

Mémoire au comité ETHI : L'utilisation et l'impact de la technologie de reconnaissance faciale

Prof. Rob Jenkins, Département de psychologie, Université de York, Royaume-Uni

Le 4 avril 2022

1. Résumé des données probantes

Plus de cinq décennies de recherche ont donné lieu à une importante littérature scientifique sur la reconnaissance faciale par les humains et la technologie. Ces recherches ont permis d'établir plusieurs faits essentiels qui sont cruciaux pour cette enquête.

1.1. Reconnaissance faciale par les humains

1.1.1. La reconnaissance des visages familiers et non familiers est fondamentalement différente. La reconnaissance des visages *familiers* (visages que nous connaissons) est très précise [1]. En revanche, la reconnaissance des visages *non familiers* (visages que nous ne connaissons pas) est étonnamment imprécise [2]. La quasi-totalité des applications de la reconnaissance faciale concerne les visages non familiers. La technologie de reconnaissance faciale (TRF) imite les principes de la reconnaissance des visages non familiers par les humains.

1.1.2. L'intuition est un mauvais prédicteur de la précision de la reconnaissance faciale. Nous sommes de mauvais juges en ce qui concerne les visages non familiers qui sont faciles ou difficiles à identifier [3]. La plupart des gens ne sont pas doués pour cette tâche et croient pourtant qu'elle est plus simple qu'elle ne l'est en réalité [4].

1.1.3. La prise de décision humaine est sujette à des biais dans lesquels les erreurs sont influencées par l'information contextuelle [5]. Par exemple, des biais peuvent être induits par l'information de « note de correspondance » montrée aux opérateurs par la TRF [6], qui peuvent potentiellement conduire à des erreurs d'identification de confiance.

1.1.4. Il existe de grandes différences individuelles dans la capacité de reconnaissance faciale. Il s'agit d'un trait cognitif stable qui est largement déterminé par des facteurs génétiques [7].

1.1.5. Les visages de sa propre race sont mieux reconnus que les visages d'une autre race. C'est le constat fait dans le monde entier pour toutes les combinaisons de races connues. L'effet des facteurs démographiques sur la précision est faible. Cependant, les gens s'attendent à ce qu'il soit important [8]. Les disparités entre groupes démographiques dans la précision de la reconnaissance faciale reflètent le régime social des visages de l'observateur. Cela est important pour au moins deux raisons. Premièrement, nous devrions nous attendre à des disparités de rendement même en l'absence de préjugés de groupe. Deuxièmement, le « régime social des visages » a un équivalent clair dans la TRF – plus précisément, la composition des ensembles d'images qui sont utilisés dans la formation.

1.2. Technologie de reconnaissance faciale (TRF)

1.2.1. Les tests d'évaluation comparative de la précision des algorithmes sont généralement effectués dans des conditions idéales. La qualité de l'image dans ces tests est exagérément bonne en raison des conditions contrôlées de capture d'image, qui ne sont pas possibles dans les enquêtes criminelles et la surveillance.

1.2.2. Nous ne connaissons pas la précision de la TRF dans des contextes appliqués réalistes. Il y a peu de preuves de la précision de la TRF dans les tests qui reflètent la gamme des déploiements opérationnels du monde réel. Une évaluation correcte de bout en bout doit comprendre la mesure de la précision du système global, qui comprend à la fois la TRF et les personnes qui l'utilisent.

1.2.3. Nous ne savons pas comment la police utilise actuellement la TRF. Il existe un manque d'information et de compréhension quant à la manière dont la TRF est utilisée dans les enquêtes policières. Le manque de clarté peut masquer le potentiel d'amélioration et empêcher l'examen public des coûts et des avantages de la TRF. Aux États-Unis, il est apparu récemment que les forces de police ne tiennent pas de registres et ne vérifient pas l'utilisation de la TRF, malgré son adoption généralisée [9, 10].

1.2.4. Même dans des conditions de test idéales, les erreurs sont plus

probables pour certains groupes démographiques que pour d'autres [11].

Les disparités du taux d'erreur entre les groupes démographiques reflètent la composition des ensembles d'images utilisés pour former la TRF.

1.3. Les facteurs humains dans la TRF

1.3.1. La TRF nécessite une surveillance humaine sous forme d'intervention, d'interprétation et de contrôle. En effet, dans les applications juridiques et médico-légales, la TRF ne « reconnaît » pas les visages. Au lieu de cela, la recherche automatisée dans les bases de données fournit une « liste de candidats » de correspondances potentielles. Les décisions finales concernant l'identité des visages sont prises par des opérateurs humains qui sélectionnent les visages dans la liste des candidats et les comparent à la cible de la recherche [12].

1.3.2. Les opérateurs humains commettent fréquemment des erreurs lors de l'examen des listes de candidats. Pour des listes de huit visages seulement, les opérateurs se sont trompés environ 50 % du temps [13]. Ce niveau de rendement est conforme aux recherches sur l'identification par témoin oculaire, dont on sait qu'elle est peu fiable, les témoins bien intentionnés identifiant souvent par erreur des suspects innocents [14].

1.3.3. La surveillance humaine est intégrée dans la réglementation. Le Code de pratique des caméras de surveillance du Royaume-Uni [15] stipule que l'utilisation de la TRF « doit toujours comprendre une intervention humaine avant que ne soient prises des décisions qui affectent un individu ». Un principe similaire de surveillance humaine a été adopté publiquement par le gouvernement fédéral australien, selon lequel l'identification d'une personne ne sera jamais réalisée par la seule technologie.

1.4. Approches visant à optimiser le rendement humain

1.4.1. Sélection de personnes ayant une haute aptitude. La probabilité d'erreur peut être considérablement réduite (i) en sélectionnant des personnes ayant une grande capacité de reconnaissance faciale pour devenir des opérateurs de TRF, et (ii) en leur fournissant une formation appropriée [13]. La capacité de reconnaissance faciale est indépendante du QI et de la capacité visuelle générale [17], mais peut être mesurée de manière fiable à l'aide de tests qui ciblent spécifiquement la reconnaissance faciale [18].

1.4.2. Analyse de la sagesse des foules. Des recherches récentes montrent que l'agrégation des jugements d'identité des visages effectués indépendamment par

plusieurs humains ou algorithmes améliore considérablement la précision de la reconnaissance faciale. Conjuguée aux récentes avancées dans la compréhension des différences individuelles dans la capacité de reconnaissance faciale, cette approche de la « sagesse des foules » est une voie prometteuse pour réduire au minimum les erreurs associées à l'utilisation de la TRF [19].

1.4.3. Une formation efficace. Des tests en laboratoire indiquent que certaines méthodes de formation apportent des avantages à l'identification des visages, mais ceux-ci sont faibles par rapport à la sélection et au recrutement [20]. Les cours de formation commerciaux sont souvent inefficaces malgré les commentaires positifs des stagiaires [21], et il est donc important que l'efficacité des cours de formation sur la TRF soit évaluée officiellement.

2. Principes d'utilisation de la TRF dans un cadre médico-légal

2.1. Une surveillance humaine appropriée et une attention aux opérateurs humains dans la conception et la mise en œuvre des systèmes de reconnaissance faciale. Les humains sont un élément essentiel pour garantir la précision de la TRF. Des efforts délibérés doivent être faits pour s'assurer que les humains participant aux décisions de reconnaissance faciale sont hautement qualifiés, soit par un recrutement ciblé, soit par une formation fondée sur des données probantes. Des tests scientifiquement validés de la capacité d'identification des visages peuvent être utilisés pour évaluer l'aptitude des opérateurs de TRF et surveiller leur rendement dans le temps. En outre, la conception des systèmes de reconnaissance faciale doit intégrer des contrôles appropriés afin de réduire au minimum le risque d'erreurs importantes.

2.2. Transparence. L'utilisation de la TRF dans le système juridique devrait s'accompagner d'une divulgation transparente de l'information relative à la précision, aux points forts, aux limites et au fonctionnement de cette technologie. Il est également nécessaire de comprendre la façon dont la technologie est utilisée et de concevoir une réglementation qui garantit une utilisation appropriée.

2.3. Perfectionnement d'une main-d'œuvre experte en reconnaissance faciale. Si la TRF doit être adoptée dans la pratique médico-légale, de nouveaux types de chercheurs et de praticiens spécialisés sont nécessaires pour concevoir, évaluer, superviser et expliquer les systèmes d'identification des visages qui en résultent.

Étant donné que ces systèmes intègrent des décisions prises par les humains et l'IA, il faut des personnes possédant une vaste expertise dans des disciplines connexes.

3. Autres renseignements complémentaires

3.1. Précision et erreurs

3.1.1. Évaluation comparative. Les tests d'évaluation comparative normalisés – du type de ceux cités par les vendeurs d'algorithmes lorsqu'ils sont interrogés sur la précision de leur TRF – sous-estiment l'erreur qui serait constatée dans la grande majorité des applications médico-légales. Dans les tests normalisés, la qualité des images est souvent beaucoup plus élevée et les bases de données consultées sont souvent beaucoup plus petites que dans les enquêtes criminelles. Les images des tests sont également celles de sujets « dociles » qui n'essaient pas d'éviter l'identification et qui regardent droit dans la caméra sous un bon éclairage. Ces conditions ne sont pas représentatives de celles que l'on rencontre habituellement dans les enquêtes.

3.1.2. Listes de candidats. Dans les tests d'évaluation comparative, la précision est souvent mesurée comme la probabilité qu'un visage correspondant d'une base de données en particulier soit renvoyé en tant que correspondance possible la mieux classée. Toutefois, comme cette probabilité est réduite pour les images de faible qualité utilisées dans les enquêtes criminelles, cela signifie que les opérateurs humains doivent examiner de grandes listes d'images – parfois plus de 100 – pour s'assurer de ne pas manquer une correspondance. Des études récentes ont testé la précision de la correspondance des visages du personnel utilisant la TRF dans son travail quotidien et ont trouvé des erreurs dans 50 % de ses décisions d'examen de liste de candidats [13], malgré l'affichage de listes de seulement huit images. Les erreurs indiquent que la probabilité que des opérateurs humains sélectionnent une personne innocente à partir d'un visage de la liste des candidats de la TRF – une identification constituant un « faux positif » – est alarmante.

3.1.3. Erreurs en cascade. Il peut être tentant de conclure que, puisque la TRF est utilisée pour générer des pistes d'enquête, et non pour porter des jugements définitifs sur l'identité, le problème des « faux positifs » dans ce contexte n'est pas grave. Cependant, une meilleure compréhension des influences du biais cognitif dans la prise de décision médico-légale – et des effets cumulatifs qu'une erreur à un stade de l'enquête peut avoir sur l'interprétation des sources d'information

suivantes [5] – indique que même des pistes d'enquête erronées peuvent conduire à des résultats graves tels qu'une arrestation injustifiée [22].

3.2. Mise en œuvre

3.2.1. Mise en œuvre du système. La mise en œuvre correcte des systèmes de reconnaissance faciale est plus complexe que le simple achat du dernier algorithme. Si les vendeurs d'algorithmes ne doivent pas se contenter d'installer le logiciel, on ne peut pas attendre d'eux qu'ils assurent la surveillance nécessaire. Pour cela, il est nécessaire de procéder à des tests « internes » en utilisant des images, des procédures et des outils représentatifs du travail de l'organisation, plutôt que de s'appuyer sur les taux de rendement des vendeurs ou sur des tests d'évaluation normalisés. Il est essentiel que ces tests tiennent compte à la fois de la précision de la TRF et du traitement humain. Il faut que du personnel spécialisé soit responsable de l'ensemble du système, y compris les algorithmes, la conception du flux de travail et la délégation des tâches aux humains. Le personnel doit avoir l'expertise nécessaire pour mesurer la précision et signaler les risques associés au déploiement d'algorithmes pour des utilisations précises.

3.2.2. Soutien du système. Le coût pécuniaire de la mise en œuvre d'un système de reconnaissance faciale dépasse de loin celui de l'achat et de l'installation de la TRF dans les systèmes de TI et de son intégration dans le flux de travail. Les allocations budgétaires pour la TRF doivent comprendre des fonds pour assurer une utilisation et une surveillance appropriées du système, y compris des tests réguliers de la précision du système de reconnaissance faciale dans le cadre d'un déploiement opérationnel, la sélection d'opérateurs humains, la formation du personnel, le perfectionnement professionnel des équipes de spécialistes de la TRF, la surveillance et la vérification de routine de l'utilisation de la TRF, et la conception de l'expérience utilisateur.

Références

[1] Burton, A.M., R. Jenkins et S.R. Schweinberger (2011). « Mental representations of familiar faces », *British Journal of Psychology*, vol. 102, n° 4, p. 943-958.

[2] Burton, A.M., et R. Jenkins (2011). « Unfamiliar face perception », *The Oxford Handbook of Face Perception*, 28, p. 287-306.

- [3] Zhou, X., et R. Jenkins (2020). « Dunning–Kruger effects in face perception », *Cognition*, vol. 203, art. 104345.
- [4] Ritchie, K.L., F.G. Smith, R. Jenkins, M. Bindemann, D. White et A.M. Burton (2015). « Viewers base estimates of face matching accuracy on their own familiarity: Explaining the photo-ID paradox », *Cognition*, vol. 141, p. 161-169.
- [5] Dror, I.E., R.M. Morgan, C. Rando et S. Nakhaeizadeh (2017). « The bias snowball and the bias cascade effects: Two distinct biases that may impact forensic decision making », *Journal of Forensic Sciences*, vol. 62, n° 3, p. 832-833.
- [6] Howard, J.J., L.R. Rabbitt et Y.B. Sirotin (2020). « Human-algorithm teaming in face recognition: How algorithm outcomes cognitively bias human decision-making », *PLoS ONE*, vol. 15, n° 8, p. e0237855.
- [7] Wilmer, J.B., L. Germine, C.F. Chabris, G. Chatterjee, M. Williams, E. Loken, K. Nakayama et B. Duchaine (2010). « Human face recognition ability is specific and highly heritable », *Proceedings of the National Academy of Sciences*, vol. 107, n° 11, p. 5238-5241.
- [8] Zhou, X., A.M. Burton et R. Jenkins (2021). « Two Factors in Face Recognition: Whether you know the person’s face and whether you share the person’s race », *Perception*, vol. 50, n° 6, p. 524-539.
- [9] *Facial Recognition Technology: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks*, États-Unis, Government Accountability Office. Sur Internet : <https://www.gao.gov/products/gao-21-518>; voir aussi <https://www.perpetuallineup.org/>, <https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021> (chapitre 9).
- [10] <https://www.flawedfacedata.com>
- [11] Grother, P., et coll. (2019). *Face Recognition Vendor Test (FRVT) – Part 3: Demographic Effects*. Sur Internet : <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf>
- [12] Note en bas de page 1, p. 7-9; Davies, B., et coll. (2018). *An Evaluation of South Wales Police’s Use of Automated Facial Recognition*. Sur Internet : <https://crimeandsecurity.org/feed/afr>
- [13] White, D., J.D. Dunn, A.C. Schmid et R.I. Kemp (2015). « Error rates in users of automatic face recognition software », *PLoS ONE*, vol. 10, n° 10, p. 1-14.
- [14] Steblay, N., J. Dysart, S. Fulero et R.C. Lindsay (2001). « Eyewitness accuracy

rates in sequential and simultaneous lineup presentations: A meta-analytic comparison », *Law and Human Behavior*, vol. 25, n° 5, p. 459-473.

[15] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1010815/Surveillance_Camera_Code_of_Practice__update_.pdf

[16] <https://www.theguardian.com/technology/2019/sep/29/plan-for-massive-facial-recognition-database-sparks-privacy-concerns>

[17] Wilmer, J.B., L. Germine, C.F. Chabris, G. Chatterjee, M. Gerbasi et K. Nakayama (2012). « Capturing specific abilities as a window into human individuality: The example of face recognition », *Cognitive Neuropsychology*, vol. 29, n° 5-6, p. 360-392.

[18] White, D., D. Guilbert, V.P. Varela, R. Jenkins et A.M. Burton (2022). « GFMT2: A psychometric measure of face matching ability », *Behavior Research Methods*, vol. 54, n° 1, p. 252-260.

[19] Phillips, P.J., A.N. Yates, Y. Hu, C.A. Hahn, E. Noyes, K. Jackson, J.G. Cavazos, G. Jeckeln, R. Ranjan, S. Sankaranarayanan, J. Chen, C.D. Castillo, R. Chellappa, D. White et A.J. O'Toole (2018). « Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms », *Proceedings of the National Academy of Sciences*, vol. 115, n° 24, p. 6171-6176.

[20] Towler, A., R.I. Kemp et D. White (2021). « Can face identification ability be trained? Evidence for two routes to expertise », dans M. Bindemann (dir.), *Forensic face matching: Research and practice*, Oxford University Press.

[21] Towler, A., R.I. Kemp, A.M. Burton, J.D. Dunn, T. Wayne, R. Moreton et D. White (2019). « Do professional facial image comparison training courses work? », *PLoS ONE*, vol. 14, n° 2, p. e0211037.

[22] <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>