votre ordinateur portable à la maison. Cela ne fonctionnera probablement pas, et vous risquez de surchauffer votre ordinateur en quelques minutes.

### *Slide 3*

Donc, sans plus tarder. Cette diapositive montre simplement quelques exemples des types d'images que nous avons à notre disposition. Elles proviennent toutes de caméras publiques disponibles sur Internet, généralement, comme des gouvernements et des particuliers ou d'autres organisations comme des parcs nationaux qui ont ces caméras et les intègrent sur leurs pages web. Ils veulent les rendre disponibles à d'autres, pas nécessairement à des fins de recherche, mais ce sont des caméras publiques. Contrairement au réseau massif de caméras en réseau fermé qui existent dans de nombreuses municipalités, par exemple, ce sont des sources qui peuvent être légitimement utilisées à des fins d'étude sans soulever de préoccupations significatives en matière de confidentialité et de sécurité. Donc, dans notre projet de recherche global, le projet Cam Squared, que j'ai mentionné dans le titre mais dont je n'ai pas beaucoup parlé dans le souci de respecter le temps imparti, nous avons en fait un processus pour découvrir ces caméras. Nous avons des articles sur ce sujet. Et l'étude que nous avons réalisée consiste essentiellement à collecter des données en continu auprès de toutes les caméras de notre réseau, soit environ 36 000 caméras que nous avons découvertes grâce à ce processus automatique que nous avons créé. Nous avons collecté des données entre avril 2020 et mars 2021. Comme beaucoup de personnes travaillant sur la COVID, nous avons pensé au début que cela allait être une courte période, puis nous nous sommes dit que cela semblait durer, alors peut-être devrions-nous commencer quelques travaux de supercalcul pour simplement collecter des données de l'ensemble de notre réseau de caméras et voir ce qui se passe. Peut-être que nous arriverons à les analyser à un moment donné, et c'est ce que nous avons fait. Nous utilisons un supercalculateur à l'Argonne National Laboratory, où j'ai une affectation de visiteur, et nous avons accès à des ressources incroyables là-bas. Mais notre réseau de stockage est capable de stocker des pétaoctets de données. Il vient d'être mis à niveau à 200 pétaoctets. C'est plus que ce que la plupart des gens peuvent imaginer, mais nous avons pu collecter environ 70 téraoctets de données uniquement pour les caméras que nous examinons ici. Difficile à mettre sur votre

ordinateur portable. Ensuite, ce que nous avons fait, c'est que nous avons essentiellement un travail qui s'exécutait tous les jours, généralement cinq à six fois par jour, et nous prenions un instantané de toutes ces caméras simultanément à l'aide d'un travail de calcul en grappe. Il faut environ 30 minutes pour collecter des données simultanément de toutes ces caméras - presque simultanément est une meilleure façon de le dire. Mais c'est une chose assez extraordinaire en soi. Cela a simplement démontré comment nous pouvons collecter autant de données visuelles sur des supercalculateurs. Ensuite, nous avons regardé les nombres d'humains et de véhicules au fil du temps, car nous voulions examiner la mobilité humaine de manière plus générale, pas seulement les piétons, mais aussi les gens montant dans des voitures, les gens qui ne sont pas dans des voitures et d'autres véhicules. Nous avons également examiné comment cela corrélait avec l'indice de rigueur d'Oxford, qui mesure ce qui se passe en matière de politiques dans divers endroits. La conclusion de notre étude est que nous pensons que les données visuelles, surtout dans les futures pandémies, seront une méthode qui sera utilisée car elle montre déjà son efficacité pour comprendre les politiques, comme je vous le montrerai dans quelques graphiques vers la fin.

#### *Slide 4*

D'accord, c'est juste un peu sur le flux de travail que nous utilisons pour cela. J'ai déjà mentionné le processus de découverte de la caméra, puis ce que nous avons fait, c'est que nous avons appliqué quelques filtres juste pour nous assurer que nous nous concentrons réellement sur les caméras qui vont vraiment aider à l'étude, car il y a beaucoup de caméras auxquelles nous avons accès qui sont un peu inintéressantes. Voici un exemple de caméra inintéressante. Nous avons trouvé une caméra qui est utilisée pour surveiller comme un panneau de fermeture de route pour s'assurer que les lumières continuent de clignoter. Cela ne nous en dira pas beaucoup sur la mobilité humaine. Et donc, nous avons fait quelques pré-traitements simplement pour savoir quelles caméras nous fournissent des données liées à la mobilité, que ce soit des humains, des véhicules, etc. Ensuite, après avoir trouvé les caméras qui répondent réellement à ces critères, nous passons effectivement à l'analyse ici. Zoom bloque en fait ma dernière partie du diagramme, mais je vais - les parties clés sont d'abord de trouver les modèles pertinents. Nous avons deux modèles différents que nous utilisons. L'un d'eux s'appelle Pedestron, qui est un modèle éprouvé dans le temps pour pouvoir examiner le trafic de type piéton et d'autres types de trafic humain, pas toujours un piéton marchant comme sur un passage pour piétons, mais dans divers contextes. Et puis nous avons YOLOv3, qui est un détecteur d'objets général. Nous l'utilisons pour - et nous l'avons formé pour être capable de regarder les véhicules. Et puis il y a une certaine intelligence intégrée pour agréger les données par emplacement. Nous utilisons essentiellement un certain nombre de services de géolocalisation pour le faire, et c'est ainsi que nous avons pu en fait élaborer ces analyses par pays ou par État. Bien sûr, le résultat final est que nous voulons créer des graphiques assez sympas, et j'espère qu'ils seront assez sympas pour cette présentation, nous finalisons encore cela.

#### *Slide 5*

D'accord, de toute façon, une des choses que nous faisons est que nous avons des caméras dans de nombreux endroits. Dans de nombreux cas, nous pouvons prendre les coordonnées géographiques que nous avons pour ces caméras, puis les corroborer avec ce que Google Street View nous dit, et il y a une certaine implication humaine dans ce processus. Nous avons donc dû réellement, vous savez, pour nos caméras représentatives que nous avons examinées, nous devons réellement les corrélérer avec Google Street View, et comme vous pouvez le voir ici, une de ces caméras montre - ouais, montre la nôtre,

d'accord, et une autre montre celle de Google. D'accord, et nous voyons certains des mêmes signes - ce ne sont pas les mêmes mots sur les panneaux, mais nous voyons définitivement le même emplacement des panneaux là-bas, et nous avons une assez grande confiance que cette caméra est là où nous pensons qu'elle est.

#### *Slide 6*

L'autre chose à faire pour ce genre de travail est de configurer un ensemble de données de validation qui va être utilisé pour nous aider à faire une analyse un peu meilleure avec les deux modèles que je vous ai mentionnés. Donc, ce que nous avons fait, c'est que nous avons pris en fait une partie de nos images, une partie de nos images, et les avons simplement utilisées pour fournir une étiquetage un peu meilleur, vous savez, c'est un piéton, c'est une voiture, et ainsi de suite. Et l'un des défis auxquels nous sommes confrontés dans notre ensemble de données est que, comme ce sont des caméras publiques, elles sont à tous les angles aléatoires, elles sont placées partout, certaines sont loin, certaines sont proches. Et donc, cela nécessite en fait que vous regardiez différents scénarios et fassiez un peu de formation pour vous assurer que lorsque vous regardez réellement n'importe quelle image dans notre ensemble de données, vous avez une plus grande chance de pouvoir la classer de manière précise.

#### *Slide 7*

Une chose, je pense - je ne sais pas ce qui s'est passé ici. Je n'ai tout simplement pas mentionné une chose, mais je n'ai tout simplement pas eu l'occasion de mentionner brièvement que vous pouvez voir ici dans la partie inférieure de l'ensemble d'images. Ce sont celles où nous appliquons nos détecteurs d'objets, et nous les avons essentiellement - ces détecteurs d'objets vont généralement nous donner - le vert est le faux positif, puis nous avons le bleu qui montre certains faux positifs, et le rouge sont les négatifs. Parce que vous pouvez voir que ces détecteurs d'objets font vraiment bien pour classifier. Cela montre simplement des piétons ici, mais ils font un très bon travail pour identifier les piétons, même des objets éloignés, vous savez, parviennent à être classifiés assez précisément comme des piétons, d'accord, ou des personnes.

#### *Slide 8*

D'accord, quelles sont les principales conclusions ? C'est de cela dont je veux parler un peu, d'accord. Je vais- j'ai juste laissé tomber quelque chose par terre. D'accord, bon, les principales conclusions sont que cela fonctionne vraiment bien. Donc, je voulais vous montrer un peu comment ces graphiques sont organisés ici. Fondamentalement, nous avons une ouverture, d'accord, c'est quand il y a une ouverture qui se produit. C'est quand il y a un confinement, donc malheureusement, il y a une certaine pixellisation ici quand je suis passé de .pdf à .png, mais c'est une ouverture en France, donc O de la France, confinement en France, et puis une autre ouverture en France. Et on voit clairement que, surtout dans les endroits où il y avait des politiques cohérentes d'ouverture et de fermeture, on voit vraiment une augmentation de l'activité à la fois du trafic routier et piétonnier, et les caméras que nous avons dans ces endroits, d'accord, augmentent vraiment. Bien sûr, lorsque de mauvaises nouvelles commencent à arriver, vous savez, nous commençons à voir les gens se retirer même avant le confinement qui a lieu, par exemple, en France. D'accord, nous savons tous que les États-Unis sont un peu une histoire intéressante et l'une des choses qui est assez fascinante, surtout avec nos graphiques pour des endroits comme la Géorgie, c'est que parce que l'indice de sévérité est, vous savez, assez plat ici, on voit que les motifs correspondent à- d'accord, je suis désolé. J'ai perdu un peu le fil, ouais- donc on voit qu'il y a un

confinement et une ouverture en Géorgie, et puis cette ouverture continue indéfiniment, et bien sûr, l'activité, vous savez, surtout l'activité des véhicules, continue d'augmenter probablement à mesure que l'économie s'ouvre, mais vous savez, on voit sinon une histoire un peu inconclusive là-bas. Mais certainement, en ce qui concerne les endroits avec des politiques cohérentes, et bien sûr, l'Europe a été assez bonne à cet égard, on voit que les motifs d'ouvertures, de confinements, et s'il y a une deuxième ouverture qui se produit, d'accord, nous n'avons pas eu de réouverture à temps pour l'Allemagne, vous savez, au moment de cette étude. Ouais, on obtient tout cela à partir de ce qui est publié sur les ouvertures- les confinements en particulier. Mais ouais, on voit que les pays européens s'en sortent assez bien. On voit que l'Australie, vous savez, et ensuite les États-Unis sont un peu moins cohérents. Hawaï, nous essayons toujours de comprendre davantage ce qui se passe là-bas, mais une chose que nous savons, c'est qu'à Hawaï, il y avait un certain nombre de personnes voyageant à Hawaï et elles suivaient essentiellement des règles différentes de la plupart des personnes qui vivaient réellement à Hawaï. Donc, vous savez, le schéma ressemble beaucoup à la Géorgie, mais peut-être en fait un peu mieux que la Géorgie en termes de ce qui se passe réellement. Donc, la conclusion, c'est que, vous savez, au moins lorsque nous voyons, vous savez, une politique cohérente, nous sommes en mesure de voir une connexion assez forte entre, vous savez, l'analyse des données visuelles de ce qui se passe réellement, à la fois avec le trafic piétonnier et routier. Bien sûr, l'espoir est qu'à l'avenir, il y aura un peu plus de cohérence, vous savez, en ce qui concerne la politique COVID-19 et- c'est l'une des raisons pour lesquelles nous avons entrepris notre étude en premier lieu, c'est que nous voulions avoir une idée de comment les gens réagissent réellement à quelles que soient les politiques présentes. Et bien sûr, je sais que je vais avoir une question sur le fait de savoir si nous avons analysé d'autres États. Oui, nous l'avons fait, mais nous voulions aussi nous assurer que nous avions suffisamment de données, en particulier des données visuelles, pour les endroits que nous analysions parce que certains États, par exemple aux États-Unis, nous n'avons pas autant de caméras que d'autres.

### *Slide 9*

Je voudrais simplement dire quelques mots à ce sujet : nous avons une équipe de recherche très importante. C'est une équipe qui comprend probablement - nous avons eu une taille variable. Nous avons environ 15 étudiants de premier cycle, nous avons eu, vous savez, trois étudiants diplômés impliqués, et bien sûr, nous avons quatre membres du corps professoral impliqués, vous savez, moi-même et Yung-Hsiang Lu étions les chefs de file du corps enseignant pour ce travail. Et nous avons David Shoham, qui a également collaboré avec nous du département de santé publique. Il était à Loyola et a déménagé à l'Université de l'État de l'Est du Tennessee, et Wei Zakharov de la Science de l'Information. Donc, il ne serait même pas possible de réaliser une étude comme celle-ci sans ces étudiants et membres du corps professoral hautement talentueux, donc je tiens à m'assurer de les mentionner.