Facebook développe la reconnaissance faciale… de dos ! | Le Net Expert Informatique

Capture d’écran du film « Mon nom est personne » (1973) (Tonino Valerii)	**Facebook développe la reconnaissance faciale… de dos !**

Après la reconnaissance vocale, la biométrie et son cortège (empreinte digitale, ADN, iris…), nous pourrons être reconnus même visage caché. Une technique qui devrait surtout servir à (encore plus de) la surveillance.

Au Far-West, dans les westerns il y avait une règle d'or, un code de bonne conduite, un code d'honneur : on ne tire pas sur les gens de dos. Par exemple dans le chef d'œuvre « Mon nom est personne », Jack Beauregard (aka Henry Fonda) déclare :
« D'accord, Nevada était mon frère, mais c'était aussi un salaud de la plus belle espèce. Pour une poignée de dollars, il tirait dans le dos d'un ami, et je ne vais pas risquer ma peau pour le venger. »
Comme on n'arrête pas le progrès, après avoir développé ceux déjà redoutablement efficaces de la reconnaissance faciale, voici que Facebook annonce avoir mis au point un nouvel algorithme capable de vous reconnaître sur une photo ou dans une vidéo même si vous êtes de profil ou … de dos. Oui oui.

La reconnaissance faciale de dos
L'article dans lequel les chercheurs du labo IA (intelligence artificielle) de Facebook expliquent comment ils cherchent à aller au-delà de la seule reconnaissance faciale au travers de la détection d'une multitude d'autres « indices » est disponible en ligne [PDF]. La méthode utilisée par l'algorithme a même un petit nom sympatoche : elle s'appelle PIPER pour « Pose Invariant PErson Recognition ». En gros : « Les poses identiques pour la reconnaissance de personnes. »
Très précisément, il ne s'agit pas de « reconnaître quelqu'un de dos » (mais avouez que ça fait un bon titre), ni même de reconnaître quelqu'un uniquement sur la base de ses mouvements ou attitudes, mais – c'est déjà pas mal – de se servir de la reconnaissance desdites attitudes (baptisées « poselets ») pour améliorer la reconnaissance faciale.
« Nous utilisons la méthode "PIPER", qui agrège les indices recueillis par un système de reconnaissance d'attitudes, entraîné par des réseaux à convolution pour lisser les variations de la pose, combiné avec une reconnaissance de visage et une reconnaissance globale. » [PDF]

Image extraite de l'étude de Facebook (DR)

Tels que présentés dans l'étude, les résultats sont assez bluffants :
« 83,05% de réussite pour les 581 identités du corpus (1 identité étant décrite par plusieurs photos) dans lesquelles on ne dispose pas de photo "de face". De plus, quand une photo de face est disponible, le taux de réussite de DeepFace (l'algo de reconnaissance faciale) passe de 89,3% à 92,4% faisant baisser de 40% le taux d'erreur relatif. » [PDF]
A noter enfin que les datasets utilisés sont directement piochés dans la base Creative Commons de Flickr [PDF].

Image extraite de l'article en question (DR)

Le poids des mots, le choc des photos
Nous sommes en train de vivre, à l'échelle de l'image et de la vidéo, la même révolution scientifique et technologique que celle que nous vécûmes au début des années 80 à l'échelle de l'ingénierie linguistique et de la fouille de corpus de textes avec l'invention du TAL (traitement automatique des langues http://fr.wikipedia.org/wiki/Traitement_automatique_du_langage_naturel) dont Jean Véronis fut un des pionniers.

Le côté positif c'est que grâce à ces progrès nous avons aujourd'hui Google, des livres numériques, nous pouvons travailler à l'échelle de corpus considérables, chercher des mots-clés facilement dans à peu près n'importe quel texte, bénéficier de services de traduction automatique de plus en plus efficaces, de correcteurs d'orthographe et de syntaxe, etc. Le côté obscur c'est que l'ensemble des technologies de surveillance dont on débat aujourd'hui au travers des différents scandales d'écoute ou de flicage des populations reposent sur les mêmes progrès de l'ingénierie linguistique.

La même chose se produira, est déjà en train de se produire, à l'échelle des technologies de ce que nous pourrions baptiser le TAIFA (traitement automatique des images fixes et animées). Et de la même manière que nous en retirerons de grands services, elles nous exposeront simultanément à de grandes dérives liées à la surveillance et au contrôle.

Il faut se souvenir qu'il y a à peine 15 ans, au début des années 2000, la plupart des spécialistes de ces questions butaient sur d'immenses difficultés pour simplement parvenir à faire avec les images ce que l'on avait réussi à faire à peu près correctement avec les textes, c'est-à-dire parvenir à les indexer.

Quinze ans plus tard, non seulement l'indexation des images et des vidéos se fait à l'identique ou sans poser guère plus de problèmes que celle du texte, mais l'on est également capable, comme pour le texte, de descendre à des niveaux de granularité très fins dans cette indexation, grâce donc notamment aux technologies de reconnaissance faciale.

Autre exemple, à l'aide à la fois de technologies relevant du « Deep Machine Learning » croisées avec les métadonnées associées à notre navigation qui permet de déterminer avec un taux de précision assez étonnant (même si certains résultats sont encore très … aléatoires), l'âge d'une personne en se basant simplement sur une de ses photos comme en témoigne le projet « How Old » de Microsoft (http://how-old.net).

Un exemple d'utilisation de « How old » sur Henry Fonda, qui avait 68 ans sur cette photo (mais était maquillé pour avoir l'air plus vieux)
Textes, images, sons, vidéos sont donc désormais indexés et des programmes sont capables d'y retrouver aussi bien des mots-clés que d'y reconnaître des visages et d'en déterminer l'âge.

Who's next ? Indexer les attitudes
Après les textes, les images, les sons, les vidéos, que reste-t-il encore à « reconnaître » ou à « retrouver » ? Précisément, les « attitudes », nos attitudes. Étendre la fouille textuelle, le « search and retrieve », jusqu'à parvenir à des niveaux très fins de compréhension, niveaux permettant à leur tour la production de nouveaux textes. Étendre la reconnaissance faciale, le « look and find », jusqu'à des niveaux très fins d'identification, niveaux permettant à leur tour une automatisation et une systématisation de logiques qu'il faut bien désigner comme s'apparentant pour l'essentiel à des logiques de… surveillance.

A la recherche de toujours plus de singularité et d'essentialisation comme s'efforcent de le faire, en parallèle, les techniques de biométrie en permettant de nous loguer ou de débloquer notre smartphone avec notre empreinte digitale, demain peut-être avec notre iris ou pourquoi pas avec l'analyse d'un échantillon de notre ADN (OK, faudrait alors lécher l'écran ou cracher dessus mais avouez que ce serait rigolo, et c'est… inéluctable).

Car par-delà la tentative d'isoler chacune de nos attitudes pour mieux nous « reconnaître », d'autres techniques utilisent la même approche aux fins cette fois de caractériser ce qui correspond à une attitude gestuelle dans notre comportement en ligne, c'est-à-dire l'activité de navigation. L'idée est d'utiliser – en gros – notre historique de navigation pour remplacer les innombrables mots de passe qui nous sont demandés par différents services : au lieu de taper « azerty » pour accéder à votre e-mail et « administrateur » pour ouvrir une session parentale sur votre ordinateur, on vous demanderait « quel est le film que vous avez visionné hier soir ? », « à qui avez-vous envoyé votre dernier SMS ? », « quel album avez-vous ajouté ce matin à votre playlist Deezer ? », etc. Bref on ne vous reconnaîtrait plus par votre mot (de passe) mais par votre comportement (de navigation).

Et là encore, le parallèle avec l'ingénierie linguistique est frappant. Dans les premières années de son développement – celui de l'ingénierie linguistique – on se contentait et s'émerveillait d'être capables d'aller simplement « retrouver » un mot dans un texte. Puis on commença à s'intéresser à la possibilité de retrouver ce mot mais également les mots de la même famille, puis ses synonymes, puis la totalité du champ lexical, y compris métaphorique, rattaché à ce mot ou à ce contexte, et ainsi de suite jusqu'aux derniers progrès dans le domaine de la reconnaissance et de l'extraction des entités nommées et de tout ce qu'elles permettent de faire.

Faux positif attitude
Même chose donc, même progrès dans le domaine des images et de la vidéo : après avoir détecté et reconnu des visages, on s'efforce de détecter et de reconnaître des attitudes. C'est fascinant, c'est vertigineux et c'est bien sûr dangereux.

C'est dangereux car l'une des différences de taille entre les technologies du TAL et celles du TAIFA c'est que pour les premières on disposait – et on dispose encore – d'un volet d'applications très large, même s'il incluait également des pratiques plus que contestables de surveillance, alors que pour les secondes, l'essentiel des applications qui en résulteront seront d'abord orientées vers des pratiques très discutables de surveillance.

Pour le dire différemment, être capable de reconnaître des mots peut permettre de faire plein de choses, et accessoirement de mieux écouter des conversations ; être capable de détecter des visages et des attitudes, à part vous aider à trier automatiquement vos milliers de photos – ce qui n'est déjà pas si mal – on ne voit pas très bien à quoi ça peut servir d'autre, à part en effet à développer des technologies de surveillance et de contrôle plus efficaces, plus intrusives, plus omniprésentes.

Un danger que renforce en même temps qu'il le souligne et le met en évidence le nouveau fétichisme du fichier et son cortège d'algorithmies permettant de « détecter » tout type de comportement.

Nombre de nos comportements, de nos attitudes en ligne relèvent de pratiques jaculatoires. De « jaculations » au sens premier « d'élan d'enthousiasme » ou d'éjaculations aux sens figurés non pas de « prières » mais de « statuts courts, émis à intervalles réguliers, avec force et un débit rapide », ou bien alors de « propos courts généralement insultants ou vulgaires » (cf., entre autres, le compte Twitter de Nadine Morano), ou bien enfin de « production ou manifestation spontanée et qui a généralement une certaine force, ou qui se manifeste violemment » (un des ressorts du principe de viralité sur les réseaux). Les algorithmes détectaient jusqu'à présent sans peine la moindre de ces jaculations. Ils viennent d'étendre cette détection aux jaculations … faciales.

Méritons-nous une bonne correction ?
Plus sérieusement, à la personnalité des algorithmes ou aux logiques de personnification succédant à celles de personnalisation, aux détections algorithmiques de nos moindres comportements, s'ajoutent désormais de nouvelles « couches » (la reconnaissance faciale, celle de nos attitudes) qui favorisent le déploiement de « technologies de l'empathie » et de leur angoissant cortège de correcteurs comportementaux là où nous ne connaissions jusqu'ici que la tyrannie et les affres des correcteurs orthographiques. Lesquels correcteurs orthographiques ont commencé par régler effectivement quelques problèmes avant de nous suggérer des recherches ou des réponses avant que nous ne leur ayons soumis la moindre question, la « libération » promise sur la gestion orthographique se transformant très vite en aliénation subie de nos processus de « requêtage » et de navigation.

Une évolution qui sera vraisemblablement la même mais cette fois à l'échelle encore plus problématique d'une « correction de nos comportements ». Des algorithmes reconnaissant l'ensemble des nos attitudes et de nos postures, des algorithmes dotés de « personnalités » s'adaptant à ce qu'ils supposent ou infèrent être la nôtre – de personnalité –, des algorithmes, enfin, corrigeant nos comportements pour les rendre plus… pour les rendre plus… pour les rendre plus… Et oui. Tout le problème est là. Dans ces quelques points de suspension et dans les logiques commerciales, politiques ou idéologiques qui les façonneront. Sans que nous n'ayons plus les moyens ni le temps d'y déceler les moyens, l'ampleur ou même les vrais acteurs à l'origine de ce phénomène de manipulation parfaitement inédit à cette échelle.

Manipulations singulières
Les médias traditionnels du XXe siècle avaient érigé en champ d'étude la manipulation des masses et les différentes techniques de propagande. Les algorithmes du XXIe siècle nous promettent une nouvelle forme de manipulation qui reste encore largement à étudier et à décrire, une manipulation ne reposant plus sur « les masses » mais sur des agrégats volatiles faits de requêtes, de comportements, de visages et d'attitudes permettant d'essentialiser chaque individu sorti de la masse. Ils conjuguent le verbe manipuler au singulier. Du latin « manipulare : conduire par la main ». De l'index blanc qui navigue au pouce bleu qui Like, nous sommes entre leurs mains. Nous nous sommes pris les doigts dans le digital.

Les algos nous racontent la fable du cyclope rendu aveugle. Mais ils ont choisi le rôle d'Ulysse pour nous laisser celui de Polyphème.

Le Web n'est pas le Far-West, mais aucune règle n'empêche de reconnaître un homme dans son dos. Le fait qu'il soit aujourd'hui impossible – ou réservé à quelques geeks – d'y être « personne » est à la fois la cause et la conséquence du problème. N'est-ce pas, Jack ?

Bientôt, Facebook pourra vous reconnaître même si votre visage est dissimulé. Une équipe du laboratoire Facebook IA Research (assistée de chercheurs de l'Université de Berkeley) vient de publier une étude sur un nouvel algorithme (ô joie) qui permet d'identifier des personnes à partir de leur posture corporelle.
Sur son blog, Olivier Ertzscheid, maître de conférences en sciences de l'information et de la communication, s'inquiète des applications pratiques de cette technologie. Nous reproduisons ce texte avec l'aimable autorisation de son auteur.
Le titre a été modifié. Une citation en anglais tirée de l'étude a été traduite et certains intertitres raccourcis. Rémi Noyon

---

Source : http://rue89.nouvelobs.com/2015/05/12/facebook-developpe-reconnaissance-faciale-dos-259134
Par Olivier Ertzscheid Enseignant chercheur