SciencesPo
CITIES AND DIGITAL TECHNOLOGY CHAIR

Working Paper

> N°05/2019

Rendre « intelligentes » les caméras :

déplacement du travail des opérateurs de vidéosurveillance et redéfinition du soupçon

Florent Castagnino

SciencesPo

CITIES AND DIGITAL TECHNOLOGY CHAIR

The “Cities and Digital Technology” Chair of Sciences Po’s Urban School has been launched in March 2017 to better grasp the impact of digital technologies on urban governance. Funded by four sponsoring firms (Cisco, La Poste, RTE, Caisse des Dépôts), the Chair aims to create new research fields exploring the interaction between digital technology and cities in an empirical and comparative perspective.

Rendre « intelligentes » les caméras : déplacement du travail des opérateurs de vidéosurveillance et redéfinition du soupçon

Florent Castagnino, sociologue, chercheur contractuel à Télécom Paris (UMR i3)
florent.castagnino@telecom-paris.fr

Résumé

Ce papier défend la thèse que les effets, supposément dissuasifs et normalisateurs, des pratiques de surveillance ne sont pas automatiques, et ce quel que soit leur degré d'automatisation. En prenant le cas de la vidéosurveillance dans les gares et les trains en France, l'article met en avant les limites tant cognitives que matérielles que rencontrent les opérateurs travaillant « derrière » les caméras. Notre travail montre que l'introduction des techniques d'intelligence artificielle ne permet pas *ipso facto* de dépasser ces limites. En effet, l'automatisation supposément induite par l'intelligence artificielle ne supprime pas le travail des opérateurs, mais le déplace. Ce déplacement de l'objet doit ainsi conduire au déplacement de l'enquête empirique vers l'étude du travail des informaticiens. Le papier montre alors comment ces derniers doivent formaliser mathématiquement une partie du travail des opérateurs afin de le stabiliser dans des règles algorithmiques. Dans cette tâche d'abstraction, ils opèrent alors des choix plus ou moins implicites de ce qui est « suspect », et ainsi de ce qu'il « faut surveiller ».

Mots clés : vidéosurveillance, intelligence artificielle, gares, opérateurs, automaticité

Introduction

Les récents progrès du traitement automatisé des images permettent désormais à des entreprises de proposer des systèmes de « vidéosurveillance intelligente », que ce soit aux municipalités, aux opérateurs de transport ou encore aux gérants de centres commerciaux[1]. L'apport de l'intelligence artificielle (IA) serait censé venir pallier les insuffisances matérielles et cognitives des opérateurs humains pour rendre, enfin, la vidéosurveillance « efficace »[2]. Pour comprendre ce que l'intelligence artificielle fait à la vidéosurveillance (et à ses opérateurs), il importe selon nous de partir des résultats que les sciences sociales ont déjà fournis sur la vidéosurveillance « non intelligente ». Ce préalable nous paraît nécessaire à deux égards : il permet premièrement de ne pas reproduire les biais méthodologiques de certaines études ; il permet également de bien mesurer les changements que cette technologie implique dans le travail des opérateurs de vidéosurveillance.

Sur le plan méthodologique, de nombreuses études postulent trop rapidement que les systèmes de vidéosurveillance remplissent automatiquement les objectifs pour lesquels ils ont été mis en place (Smith, 2012). Ce biais fonctionnaliste et déterministe – contre lequel Armstrong et Norris mettaient déjà en garde dans leur ouvrage séminal (Norris et Armstrong, 1999) – se retrouve plus largement dans les travaux relevant des *surveillance studies* (Castagnino, 2018 ; Kroener et Neyland, 2012). Smith note plus particulièrement que « la plupart des auteurs semblent presque oublier que les caméras de surveillance ne sont pas conscientes ni autonomes et requièrent, pour être effectives, un contrôle constant par des êtres humains en situation de travail » (Smith, 2004, p. 377)[3]. Pour eux, tout se passe comme si les objets remplissaient leur script (Akrich, 1987) de manière automatique, oubliant alors d'étudier la « technologie en usage » (Orlikowski, 1992). Les opérateurs de vidéosurveillance sont ainsi doublement invisibilisés : par leur condition de travail (dans des salles fermées, derrière leurs écrans) et par le déterminisme technologique de certaines approches scientifiques.

---

[1] Comme l'attestent plusieurs articles de presse professionnelle ou généraliste : *La reconnaissance faciale progresse, sous la pression des industriels et des forces de l'ordre*, Le Monde, 14/10/2019, https://www.lemonde.fr/pixels/article/2019/10/14/sous-la-pression-des-industriels-et-des-forces-de-l-ordre-la-reconnaissance-faciale-progresse_6015370_4408996.html ; *Amazon vend sa reconnaissance faciale à la police, les experts s'alarment*, Le Big Data, 04/05/2019, ; https://technologie.securitas.fr/decryptage/videosurveillance-intelligence-artificielle-service-surete ; Voir également cette note du Centre de Recherche de l'Ecole des Officiers la Gendarmerie Nationale : *Reconnaissance faciale et contrôles préventifs sur la voie publique, l'enjeu de l'acceptabilité*, CREOGN, https://www.gendarmerie.interieur.gouv.fr/crgn/Actus/Reconnaissance-faciale-et-controles-preventifs-sur-la-voie-publique-l-enjeu-de-l-acceptabilite.

[2] Plusieurs études rappellent régulièrement la faible efficacité de la vidéosurveillance pour réduire la délinquance et la criminalité. En sommant 22 études par exemple, Welsh et Farrington (2003) montrent que l'utilisation de la vidéosurveillance a permis une baisse générale de la criminalité de seulement 4 %. Dans une revue de littérature du Home Office de Londres, sur 14 études, une seule montre une baisse de la criminalité (Gill et al., 2005).

[3] Notre traduction de l'original : « most writers seem almost to forget that, by and large, CCTV cameras are neither conscious, nor autonomous, and require, in order to be effective, constant monitoring and control by human beings in a work-like situation ».

Concernant les résultats des travaux s'étant penchés empiriquement sur le travail des opérateurs, nombreux sont ceux qui se sont concentrés sur leurs pratiques discriminatoires (Coleman et Sim, 2000 ; Walby, 2005)[4]. D'autres études plus ethnographiques s'intéressent aux gestes, aux attitudes, à l'ambiance de travail des opérateurs (Klauser, November et Ruegg, 2006 ; Le Goff, 2013 ; Smith, 2004). C'est plutôt dans cette dernière veine de travaux que l'analyse présentée ici se situe. Que font concrètement les opérateurs de vidéosurveillance, qu'est-ce que « vidéosurveiller » ? L'un des résultats qui ressort de ces diverses études est que les opérateurs ne font pas que surveiller. Du moins, l'activité qui consiste à fixer un écran et à rechercher de manière active quelqu'un en train (ou qui serait susceptible) de commettre un acte répréhensible ne constitue pas la majeure partie du travail des opérateurs (Bonnet, 2012 ; Helten et Fischer, 2004 ; Le Goff, 2013)[5]. Que font donc les opérateurs de vidéosurveillance ? Les études de Fischer et Helten et celle de Bonnet convergent pour dire que dans la grande majorité des cas, les caméras sont utilisées pour la gestion de l'espace (réagir en cas d'alarme incendie, contrôle de la bonne tenue d'un équipement, repérage des problèmes dans les parkings, des problèmes de propreté, de ventilation, etc.). Ainsi, l'activité de surveillance proprement dite apparaît presque secondaire dans l'activité de travail des opérateurs. Cette répartition des tâches ne peut s'expliquer *a priori* par un manque de professionnalisme des opérateurs. S'ils font majoritairement autre chose que surveiller le comportement des gens, c'est que globalement il ne se passe pas grand-chose de répréhensible sous l'œil des caméras, et que la probabilité pour qu'un événement soit repéré et donne lieu à une intervention policière est relativement très faible. Toutes les études qui ont cherché à mesurer l'effet de la vidéosurveillance dans la lutte contre la délinquance et la criminalité de rue concluent à un impact très faible, voire inexistant[6]. Comme nous le verrons dans cet article, cette faible efficience s'explique en partie par les capacités limitées de détection et d'attention visuelle des opérateurs, mais également par des problèmes de maintenance des caméras, ainsi que des difficultés dans l'extraction et le stockage des images.

[4] Depuis les travaux pionniers de Norris et Armstrong au Royaume-Uni (1999), la recherche qualitative a bien démontré que les opérateurs incorporaient et réifiaient les valeurs culturelles dominantes et les stéréotypes dans leur classification de la réalité sociale. La suspicion est en effet un construit social et plusieurs études ont montré que l'attention des opérateurs était plus déterminée par les catégories (sociales, raciales, de genres) prêtées aux individus que par leur comportement.

[5] Dans leur étude sur les opérateurs de vidéosurveillance dans des centres commerciaux à Berlin, Helten et Fisher montrent que cet usage de la vidéosurveillance ne représente qu'environ 16 % du temps de travail (Helten et Fischer, 2004). Dans son étude sur le travail des opérateurs municipaux de vidéosurveillance, Le Goff estime que le balayage général des caméras pour vérifier que tout est en ordre constitue environ 20 % du temps de travail d'un opérateur (Le Goff, 2013), auquel il faut rajouter une utilisation plus active des caméras où l'opérateur est à la recherche du flagrant délit. Dans son étude sur les gares et les centres commerciaux, Bonnet rejoint également ce résultat en montrant que les usages « sécuritaires de la vidéosurveillance» (directement liés à la lutte contre la délinquance) sont largement minoritaires dans le travail des opérateurs (Bonnet, 2012).

[6] Outre les résultats déjà mentionnés dans la note 2, rajoutons que dans une nouvelle méta-évaluation de 44 études internationales, Welsh et Farrington notent qu'un effet positif de la vidéosurveillance sur la réduction de la délinquance n'est prouvé que dans la surveillance des parkings (Welsh et Farrington, 2007, p. 8). Sur une sélection de 44 études (légèrement différentes de celle de Welsh et Farrington), des chercheurs californiens concluent que 59 % estiment que les caméras n'ont pas d'effet, un effet incertain ou des effets indésirables sur la réduction de la délinquance. Sur les 19 études prouvant qu'il n'y a pas d'effet statistique sur la réduction de la délinquance, 52,6 % relèvent même une augmentation des crimes (Cameron et al., 2008, p. 4).

Les tenants de la « vidéosurveillance intelligente » promettent alors des systèmes qui automatiseraient une grande partie du travail des opérateurs. On retrouve dans leurs arguments les topiques propres aux promoteurs des procédés algorithmiques nourries aux « big data » (Ayres, 2007 ; Castro, 2016 ; Milgram, 2013) : une plus grande capacité de traitement (ici des images de vidéosurveillance) et une plus grande neutralité que ne le permet l'activité humaine (l'algorithme n'étant pas censé être influencé par sa position sociale). Cette nouvelle technologie permettrait alors de rendre la vidéosurveillance réellement « efficace », en assurant l'effet supposément dissuasif des caméras et en fournissant une aide aux interventions des forces de l'ordre sur le terrain. Le développement de la « vidéosurveillance intelligente » pourrait en effet laisser croire que l'on se dirige vers des systèmes où l'opérateur humain tend à disparaître ou du moins vers des systèmes où l'opérateur humain est dépossédé d'une partie importante de son travail. Il convient selon nous de ne pas faire preuve de fonctionnalisme, y compris dans l'étude de systèmes fortement automatisés. L'un des principaux arguments développés dans ce papier consiste à montrer que les procédés d'automatisation (qu'ils soient organisationnels ou algorithmiques) ne suppriment pas le travail des opérateurs, mais le déplace. Dès lors, c'est ce déplacement et ses effets qui doivent être étudiés. Pour tenter de rendre plus efficient le travail des opérateurs, plusieurs parties importantes de leur travail sont déplacées : le choix des images à surveiller tend à être déplacé dans des procédures de visionnages automatisées, tandis qu'une partie du jugement sur les faits détectés (sont-ils importants ? faut-il prévenir les forces d'intervention ?) peut être déplacé dans un logiciel générant automatiquement des alarmes que l'opérateur doit traiter. Ainsi, les gestionnaires de systèmes de vidéosurveillance tentent d'automatiser une partie du nécessaire « travail d'enquête » qu'effectuent les opérateurs. Ces déplacements d'une partie du travail supposent cependant un accord sur les bonnes pratiques ainsi que sur les priorités d'action (que faut-il repérer ?), afin de les formaliser dans des procédures organisationnelles ou algorithmiques. Cette tâche d'abstraction du travail n'a rien d'évident pour celles et ceux qui la mettent en œuvre. Nous verrons ainsi que les automatismes permis par ce travail d'abstraction n'assurent pas mécaniquement l'automaticité des effets des pratiques de surveillance : le travail d'enquête des opérateurs n'est pas supprimé, mais seulement déplacé.

Cet article propose d'explorer les effets d'une automatisation d'une fonction de surveillance, à partir d'une enquête de terrain réalisée auprès d'opérateurs de vidéosurveillance de gares ferroviaires et de train en France (voir l'encadré méthodologique). Dans un premier temps, l'article analyse leur travail de « surveillance active », en se centrant sur les pratiques de rationalisation de leur activité de visionnage. Il explore également le « travail d'enquête » que les opérateurs doivent mener, au-delà de la simple détection d'un élément ou comportement suspect. Dans un second temps, le papier montre que l'effectivité des détections et leur plus-value dans le travail de surveillance sont également minées par les limites matérielles des caméras (maintenance, accès et stockage essentiellement), enjeux rarement traités dans les études disponibles. Enfin, la troisième partie montre que l'automatisation d'une partie de l'activité des opérateurs via des systèmes « intelligents » ne supprime en rien le travail d'enquête qu'ils doivent mener : il le déplace, en partie, vers le travail des informaticiens, ces derniers venant alors modifier les définitions professionnellement admises de ce qui est « suspect », et ainsi de ce qu'il « faut surveiller ».

Encadré méthodologique : enquêter sur les opérateurs de vidéosurveillance dans le milieu ferroviaire en France

Ce papier s'appuie sur une enquête de terrain menée dans le cadre d'une recherche doctorale (2012-2017), comparant les dispositifs de prévention des accidents et les dispositifs de prévention de la délinquance, dans le milieu ferroviaire en France (Castagnino, 2017). Les données proviennent essentiellement d'entretiens avec les professionnels de la SNCF (ou travaillant pour l'entreprise) et des observations de leurs pratiques de travail. Plus précisément 43 entretiens semi-directifs furent menés avec des agents de la Surveillance Générale (le service interne de sûreté de la SNCF, dite Suge), des policiers intervenants en gare, et des agents de sécurité privée embauchés par la SNCF. Si les données mobilisées dans ce WP concernent avant tous les entretiens et observations menées avec des acteurs de la vidéosurveillance, l'analyse est également nourrie des données recueillies auprès des autres professionnels de la sûreté.

Les deux premières parties de l'article examinent plus précisément le travail de trois types d'opérateurs de vidéosurveillance, qui se distinguent par leurs conditions d'exercice et leur statut. La troisième partie repose elle sur l'analyse d'un projet de développement de « vidéosurveillance intelligente ». Nous détaillons ci-dessous l'environnement professionnel des différents acteurs dont nous analysons l'activité de travail tout au long de l'article.

La première partie du papier revient sur le travail de visionnage des images, qui est documenté à partir de l'analyse du travail de deux types d'opérateurs vidéo :

1) Les « vidéopatrouilleurs » de la Suge, présents au PCNS

La Suge est le service de sûreté interne de la SNCF, avec des agents en uniforme, armés et disposant de certains pouvoirs de police judiciaire. La Surveillance Générale est organisée territorialement en Zones Sûretés, la plupart des agents étant affectés à des missions de patrouille et de sécurisation des gares et des trains. Certains agents sont spécifiquement dédiés à la vidéosurveillance, et travaillent au Poste de Commande National Sûreté (PCNS). Le PCNS n'est pas qu'une salle de vidéosurveillance. C'est également une salle de gestion de crise en cas d'événement sûreté et c'est surtout le centre de pilotage et d'assistance de toutes les équipes de la Suge déployées sur le réseau ferroviaire. Concrètement le PCNS est une grande salle de commandement avec plusieurs « tables » (ensemble de postes de travail) qui traitent les appels d'urgence et sollicitations des équipes de la Suge sur un territoire particulier. À côté des agents travaillant « sur table », on trouve les opérateurs vidéo, appelés en interne « vidéopatrouilleurs ». Ces derniers effectuent des rondes virtuelles en regardant les images de vidéosurveillance. Ils peuvent également assister les équipes de la Suge lors de missions spéciales.

2) Les opérateurs du Centre de Surveillance de la gare du Nord

La gare de Nord dispose d'un Centre de surveillance où sont regroupés trois services : maintenance, sécurité incendie, vidéosurveillance. Plus précisément,

Gares & Connexions, gestionnaire de la gare et assurant directement la maintenance, sous-traite les missions de sécurité incendie et de visionnage des caméras de vidéosurveillance à deux sociétés privées. Ce sont les opérateurs de vidéosurveillance présents dans ce Centre de surveillance que nous avons pu observer pendant 3 jours.

La deuxième partie du papier analyse les enjeux matériels que pose l'usage de caméras de vidéosurveillance. C'est au sein de la Zone Sûreté Sud-Est que nous avons pu le plus échanger avec les référents vidéo qui sont en charge du travail d'extraction des bandes de vidéosurveillance en cas de réquisition policière, ainsi que de la supervision de la maintenance des systèmes de vidéosurveillance.

La troisième partie porte sur les tentatives d'automatisation du travail de visionnage grâce aux apports de l'intelligence artificielle. Nous avons pu faire des entretiens avec 3 chercheurs en informatique de l'INRIA travaillant sur un projet de recherche de vidéosurveillance intelligente avec la SNCF. Nous avons également analysé certaines de leurs publications liées aux résultats du projet de recherche.

1 Les techniques de rationalisation du travail de visionnage

Afin de bien saisir les effets de l'introduction des techniques d'intelligence artificielle dans les systèmes de vidéosurveillance, il convient d'abord de décrire précisément la façon dont les opérateurs travaillent. Nous nous appuierons pour cela sur l'analyse du travail de deux types d'opérateurs vidéo : ceux de la Surveillance Générale, le service interne de sûreté de la SNCF (appelés « vidéopatrouilleurs », et qui travaillent au PCNS) et ceux de la société de sécurité privée employés par *Gares & Connexions* à la gare du Nord, à Paris (voir l'encadré méthodologique). Puisque les techniques d'intelligence artificielle s'offrent comme solution pour traiter des volumes considérables de données, regardons en premier lieu comment les opérateurs et leur encadrement tentent déjà d'optimiser le visionnage des images de vidéosurveillance : nous présenterons d'abord des opérations de réduction propres aux opérateurs (1.1), puis la technique de séquençage automatique qui oriente le travail des agents (1.2).

Les opérateurs de vidéosurveillance en gare que nous avons pu observer sont placés dans une situation de travail inconfortable. Alors qu'ils ont à leur disposition un volume important de données (constitué par l'ensemble des images provenant des caméras et diffusées sur plusieurs moniteurs), ils se retrouvent souvent incapables de repérer des éléments jugés pertinents dans leur tâche de surveillance. Le visionnage d'images de vidéosurveillance en temps réel est en effet une activité ingrate et difficile au terme de laquelle peu d'éléments sont effectivement repérés ou mènent à une action sur le terrain. En effet, lorsque l'on s'intéresse

à l'activité de travail proprement dite (Bidet, 2006 ; Bidet et Vatin, 2016)[7] des opérateurs, les limites cognitives de la détection visuelle sont à considérer. La multiplication des caméras de vidéosurveillance n'implique pas la multiplication des opérateurs qui les visionnent effectivement. Le ratio « caméras visionnées / caméras en place » est en général très faible. En gare du Nord par exemple, l'opérateur de vidéosurveillance dispose de 1008 caméras dans la gare, son équipement lui permettant seulement d'afficher 12 écrans. Avec un système de séquençage automatique, l'opérateur peut faire défiler une quarantaine de caméras en 5-10 min environ. Ainsi, par tranche de 5-10min, l'opérateur visionne environ moins de 4 % des caméras. Si le cas de la gare du Nord minimise le plus ce ratio, dans les autres gares équipées que nous avons pu observer, l'ordre de grandeur est similaire.

Dans ces conditions, comment les opérateurs effectuent-ils leur travail ? Nous ne reviendrons pas sur les différents stéréotypes et catégorisations qui orientent indéniablement le travail des opérateurs vidéo (cf. *supra*). Nous nous concentrons ici sur les façons de repérer que développent les acteurs, au-delà des difficultés que nous venons de mentionner.

1.1. Opérations de réduction

D'une manière similaire au travail policier, les opérateurs de vidéosurveillance (agents de la SNCF ou agents privés) vont procéder à des catégorisations de leurs « clientèles » afin de pouvoir concentrer leurs ressources et leur capacité d'attention. Une séparation claire s'établit entre les personnes « qui ne posent pas de problèmes » et « les autres », entre les « vrais voyageurs » et les « vrais parasites », comme les policiers peuvent définir leurs « vrais », « petits » ou faux « clients » (Boussard, Loriol et Caroly, 2006). Ainsi, les catégories ethniques et sociales du sens commun vont constituer « des instruments de travail et font partie de cet ensemble de connaissances pratiques qui forme l'arrière-plan, la référence du travail policier » (Lévy, 1987, p. 31), mais également des opérateurs de vidéosurveillance. Au-delà de ces catégories et représentations disponibles (typiquement, dans le cas des gares ferroviaires, les catégories de « jeunes », « SDF », « Roumains », etc.)[8], les opérateurs de vidéosurveillance vont procéder à des réductions comportementales. Ces opérations de réduction permettent d'une certaine façon de rationaliser le travail de visionnage des écrans de vidéosurveillance. Elles permettent aux opérateurs de focaliser leur attention sur des éléments précis lors du balayage des écrans, moins susceptibles de produire des suspicions discriminatoires que les catégorisations *a priori*.

La première réduction est celle du mode opératoire. Les agents apprennent à repérer les façons de procéder, de « travailler » – pour reprendre leur terme – des personnes

[7] La sociologie de l'activité se distingue de la sociologie classique du travail, d'inspiration friedmanienne, en voulant étudier plus concrètement le travail *en actes*. Ce parti pris permet notamment de considérer les questions temporelles et la technicité du travail, souvent abstraites des études classiques. Cette sociologie de l'activité est ainsi attentive au caractère distribué de l'action, aux processus de définition et de redéfinition des situations par les acteurs, aux différents appuis qu'ils mobilisent pour s'orienter. Il ne s'agit pas de nier les contraintes des structures et des organisations, mais de toujours faire attention aux capacités critiques des acteurs pour les détourner, les adapter, les adoucir, les façonner : si elles leur préexistent, leur permanence, tout comme leur changement, est en partie à rechercher dans les ajustements des acteurs.

[8] Pour en savoir plus sur ces pratiques de catégorisations, cf. Castagnino, 2017

délinquantes. Toute une série de « petits détails » est effectivement associée, dans l'esprit des opérateurs, à un type d'actes répréhensibles. À propos des vols de téléphones portables par exemple, l'un des opérateurs du PCNS, en remontant dans les bandes vidéo (les bandes sont accessibles jusque dans les 72h), commente :

> *Quand on voit trois individus comme ça là, qui arrivent, habillés pareil, ils se dispersent sur le quai, ça ça attire l'attention, c'est des petits trucs. On les voit, le train s'arrête en gare, ils bloquent la porte, on sait qu'il y a quelque chose qui se prépare derrière, donc on peut anticiper*
> *(Vidéopatrouilleur 1, PCNS)*

Les pickpockets, bien que tous les opérateurs rencontrés admettent l'extrême difficulté à les repérer, sont toutefois eux aussi réduits à certaines attitudes :

> *Les pickpockets, sur le quai, ils vont regarder à droite, à gauche, ils vont se mettre derrière les touristes pour essayer...ça c'est des petits trucs qu'on acquiert et qu'on utilise après pour les localiser, les repérer. Ça change tout le temps, mais on s'adapte au fil du temps*
> *(Vidéopatrouilleur 2, PCNS)*

Ainsi, le repérage de méthodes de travail (gestes spécifiques, heures d'activité) est une technique des opérateurs de vidéosurveillance afin de réduire la focale de ce qu'ils doivent scruter.

La deuxième réduction concerne l'apparence vestimentaire des « clients ». Il s'agit ici de repérer « *les nouvelles têtes qui arrivent* » et de connaître les régulières.

> *En général, les individus qui commentent les méfaits, ils reviennent en général. Comme je disais, celui qui a volé ce matin, il va revenir peut-être cet après-midi. Il sera pas habillé pareil, mais il aura peut-être les mêmes chaussures, un petit truc qui va faire que ça va tout de suite faire tilt : lui on le connaît, on l'a déjà vu.*
> *(Vidéopatrouilleur 2, PCNS)*

> *des fois ils changent de vêtements, mais ils oublient la casquette, les baskets, une petite montre rouge. C'est un détail, mais ça permet de voir, de repérer. Et on peut revenir après sur la vidéo et faire une comparaison. Et après on peut le suivre voir s'il travaille, et anticiper avec l'équipe Suge, se placer en amont et voir ce que ça donne*
> *(Vidéopatrouilleur 1, PCNS)*

La mémoire est donc également une compétence valorisée dans ce travail. En effet, il n'est pas autorisé, par exemple, de faire une capture d'écran d'un individu et de l'imprimer en guise de « cible » à repérer, comme peuvent le faire certains opérateurs municipaux (Le Goff, 2013, p. 97). La transmission de la mémoire entre agents s'effectue plutôt par écrit. Les agents de la matinée peuvent noter des indications et descriptions d'individus sur leur main courante qui est à disposition des agents prenant la relève. Les descriptions écrites vont ainsi servir à aiguiller le regard des nouveaux agents lors de leur balayage des caméras.

Quoique ces deux techniques de réduction soient communes aux différents opérateurs que nous avons rencontrés, la réalité du travail est loin d'être similaire entre l'opérateur de la société de sécurité privée et l'opérateur de la Suge. Cette différence s'explique avant tout par les statuts professionnels de ces deux types d'opérateurs vidéo. Les opérateurs du Centre de Surveillance de la gare du Nord sont salariés d'une entreprise de sécurité privée, à qui est

sous-traitée une partie du travail de surveillance. Comme a pu le montrer Péroumal, les entreprises de sécurité privée sont essentiellement composées d'agents d'exécution, souvent faiblement qualifiés, se retrouvant dans ces emplois de surveillance et de gardiennage par nécessité (Péroumal, 2009)[9]. Leurs conditions de travail diffèrent largement des opérateurs de vidéosurveillance directement salariés par la SNCF, en tant qu'agent de la Surveillance Générale. Si les trajectoires professionnelles de ces deux groupes se recoupent en partie (notamment pour les agents provenant déjà du secteur de la sécurité : police, armée, pompier), leur statut de salarié interne et leurs attributions dérogatoires en tant qu'agents de la Suge (ils sont notamment armés et disposent de pouvoirs contraignants les rapprochant des forces de police ou de gendarmerie) leur confèrent un prestige social et un confort matériel supérieur aux agents de sécurité privée. Cette différenciation se traduit également dans la répartition des missions : les opérateurs de vidéosurveillance de la Suge travaillent au Poste de Commande National Sûreté (PCNS) et sont beaucoup plus intégrés au travail de patrouilles de leur collègue. Comme nous allons le voir, les incidences de cette dichotomie professionnelle se repèrent jusque dans le « travail d'enquête » que doivent mener les opérateurs.

1.2. Séquençage automatique et « travail d'enquête »

Au-delà du travail fastidieux de repérage, les opérateurs doivent également fournir une évaluation de la situation afin de décider s'il faut prévenir les services (de la Suge, de la police, des pompiers, d'autres personnels SNCF, etc.) susceptibles d'intervenir sur place. Les ressources de ces services d'intervention étant limitées, il s'agit pour les opérateurs de vidéosurveillance de ne pas les solliciter pour des problèmes que ces services jugeraient trop mineurs ou qui ne seraient pas de leur ressort[10]. Les opérateurs doivent donc participer à la « levée de doute » des événements ou éléments qu'ils détectent. Cette levée de doute peut être interprétée comme une enquête au sens de Dewey soit « la transformation contrôlée ou dirigée d'une situation indéterminée en une situation qui est si déterminée en ses distinctions et relations constitutives qu'elle convertit les éléments de la situation originelle en un tout unifié » (Dewey, 1967, p. 169)[11]. L'enquête est un processus au terme duquel l'incertitude est

[9] Comme le précise Péroumal, « Le gardiennage et la surveillance ne sont pas des professions qui s'exercent à travers une « vocation » […]. Si ces professions sont le prolongement d'une carrière exigeant une reconversion nécessaire, comme c'est le cas pour des agents provenant de la police, de l'armée ou des pompiers […], pour la plupart c'est l'urgence et la nécessité du moment qui les rabattent vers ce secteur. Dans bien des cas, les emplois de sécurité deviennent alors pour eux des « emplois refuges » assimilables à des espaces de relégation professionnelle » (Péroumal, 2009, p. 6). Il faut également noter que ces entreprises « absorbent également une main-d'œuvre dotée d'un capital scolaire relativement élevé, issue la plupart du temps d'anciennes colonies. Ces agents, qui connaissent parmi la population de la sécurité certainement les formes de déclassement les plus fortes, sont également confrontés à une vulnérabilité de leur condition qui, à l'exemple des étudiants d'origine africaine, se caractérise par une forme de salariat hybride évoluant entre plusieurs espaces (université, marché du travail, pays d'émigration et d'immigration) » (Péroumal, 2009, P ; 6).

[10] D'une sollicitation mesurée des services d'intervention dépend en partie la crédibilité des opérateurs, crédibilité qui ne va pas de soi, comme a pu le montrer Le Goff à propos des relations entre opérateurs municipaux et police nationale (Le Goff, 2013).

[11] Pour Dewey, l'enquête est une pratique qui se retrouve dans toutes les activités humaines (il en trouve le modèle dans le fonctionnement organique), ce qui lui permet de défendre une continuité entre le raisonnement du sens commun et l'activité scientifique (qui ne serait différenciés que par la formalisation des différentes

dissipée. C'est bien ce qui se joue lors des différentes levées de doute (formelles ou informelles) auxquelles procèdent les acteurs étudiés une fois un élément anormal détecté. Mais l'anormalité alors repérée n'est au début que soupçon : la situation repérée ne va pas de soi, mais il reste à le confirmer. Les levées de doute réalisées par les opérateurs de vidéosurveillance se rendent bien descriptibles par les différentes étapes que Dewey différencie dans sa théorie de l'enquête[12].
Dans cette sous-partie, nous insisterons sur les procédés techniques qui tentent de concentrer ce travail d'enquête sur des moments et des secteurs particuliers, et ce en cadrant le travail de visionnage. L'efficience de ce travail de cadrage – qui stabilise l'accord entre les services d'intervention et les opérateurs sur ce qu'il est pertinent de signaler – dépend en partie de l'intégration plus ou moins forte des opérateurs au travail des équipes d'intervention de la Suge.

Nous avons pu relever deux techniques de rationalisation du travail de visionnage (censées maximiser le nombre d'écrans visionnés par les opérateurs) que l'on retrouve au PCNS et au Centre de surveillance de la gare du Nord.

Le premier concerne l'ergonomie du poste de l'opérateur. Dans les deux cas considérés, l'ergonomie du poste est schématiquement similaire. Au second plan en hauteur, encastrés dans un mur ou surélevés par un bras mécanique, plusieurs écrans (4 dans un cas, 3 dans l'autre) servent uniquement à la visualisation. Sur le premier plan, les opérateurs disposent de moniteurs principaux qui leur permettent de commander telle ou telle caméra et d'afficher ses images à l'écran. Ce premier plan est en général constitué de deux moniteurs : l'un permet d'afficher la liste de toutes les sources vidéo, l'autre permet une visualisation. La liste peut prendre plusieurs formes : il peut s'agir d'une vraie liste, organisée en général par secteurs et où chaque caméra a un nom et un numéro, ou d'une carte des différentes gares et secteurs de gares, où sont renseignées les positions des caméras. Il suffit à l'opérateur de cliquer sur une caméra ou d'effectuer un « cliquer-glisser » entre la liste et l'écran de visualisation. Avec ce système, c'est donc l'opérateur qui a la main sur les caméras qu'il désire visionner. Ce procédé relativement intuitif est censé inciter l'opérateur à visionner le plus de caméras possible. De fait, cela place l'opérateur dans une attitude de vigilance : il n'a rien de précis à visionner, il fait appel à son expérience pour vigiler telle ou telle zone. Ainsi, si plusieurs zones sont régulièrement visionnées (dans le cas de la gare : les quais, les interconnexions, les parvis), d'autres sont quasiment non surveillées.

> *Je vais te montrer, une sortie qui donne vers la rue Maubeuge. Il n'y a pas de raison précise pour la surveiller, c'est un portail où les gens rentrent avec leurs voitures, des gens qui travaillent ici, donc on n'a pas grand-chose à surveiller. On a un agent sur place avec une guérite à côté du portail, donc on la regarde même pas*

étapes de l'enquête). Reprendre sa définition de l'enquête ne signifie pas que nous adhérions complètement à ses analyses épistémologiques. Ce qui nous intéresse surtout ici est la définition de l'enquête comme processus pratique.

[12] Ces différentes étapes étant : une situation indéterminée qui provoque un doute ; l'institution d'un problème où l'on reconnaît que la situation indéterminée nécessite une enquête ; la détermination de la solution du problème, où l'on émet les différentes hypothèses possibles ; le raisonnement, qui consiste à évaluer théoriquement les différentes hypothèses émises ; l'expérimentation, où les hypothèses sont testées et au terme de laquelle l'incertitude sur la situation initiale est supprimée

(Opérateur vidéo, Centre de surveillance en gare)

Certaines caméras ne sont ainsi pas jugées utiles, soit qu'il ne se passe rien dans la zone qu'elle couvre, soit qu'elles sont obsolètes en raison d'un autre type de surveillance déjà présent (ici la guérite).

La seconde technique de rationalisation consiste en la programmation de séquences de visionnage automatiques. Que ce soit dans le cas de l'opérateur de la Suge ou des opérateurs privés, il existe une programmation de séquences où défilent automatiquement certaines caméras, pendant un temps donné, sans action de l'opérateur (outre le démarrage d'une séquence). Ainsi, cela permet de balayer plus rapidement un secteur. Ces séquences ont soit été programmées par les responsables des agents soit par les agents eux-mêmes. Cette automatisation du travail de visionnage cible ainsi leur action sur un champ plus réduit, et donc plus maîtrisable, du phénomène à surveiller. Comme l'a déjà noté une étude sur le travail d'opérateurs de vidéosurveillance d'un réseau autoroutier : « La planification du système de surveillance et la détermination de ses capacités techniques répondent en premier lieu à une conception plus ou moins explicite des objets à risques, des situations à risques, des personnes à risques ou encore des espaces à risques. Cette planification du système rend la surveillance plus performante et plus ciblée par rapport aux risques tels qu'ils ont été imaginés au départ » (Klauser, November et Ruegg, 2006, p. 37). En effet, la réalisation de séquences automatiques suppose que l'on ait défini en avance les lieux et les moments qui posent problème, cadrant ainsi le travail d'alerte en organisant l'activité autour de routines de visionnage.

La plus-value opérationnelle de ces techniques de rationalisation n'est en revanche pas similaire entre les opérateurs de sécurité privée et les opérateurs internes. En effet, la position de l'opérateur du Centre de Surveillance et son statut professionnel limitent fortement son intégration au travail des agents de la Suge ou ceux de la police. De fait, son activité est plus mise au service des agents de maintenance de la gare et du service incendie, placés à côté de lui au sein du Centre de Surveillance. Son travail de surveillance et d'alerte se fait en collaboration avec les agents de ces deux autres services. Pour autant, le fait de pouvoir techniquement choisir quelles caméras s'affichent ou de disposer des séquences automatiques ne suffit pas à maintenir la vigilance des agents, surtout lorsque l'on sait que les périodes de travail peuvent atteindre la demi-journée (7h-19h ou 19h-7h) dans le cas des opérateurs privés en gare du Nord[13]. Lorsque nous avons pu consulter le « Cahier des interventions » (une sorte de main courante des opérateurs), nous avons remarqué que plus

[13] Selon la *Convention collective nationale des entreprises de prévention et de sécurité* datant de 1985, le temps de travail des employés de ce secteur déroge au régime général. L'article 4 de l'avenant à la convention de 1987 stipule : « Il est convenu, par dérogation aux dispositions de l'article L. 212-1, que la durée quotidienne de travail effectif ne peut dépasser 12 heures pour les services englobant un temps de présence vigilante ». La semaine de travail ne peut excéder 4 fois 12 heures et un jour de repos minimum doit être accordé après toute période de 48h de service (article 5). Source : Legigrance [https://www.legifrance.gouv.fr, consulté le 25 mai 2017].

de la moitié (pour les jours où sont étions présents du moins) des interventions inscrites relevaient des activités des deux autres services du Centre de surveillance[14].

L'activité des opérateurs de la Suge est en revanche plus rythmée, notamment parce que leur intégration au PCNS leur permet d'effectuer d'autres missions que le simple visionnage d'images. Ils sont d'ailleurs en interne appelés « vidéopatrouilleurs », soulignant ainsi leur appartenance au collectif de travail des équipes Suge de terrain. Deux types de missions leur sont singuliers. Tout d'abord, les vidéopatrouilleurs répondent aux appels extérieurs (30% de l'activité). C'est un des agents du PCNS qui fait l'interface entre les équipes Suge et les vidéopatrouilleurs. Il s'agit notamment d'effectuer des analyses sur une zone ou d'assister une équipe sur un délit ou un crime :

> *Appel extérieur au vidéopatrouilleur. Il est sollicité pour une « constatation immédiate » suite à une suspicion de vol à l'arraché. Après avoir récolté les informations sur le lieu et l'heure exacte, le vidéopatrouilleur « remonte » dans les bandes vidéo afin de donner la description : « donc ouais t'as bien un type, jvois pas bien s'il est black, mais on dirait, surtout il a un pull à capuche, uni. Bah le coup classique oui, juste avant que les portes du train se ferment, il descend, au niveau du repère « S » effectivement, et puis il se met à trottiner. Donc ouais ça peut être votre type. Il fait environ 1m80 je dirais, corpulence normale. Je peux pas vraiment dire plus »*
> *(Visite commentée, PCNS)*

Les vidéopatrouilleurs effectuent également des « missions programmées » (environ 10% de l'activité). Il s'agit la plupart du temps d'accompagner les équipes Suge qui travaillent en civil, notamment pour la répression des vols. Concrètement, il s'agit de filer des voleurs suspectés afin de les prendre en flagrant délit.

L'activité principale des vidéopatrouilleurs (60 % du temps de travail) relève cependant du visionnage d'initiative. Les vidéopatrouilleurs effectuent des rondes vidéo et tentent de repérer des activités illicites. Même si ce travail est moins prenant et rapproche fortement la condition du vidéopatrouilleur des autres opérateurs de vidéosurveillance, son implication dans des opérations plus directes de lutte contre la délinquance lui fait accepter relativement bien cette tâche de visionnage plus passive, un peu ingrate (étant donné le peu d'« affaires » qu'elle génère). En outre, la proximité avec les équipes d'intervention leur permet d'anticiper les besoins et attentes des patrouilles de terrain, ce qui facilite grandement leur travail de levée de doute et d'enquête sur les détections qu'ils effectuent.

Malgré ces différences professionnelles et statutaires, nous avons donc retrouvé les mêmes techniques de rationalisation du travail de visionnage. Cette rationalisation (du moins ces tentatives) apparaît nécessaire dans la mesure où d'une manière générale les opérateurs sont

[14] C'est notamment souvent le cas lors des alertes du service incendie (qui intervient également pour des malaises de voyageurs). À la fin d'une matinée d'observation, l'opérateur de vidéosurveillance avait répertorié 5 de « ses » interventions, en notant : « Les ADSI ont été avisés ». Or, si les Agents De Sécurité Incendie ont bien été avisés, ce fut par leur propre système d'alarme et non pas par l'opérateur vidéo. Notre enquête ne nous permet pas de dire s'il s'agit là d'une pratique singulière ou bien collective et massive. Au regard de la littérature sur les opérateurs et nos observations, cette pratique nous semble tout de même représentative du fait qu'une majorité d'opérateurs s'ennuie et ne se sent pas réellement utile dans le travail. D'où cette récupération des alertes des autres services, en guise de légitimation de leur propre activité, que ce soit auprès de leur hiérarchie ou du prestataire qu'est le gestionnaire de gare.

dans une position de submersion d'images, qu'ils n'ont absolument pas le temps de visionner. En effet, mis à part les missions ciblées dont nous venons de parler, la majorité du temps de travail de ces opérateurs oscille entre une recherche active de situations anormales (plutôt rare) et une attitude plus réactive, où il s'agit d'attendre qu'il se passe quelque chose (plutôt la norme). Même dans sa forme passive, l'activité de visionnage n'a rien d'aisé et demande aux agents une concentration qu'il n'est pas facile de maintenir, notamment lorsque l'ennui survient (Smith, 2004). Cette vigilance est d'autant plus difficile à tenir que les opérateurs doivent également faire face aux limites matérielles que leur posent les caméras.

2 Des conditions matérielles de la vidéosurveillance

Les limites de l'efficience des caméras de vidéosurveillance ne proviennent pas seulement des capacités cognitives des agents et de l'organisation du travail, mais également des conditions matérielles d'exercice. Comme l'a montré Dubbeld dans le cas de la vidéosurveillance dans les gares néerlandaises, les capacités de surveillance peuvent être limitées par la complexité technologique et la maintenance qu'elle exige (Dubbeld, 2005). Si nous venons de voir que les caméras ne sont pas toujours visionnées, et que ce travail de visionnage est fastidieux et difficile, les caméras ne sont pas toujours également en état de fonctionner ou en état d'être utilisées. C'est ce que nous allons voir avec les problèmes de maintenance des caméras (2.1) et le travail d'extraction des bandes vidéo (2.2). Nous prenons ici le cas des caméras embarquées dans les trains, qui nous semblent le plus représentatif de ces limites matérielles touchant également les caméras situées en gares.

2.1. De la fragilité technique des caméras

Comme rappelé en introduction, les caméras de vidéosurveillance – tout comme beaucoup d'objets techniques – ne fonctionnement pas d'elles-mêmes. Souvent, les techniques et les objets sont soit étudiés comme des éléments solides et durables ayant de forts effets de structuration sur l'ordre social, soit comme des éléments très controversés ou en crise. C'est bien le cas de la vidéosurveillance régulièrement accusée d'être un facteur de « purification de l'espace » (Bannister, Fyfe et Kearns, 1998), d'exclusion des classes populaires (McCahill, 1998) ou encore de standardisation des villes (Bétin, Martinais et Renard, 2003). L'objet est ici considéré dans sa toute-puissance, sa durabilité et sa solidité, bien que fortement critiqué. De telles approches et conceptions de la technique ne permettent pas de saisir les opérations sécuritaires quotidiennes. Elles contrastent fortement avec la conception que les opérateurs que nous avons rencontrés et observés ont des caméras : celles de caméras vieilles, parfois obsolètes, avec un manque d'intégration des différents systèmes rendant difficile leur exploitation, et subissant une maintenance non optimale. Comme le montrent Pontille et Denis, à travers l'exemple de la signalétique dans le métro parisien, la « fragilité technique » n'est pas un élément anecdotique dans la vie des objets (et donc de ceux qui les utilisent et les maintiennent), mais bien un mode d'existence principal (Denis et Pontille, 2015)[15]. D'où

[15] Ceci ne signifie pas que l'immuabilité et l'immobilité des objets techniques sont une illusion, mais qu'il faut compléter ce mode d'existence par celui de la fragilité.

l'importance des opérations de maintenance et de leur étude (Graham et Thrift, 2007 ; Henke, 1999).

Si nous n'avons pas observé le travail de maintenance des caméras durant notre enquête, c'est sa défaillance que nous avons pu saisir en partie au travers du travail des utilisateurs des caméras. Plusieurs dysfonctionnements concernant la base matérielle assurant le fonctionnement des caméras de vidéosurveillance peuvent être soulignés. Pour les autres corps de métier de la SNCF, le système de vidéosurveillance n'apparaît pas prioritaire. Les problèmes de maintenance reflètent bien cette situation. En effet, le système de vidéosurveillance n'est pas une priorité pour les mainteneurs des trains. La vérification du bon fonctionnement des caméras de vidéosurveillance n'est pas encore une activité totalement intégrée au travail de maintenance. La forte pression des agents du Matériel pour rendre à l'heure les rames sur lesquelles ils travaillent les incite à hiérarchiser les éléments à maintenir. À ce titre, les caméras de vidéosurveillance n'apparaissent pas prioritaires, comme l'illustre ce propos d'un mainteneur :

> *ouais les caméras, si si on doit regarder…Mais bon j'avoue que je le fais pas automatiquement quoi. Moi le train faut qu'il puisse rouler surtout, et en toute sécurité*
> *(Mainteneur matériel, Visite commentée Technicentre)*

Bien évidemment, les utilisateurs des systèmes de vidéosurveillance ont une évaluation tout autre de l'importance à accorder à cette maintenance dans la mesure où il en va de leur condition de travail. Le référent vidéo de la DZS Sud-Est explique par exemple avoir dû réaliser un vrai travail de conviction, d'enrôlement des mainteneurs :

> *Maintenant, j'ai la légitimité ce qui est plus facile pour être écouté, mais voilà quoi. Au début on avait pas mal de problèmes techniques, parce que c'était pas leur trame de vérification régulière. La vidéo, c'est marqué "voyant qui marche pas"....en gros c'est pas grave. Et au-dessus toute la ligne managériale dit : "c'est pas grave", « attends, on a des problèmes de frein c'est quand même plus... important ». Alors est-ce que c'est plus important ? Je sais pas. Un train peut pas rouler, c'est sûr c'est grave. Si y a pas de vidéo, le train peut rouler. Mais là, comme je suis en train de traiter une agression sexuelle, si on peut choper le gars, est-ce que c'est pas aussi...voilà, à défaut de l'avoir évité. C'est ça qui faut leur faire comprendre, tout doucement en leur expliquant.*
> *(Référent Vidéo, Zone Sûreté)*

En plus de ce travail de conviction, ce référent a décidé d'effectuer une recension du parc de caméras sous sa responsabilité, afin de vérifier leur bon fonctionnement. Ce travail de vérification ne peut se faire à distance et nécessite d'aller sur place. Le Référent Vidéo de la DZS Sud-Est s'est fixé un objectif de 55 gares contrôlées par trimestre, ce qui lui semble ambitieux étant donné les problèmes d'accessibilité de certaines gares :

> *Pour faire Nevers et Clermont et y a que deux systèmes vidéo à ces endroits-là bah, il faut 2 jours quoi […] Y a pas mal de vidéo sur les Alpes, et c'est pas mal de vallées, la Tarentaise, la vallée de l'Arc, on peut pas prendre de train, y a pas beaucoup de trains. Si je m'arrête à Valence, j'attends le train d'après pour monter ...c'est dur quoi, c'est intenable, c'est pas le RER jveux dire. Donc il faut y aller en voiture, ça prend moins de temps, mais c'est pas…[sous-entendu : optimal]*

Ces situations soulèvent, encore une fois, la nécessité d'investiguer les pratiques quotidiennes de surveillance. La toute-puissance postulée des caméras de vidéosurveillance est largement remise en cause dès lors que l'on s'intéresse aux pratiques de travail effectives des opérateurs

et aux conditions matérielles de leur travail. Nous avons pu observer un autre indicateur de cette fragilité technique au travers du travail d'extraction des bandes des systèmes de vidéosurveillance embarqués dans certains trains.

2.2. Le travail d'extraction des bandes vidéo : limite supplémentaire à l'efficience de la vidéosurveillance

Notons tout d'abord qu'il n'existe pas, au sein de l'entreprise de transport, de service centralisé spécifiquement en charge des caméras. Ainsi, lors d'achats de nouveaux systèmes, la cohérence avec ceux déjà existants n'est pas un problème soulevé. Personne, que ce soit au niveau de la direction des Achats ou du Matériel, n'a en charge les problèmes de comptabilité des nouveaux systèmes avec ceux déjà en place. D'où la coexistence de plusieurs systèmes de vidéosurveillance en fonction du type de matériel roulant considéré, comme l'illustre le propos de ce référent vidéo :

> *Là on a acheté de nouveaux TER (inaudible) de Bombarbier. Bombarbier on a déjà des trains équipés de vidéo. En fait, sur les trains on retire les disques durs, enfin les cassettes-disques durs. Et bien ils les ont modifiés histoire de revendre du nouveau matériel, c'est pas le même logiciel non plus…*
> *Nous on a le Régio2N, sur Bourgogne, normalement sur Auvergne ils auront du Régiolis d'Alstom. Alstom a déjà deux matériels roulants chez nous, les 24500 et les tram-trains sur l'ouest lyonnais, qui sont équipés déjà du système Alstom avec deux logiciels différents, deux formats de disques durs différents, et là ça sera encore un troisième*
> *(Référent Vidéo, Zone Sûreté)*

Ainsi, il y a un manque d'intégration des systèmes de vidéosurveillance qui a une conséquence directe sur le travail d'exploitation des images. Outre le fait qu'il faille former les opérateurs à différents logiciels, l'exploitation des bandes vidéo n'est pas optimale. En effet, dès lors que le format d'enregistrement du fichier vidéo n'est pas le même que précédemment, une opération de conversion est nécessaire pour pouvoir exploiter les images sur le logiciel précédent. Si cela peut paraître assez trivial, cela pose de réels problèmes aux opérateurs vidéo chargés d'établir des extractions (à la suite d'une réclamation de la police par exemple). En fonction du matériel, cette opération de conversion est plus ou moins évidente et rapide :

> *ah bah des fois t'as pas le temps, ou t'as pas ce qui faut pour assurer la compatibilité avec le logiciel. Bon c'est très rare, mais une fois j'ai dit qu'on n'avait pas d'enregistrement parce que c'était trop compliqué et long…C'est quand même grave tu vois !*
> *(Référent Vidéo, Zone Sûreté)*

Outre ces problèmes de compatibilité, le travail d'extraction des bandes vidéo n'a en soi rien d'évident. L'une des tâches du Référent Vidéo est d'aller récupérer les disques durs (sur lesquels sont enregistrées les vidéos) disposés dans les trains[16]. En effet, les systèmes mis en place permettent une conservation des bandes pour une durée maximum de 72h. Au-delà, les heures enregistrées sont « écrasées » par un nouvel enregistrement. C'est essentiellement une question de coût, dans la mesure où la loi autorise une conservation de 30 jours : les disques durs qui permettraient de conserver les bandes toute la durée autorisée

[16] À notre connaissance, il n'y a pas de visionnage en temps réel des caméras installées dans des trains, seulement des enregistrements des images. Selon les modèles de train, l'enregistrement est continu ou activé par le conducteur.

coûtent plus chers. En général, ce travail d'extraction est effectué à la suite d'une réquisition de la police. Il arrive cependant que le Référent Vidéo aille retirer certains disques de manière préventive, anticipant les problèmes pratiques que pose l'extraction, comme c'est le cas dans cet extrait de notre journal de terrain :

> *Le Référent Vidéo discute avec son responsable (le chef de la cellule Sécurité Economique et Financière) et lui demande l'autorisation d'aller récupérer les disques durs de la rame n°XYZ à P, suite à l'agression d'un contrôleur. En effet, pour lui – et d'après d'autres cas qu'il cite – il y a de grandes chances que le contrôleur porte plainte et qu'il y ait une réquisition judiciaire. Son responsable lui rétorque que cela va lui prendre beaucoup de temps, si la rame stationne à P. Le Référent Vidéo explique que, justement, c'est bien parce que la rame est à P, et que son retour n'est programmé que dans 4 jours. Pour éviter que les bandes ne soient « écrasées », le Référent Vidéo préfère aller retirer les bandes sur place. Le responsable donne son accord.*
> *(Visite commentée 2, Zone Sûreté)*

La spatialité du réseau rend par ailleurs complexe le travail d'extraction. En effet, comme l'illustre l'extrait précédent, il faut physiquement se rendre dans le train pour y récupérer les disques durs. Or, pour toute la Zone Sûreté en question, il n'y a qu'un seul Référent Vidéo, basé à Lyon et chargé de ces extractions. Les trains ne passent pas par Lyon tous les jours, tant s'en faut. Il faut alors se rendre sur place, souvent en voiture parce qu'il n'y a pas toujours assez de trains ou que ces derniers mettent trop de temps. En outre, l'extraction ne peut se faire dans une gare en stationnement, mais doit être effectuée dans une zone de garage, essentiellement pour des questions de sécurité :

> *Les enregistreurs sont sous un siège. Donc 1), c'est pas pratique, 2) le jour où on retire un disque et on demande aux gens [les voyageurs éventuellement présents dans le train] de se pousser, pour peu qu'il y ait un lascar dans l'équipe, voilà, le lendemain ils vont essayer de tout péter dans les trains, donc on peut pas se permettre ça*[17]
> *(Référent Vidéo, Zone Sûreté)*

Pour pallier ce problème géographique, certains agents de la Suge (de patrouille) ont été formés à l'extraction et disposent de mallettes avec des bandes d'échange et peuvent donc à l'occasion effectuer l'opération. Un système de sauvegarde automatique vient également pallier cette limite des 72h d'enregistrement maximum. Il s'agit d'une alarme qui déclenche un enregistrement sur une partie protégée du disque. Les quelques minutes de vidéo en amont et en aval du déclenchement sont ainsi préservées de l'écrasement. Cet enregistrement protégé peut être déclenché de trois façons : lorsque le conducteur du train appuie sur le bouton dédié ; lorsque le signal d'alarme (celui d'arrêt d'urgence) est actionné ; lorsque la caméra est masquée. Ce système d'enregistrement automatique semble cependant peu effectif, notamment parce que ses conditions d'activation sont dans la pratique peu atteignables, au dire du Référent Vidéo :

> *Traiter une réquisition sur alarme, ça reste exceptionnel. Parce que les gens qui se font agresser vont pas tirer le signal d'alarme, le mécano qui est pas au courant va pas appuyer, le contrôleur a pas forcément vu non plus*
> *(Référent Vidéo, Zone Sûreté)*

[17] Plus précisément, c'est parce que le temps où un train est inaccessible aux voyageurs est trop court pour effectuer la manœuvre de récupération que les gares de voyageurs sont évitées.

Il faut également noter que tous les trains ne sont pas équipés de vidéosurveillance. C'est pour cela qu'une des tâches du PCNS est de réaliser une recension des trains et gares équipés. Cela permet notamment d'accélérer la réalisation éventuelle d'extraction. Les agents du PCNS, en cas d'agression par exemple, peuvent vérifier, grâce au numéro de train, le type de rame et ainsi savoir si elle est équipée ou non en vidéosurveillance. Il faut donc d'abord reconstituer l'identité train-agressé / train-vidéosurveillé avant de procéder à l'extraction. De son côté, le Référent Vidéo rencontré a constitué une cartographie des gares équipées et des durées de conservation (dans les gares, certains vieux systèmes vont au-delà des 72h). Ceci lui permet de connaître le temps dont il dispose pour récupérer les bandes.

Ces tâches d'extraction, de recension et de vérification de l'état des caméras de vidéosurveillance démontrent que l'on est très loin d'une automaticité des systèmes de surveillance. Comme tout autre système technique, il nécessite un travail de veille, de maintenance qui ne se limite pas à la réparation une fois une panne décelée.

Les contraintes matérielles et cognitives que nous avons présentées dans les deux premières parties de l'article limitent fortement l'utilisation de la vidéosurveillance et son efficience. Les détections et enquêtes à partir d'images de vidéosurveillance en sont forcément réduites. Nos résultats sont ainsi congruents avec des travaux menés sur d'autres terrains qui mettent en avant la faible contribution des images de vidéosurveillance au travail d'investigation policière (Mucchielli, 2016a, 2016b)[18], ou son ambiguë intérêt probatoire dans le champ judiciaire (Lemaire, 2017, 2019)[19].

Pour obvier à ce manque d'efficience, industriels et chercheurs tentent de mettre au point des caméras dites « intelligentes ». Nous montrerons dans notre dernière partie que même dans ces dispositifs où l'automatisation de la détection est forte, le travail d'enquête des acteurs ne disparaît pas, mais se déplace dans la stabilisation de règles algorithmiques.

3 Vers des caméras "intelligentes" ?

Nous avons pu rencontrer trois chercheurs en informatique de l'INRIA (Institut national de recherche en informatique et en automatique), ayant travaillé dans le projet ANR VIDEO-ID, en collaboration avec l'industriel Thales et avec pour utilisateurs finaux des opérateurs de transports, dont la SNCF[20]. L'un des objectifs du projet de recherche est la confection d'un

[18] Pour donner un exemple précis : sur la vidéosurveillance municipale marseillaise, Mucchielli a montré que le nombre de cas où les images ont été utiles sur l'ensemble des faits de voies publiques était d'environ 1,5 %.

[19] Au-delà des difficultés que pose la qualité des images afin de les transformer en preuve dans un procès, Lemaire a démontré que cette fabrique de la « vidéo-preuve se réalise à travers une série d'opérations sociales de sélection déterminées par des intérêts inégaux à s'investir et à l'investir » (Lemaire, 2017, p.23)

[20] Le projet a été financé dans le cadre de l'édition 2007 du programme « Concepts, Systèmes et Outils pour la Sécurité Globale » de l'Agence Nationale de la Recherche. Le projet regroupait 3 laboratoires informatiques (Inria, Eurecom et Telecom), deux laboratoires de sciences humaines (Telecom et le CREDOF de Paris 10) et un centre de recherche industrielle de Thales. Le comité de pilotage était également composé de représentants des

algorithme d'analyse des images des bandes de vidéosurveillance afin de détecter des « situations anormales », de reconnaître des personnes (par identification de leur visage) et de les suivre dans une foule – les caméras étant situées dans des gares ferroviaires ou des stations de métro. Pour ce faire, nous allons voir que les informaticiens doivent effectuer une abstraction et une formalisation d'une partie du travail des opérateurs de vidéosurveillance. Dans ce processus, la définition des situations ou comportements « anormaux » va subir plusieurs transformations : si les informaticiens font valoir le critère de la légalité comme principe définitionnel *a priori* (3.1), le travail concret de confection de l'algorithme leur fera adopter une définition procédurale de la normalité (3.2). Ce faisant, cette définition entérine une vision particulière des normes sociales et vient se heurter aux besoins concrets des utilisateurs, forçant alors les informaticiens à complexifier leur modèle (3.3).

3.1. Le projet VIDEO-ID : une première définition juridique des situations « anormales »

Le principe retenu dans le projet est que l'algorithme détecte seul des situations ou les personnes recherchées et génère une alarme que l'opérateur doit traiter. Avoir pu saisir les caméras « intelligentes » et leur algorithme à l'état de projet de recherche permet de voir que dans la confection même de l'automaticité des alarmes (soit l'élaboration des règles de l'algorithme), c'est un travail d'enquête sur la définition du normal et de l'anormal qui se joue. Comme l'exprime clairement la présentation du projet :

> « L'identification d'un comportement dit « anormal », « étrange » ou « bizarre » suppose déjà une interprétation et un jugement sur ce qu'il convient socialement de considérer comme tel. L'automatisation solidifie et « anonymise » ce jugement dans le dispositif matériel » (Fedorczak et Bremond, 2007, p. 2)

Dans un projet de recherche, ce travail d'interprétation et de jugement devient descriptible. Il peut s'interpréter comme ce travail d'enquête au sens de Dewey, résultat d'un compromis entre les chercheurs informaticiens (porte-paroles des possibilités techniques de l'identification des personnes et de la reconnaissance des comportements par vidéo) et les besoins des utilisateurs finaux qui sont ici la SNCF et la RATP. Dans nos entretiens avec les chercheurs de l'INRIA ces questions sont soldées par des réponses légales et pratiques. À la question de savoir ce qui est « normal » ou « anormal », l'un des chercheurs pose la légalité du comportement comme définition de sa normalité.

> *le graffiti c'est pas normal parce que ça détériore les biens de la SNCF et après ça a un coût. C'est basé sur des délits légaux…je ne vois pas ce… [rires gênés]*
> *(Chercheur 1, INRIA)*

La normalité, définie par le droit, est considérée comme une chose allant de soi, sans prendre en compte les déterminants sociaux posés dans la présentation du projet de recherche.

prescripteurs (ministère de l'Intérieur) et des utilisateurs finaux : SNCF, RATP et l'Union Internationale des Chemins de fer. Le projet ANR en question était terminé au moment des entretiens, mais les trois chercheurs rencontrés étaient tous encore investis dans des recherches en continuité avec VIDEO-ID.

Puisqu'il s'agit uniquement de détecter les comportements illégaux, les questions de la détermination sociale des illégalismes est évacuée, conduisant ainsi à reproduire l'ordre établi : « *It's just to help people who need help* » (Chercheur 2, INRIA)[21].
En outre, l'opérateur humain apparaît toujours comme le juge en dernier ressort. C'est lui qui décide s'il faut intervenir ou non, après la détection du logiciel. Les possibilités techniques ne semblent pas encore permettre de véritables prédictions, ce qui règle pour les chercheurs la question de potentielle interpellation *a priori*.

> *Le logiciel fait juste des propositions. L'opérateur peut toujours aller à la pêche, aller regarder lui quelle est la caméra qu'il trouve la plus intéressante, parce qu'en fonction de l'heure et du lieu, il sait que là il a y des choses intéressantes (Chercheur 1, INRIA)*

La définition du « normal » semble ainsi réglée par sa détermination juridique : il s'agit de détecter des comportements illégaux. Pourtant, une exploration plus détaillée du projet montre que ce n'est pas aussi simple. Les acteurs impliqués ont dû s'entendre sur « ce qui est intéressant », et ont ainsi dû procéder à une enquête abstraite (mais sur des bandes de vidéosurveillance réelles), avant de fixer des règles algorithmiques.

3.2. Le déplacement du travail d'enquête : déterminer "ce qui est intéressant"

Le travail d'enquête (décider si un élément ou un comportement détecté est véritablement suspect et mérite de prévenir les forces d'intervention), jusqu'alors réalisé par les opérateurs de vidéosurveillance, doit être déplacé dans la fixation de règles procédurales. On rentre ici dans le travail des informaticiens qui doivent transformer les intérêts des utilisateurs d'un algorithme (ici, détecter des éléments ou comportements suspects) dans « l'ordre technique qui est le leur » (Cardon, 2018, p. 67)[22]. Comme l'explique Cardon, un algorithme n'a pas accès au sens des informations qu'ils calculent et n'en a pas de compréhension symbolique. Le travail des programmateurs consiste à essayer de trouver les meilleures approximations mathématiques d'un principe (ici classificatoire : suspect/non suspect), qui sera par la suite interprété par les utilisateurs comme symbolique (ceci est suspect / ceci n'est pas suspect). « Cette manière de trouver une approximation de la substance à travers une procédure constitue souvent la force des "meilleurs" algorithmes » (Cardon, 2018, p. 67-68). Comment les informaticiens rencontrés ont-ils cherché à tendre vers cette approximation procédurale ?

Une première difficulté pour eux fut de faire le point avec les utilisateurs finaux sur leurs besoins en termes de détection. Ces besoins sont apparus difficilement modélisables pour les informaticiens. Certaines situations semblent simples, par exemple :

> *Ils nous disent : « on est intéressé par ce scénario-là », bon bah c'est typiquement « bagage abandonné », « bagarre » (Chercheur 3, INRIA)*

Cette simplicité n'est cependant pas caractéristique de la majorité des cas :

[21] « Il s'agit juste d'aider les gens qui en ont besoin ». L'un des trois chercheurs n'était pas francophone.
[22] Sur le fonctionnement procédural et non substantiel du raisonnement algorithmique, et l'opacité intrinsèque qu'il peut générer, voir Burrell, 2016.

Ils ne savent pas vraiment ce qu'ils cherchent, euh…ils savent vaguement, mais précisément nous donner des exemples pour qu'on puisse développer des algo dessus ça reste quelque chose de compliqué
(Chercheur 3, INRIA)

L'une des solutions adoptées est alors de partir des possibilités techniques afin d'obvier à ce « défaut » des utilisateurs. Les chercheurs ont mis en place une technique « d'apprentissage non supervisé » où un logiciel apprend à détecter des schémas récurrents dans les images de vidéosurveillance qu'on lui fournit. Après avoir constitué une base de métadonnées vidéo (en ayant sélectionné les extraits de bandes vidéo où il y a des éléments à analyser : présence d'individus, de groupes ou de situations jugées intéressantes par les utilisateurs « *par exemple : les chutes, les attaques, les graffitis »*), le logiciel tente « d'apprendre » :

Il va essayer d'extraire l'information récurrente : quels sont les patterns de comportements récurrents ? [Par exemple :] les gens rentrent par cette zone, ils passent sur cette zone-là, ils s'arrêtent trois secondes ici pour repartir par là. Donc ça on va trouver que c'est un pattern récurrent et ça nous définit les activités fréquentes. Et à partir de ces activités fréquentes, toutes les autres c'est des activités non fréquentes, voire rares
(Chercheur 1, INRIA)

À partir de ces « patterns », l'objectif est de réaliser un « *modèle d'activité, qu'après on va réinsérer dans le système pour pouvoir le reconnaître en ligne et générer une alerte directement* » (Chercheur 1, INRIA). Nous nous trouvons ici dans ce qu'Antoinette Rouvroy et Thomas Berns ont appelé « la gouvernementalité algorithmique », soit un « type de rationalité (a)normative ou (a)politique reposant sur la récolte, l'agrégation et l'analyse automatisée de données en quantité massive de manière à modéliser, anticiper et affecter par avance des comportements possibles » (Rouvroy et Berns, 2013, p. 173). L'idéal de l'algorithme en cours d'élaboration est bien de rapporter le comportement d'un individu aux « patterns » préalablement « appris », dans un objectif d'anticipation. Si le logiciel reconnaît une séquence comportementale codée comme suspecte ou anormale (aboutissant par exemple à une rixe entre voyageurs), il génèrera une alerte (signifiant : « attention, cette personne adopte un comportement pouvant mener à une rixe »). Nous ne sommes toutefois pas exactement dans un algorithme de prédiction par des traces, comme ceux utilisés dans le web[23]. Si traces il y a (l'enregistrement des comportements), elles ne sont pas personnalisées, elles ne sont pas attachées à tel ou tel individu. Le logiciel ne peut se baser que sur le comportement en temps réel d'un individu, il n'intègre pas *ses* déplacements et comportements passés, mais les compare à ceux de tous les individus passés par la gare avant lui, pour détecter des « patterns » qu'il a construits.

Des limites techniques semblent toutefois encore importantes dans cette phase de détection. Comme l'explique ce chercheur : « *les algo vision sont assez fragiles, ils ne sont pas très fiables et puis la connaissance elle est difficilement mobilisable par les utilisateurs* (Chercheur

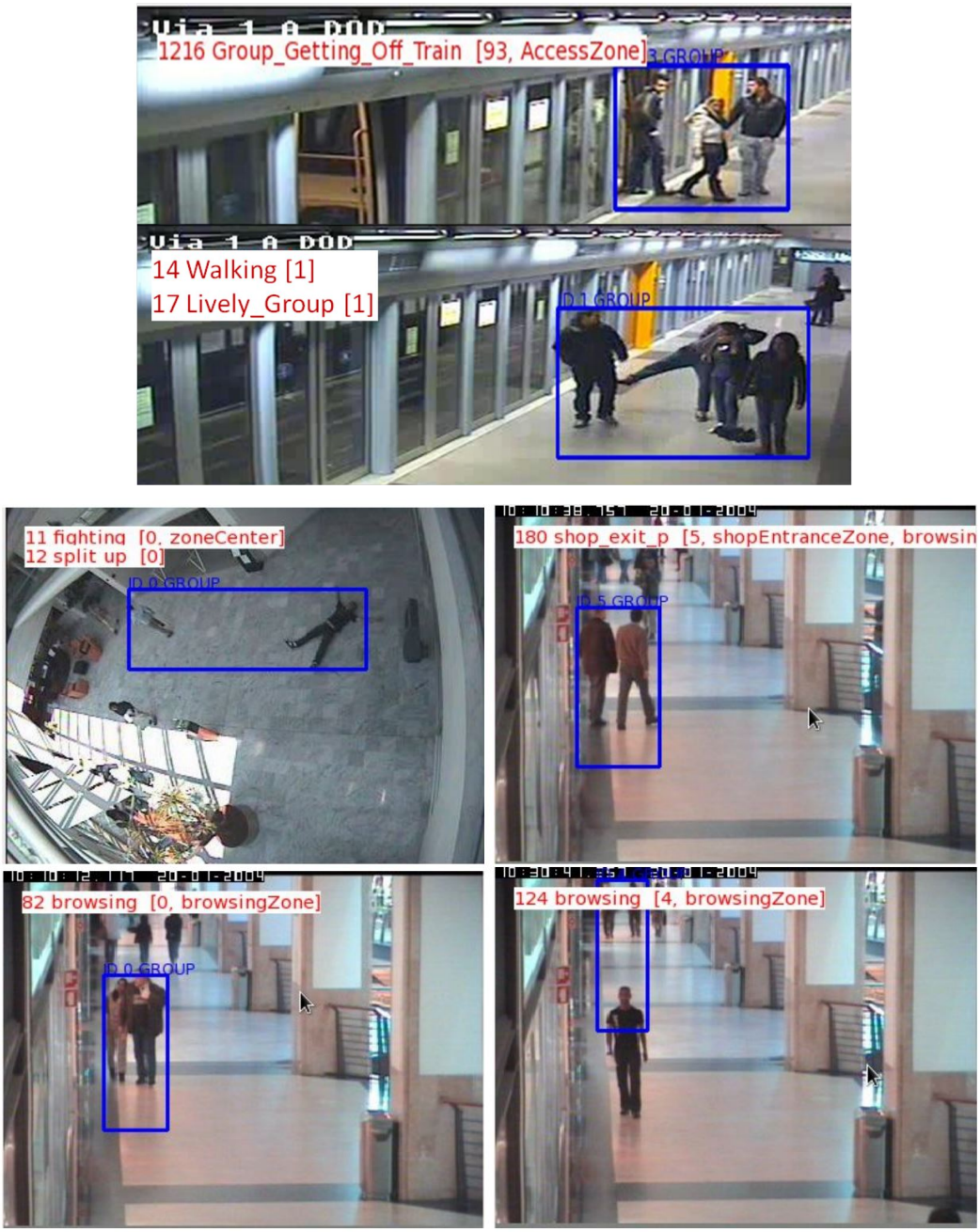
[23] Il s'agit de la quatrième famille d'algorithmes identifiée par Dominique Cardon dans son analyse du web (Cardon, 2015). Après les algorithmes calculant la popularité (le nombre de clics), l'autorité (le nombre de liens hypertextes) et la réputation (le nombre de « likes »), la dernière famille concerne la prédiction par les traces que les internautes laissent sur le web.

3, INRIA). Même sur un scénario simple tel qu'un bagage abandonné, la qualité de la détection semble très contingente :

> *Ça dépend de l'environnement. C'est-à-dire que si la caméra est de suffisamment bonne qualité, que l'éclairage dessus est de suffisamment bonne qualité, et qu'il n'y a pas trop de monde, on détecte. Après s'il y a un des paramètres en moins et bien c'est plus difficile. Ça dépend de la qualité de la scène.*
> *(Chercheur 1, INRIA)*

Concernant la détection des personnes, les chercheurs ont tout de même réussi à produire un logiciel encadrant d'un liseré de couleur un individu ou groupe d'individus et attribuant des étiquettes à leur comportement : « calme », « stressé », « court », « statique », « actif », « erratique », etc. Plus la personne ou le groupe est isolé, et plus la détection et la qualification du comportement sont aisées. Voici des exemples tirés d'une communication des chercheurs rencontrés (figure 1). .

Figure 1 - Identification et qualification des personnes sur vidéosurveillance



Source : Zaidenberg, Boulay et Bremond, 2012, p. 2-7-8

3.3. Distinguer le normal de l'anormal : la fréquence comme juge de paix

Après cette phase de détection, tout comme les opérateurs humains, le logiciel doit être capable de déterminer s'il doit générer une alarme ou non. Nous avons vu qu'une première définition de la normalité donnée par un chercheur était basée sur la légalité du comportement.

Pourtant, la technique d'apprentissage non supervisée induit une autre définition du normal, reposant sur la fréquence des comportements. Est anormal ce qui est non récurrent statistiquement.

> *Moi je dirais que tout ce qui est rare normalement dans tous les domaines, si tu veux même en sécurité, toutes les activités rares c'est sur ces activités-là qu'on doit faire attention […]. Ben un comportement rare c'est un comportement improbable statistiquement*
> *(Chercheur 3, INRIA)*

> *i'm working on groups behavior in videosurveillance. The idea is to try to recognize some dangerous behaviors in subways. Until now, we don't have so many data of very dangerous behaviors. So we are just dealing with …maybe abnormal : when they are gathered for too many times, when they are too agitated, changing their way suddenly. So strange behaviors […]. Maybe it's not really dangerous but it's not common*[24].
> *(Chercheur 2, INRIA)*

Dans ce dernier extrait, on voit un glissement : le « dangereux » devient l' « anormal » ou l'étrange », lui-même « ce qui n'est pas commun ». Ces chercheurs se rapprochent alors de la pensée des ingénieurs, telle que décrite par Vatin, c'est-à-dire une pensée tournée vers l'action : « Pour s'orienter dans le monde, il faut des points d'appui ; il faut bien les prendre où on les trouve » (Vatin, Caillé et Favereau, 2010, p. 89). Pris dans une logique d'action (il faut produire un logiciel de détection), les chercheurs travaillent avec ce qu'ils ont et ce qu'ils peuvent effectivement mesurer : la fréquence des modèles comportementaux détectés.
Dans une des communications des chercheurs de l'INRIA et de Thales, on retrouve cette équivalence entre le peu fréquent et l'anormal.

> « Dans une troisième étape, nous utilisons un algorithme d'apprentissage non supervisé pour regrouper les individus principalement selon leurs comportements et découvrons à la fois des événements *fréquents/normaux* et des événements *inhabituels/anormaux* » (Patino et al., 2011, p. 3, nous soulignons)[25]

On retrouve en effet cette idée de la fréquence comme principe de la normalité dans les premiers usages de la statistique chez les précurseurs et fondateurs des sciences sociales (Desrosières, 1988). Cependant, il y a ici une confusion subreptice entre le normal-statisticien et le normal-juge. Loin de nous l'idée que les statistiques seraient par essence dépourvues de jugements moraux. Mais l'utilisation ici de « normal », en plus de son sens descriptif (ce qui est peu fréquent) renvoie à un jugement social et moral de ce qui est répréhensible[26]. La

[24] « Je travaille sur les comportements de groupe à partir de vidéosurveillance. L'idée est d'essayer de reconnaître des comportements dangereux. Jusqu'à présent, nous n'avons pas tant de données sur des comportements très dangereux. Donc on traite avec…[des comportements] peut-être anormaux : quand ils sont rassemblés trop de temps, quand ils sont trop agités, quand ils changent soudainement de chemin. Donc des comportements étranges […]. Peut-être que ce n'est pas vraiment dangereux, mais ce n'est pas commun ».

[25] Notre traduction de l'original : « In a third step we employ a high-level clustering algorithm to group mobiles according principally to their behaviours and discover both, frequent/normal behaviours and unusual/abnormal events ».

[26] On s'éloigne ici alors de la gouvernementalité algorithmique de Rouvroy et Berns, dans le sens où il n'y a pas d'émancipation de la norme statistique au profit de normativités immanentes propres à chaque individu. On est toujours dans ce que Foucault a appelé la « normalisation » où la norme est issue du repérage du normal et de

commande sociale des chercheurs est bien de confectionner un algorithme détectant les comportements dangereux, du moins qui posent un problème aux gestionnaires des gares et des métros. Le dangereux est devenu l'anormal. Il y a là une superposition des sens statistique et moral de « normal ». Ceci n'est pas seulement un enjeu théorique, il est aussi pratique. En effet, ces chercheurs et leurs commanditaires se sont bien rendu compte que tout ce qui était rare n'était pas forcément anormal (dans le sens de dangereux ou du moins de « devant être détecté »). Pour Rouvroy et Berns, l'une des forces de la gouvernementalité algorithmique est justement que les « ratés » (quand l'algorithme se trompe) servent à améliorer l'algorithme : « le but est de ne rater aucun vrai positif, quel que soit le taux de faux positif » (Rouvroy et Berns, 2013, p. 174). Cependant, dans notre cas, l'objectif pratique de l'algorithme fait des faux positifs un véritable problème. Un trop grand nombre de faux positifs, générant une alarme, pourrait aboutir à une paralysie de l'action. En outre, il est des situations que l'on voudrait détecter qui apparaissent, au vu des caméras, tout à fait normales (au sens de fréquentes).

Les chercheurs ont alors tenté de complexifier le modèle. Premièrement, en déterminant des combinaisons de comportements normaux (fréquents) qui, selon des circonstances, peuvent devenir anormaux (dangereux).

> *Il y a des activités qui sont très difficiles à détecter en elles-mêmes […] Donc par d'autres moyens on peut détecter des comportements anormaux, qui sont pas forcément l'acte malveillant en lui-même.*
> *Exemple : ça va être très difficile de détecter que la personne prend un objet et le met dans sa poche. Ils sont spécialistes, c'est pas du tout facile. Par contre, c'est relativement facile de détecter qu'une personne reste un petit peu longtemps sur une allée, tourne, regarde à droite, à gauche que personne le regarde et après, il se dépêche de sortir à la caisse*
> *(Chercheur 1, INRIA)*

Il faut ainsi combiner la détection d'une action avec d'autres caractéristiques pour pouvoir lui donner un sens.

> *Par exemple le fait que des gens restent longtemps euh…dans une station, c'est pas vraiment anormal. Mais on s'aperçoit quand même, qu'en fonction de là d'où ils viennent, de là où ils repartent, et de comment ils restent, on peut arriver à détecter des gens qu'ont des…des comportements on va dire…anormaux*
> *(Chercheur 3, INRIA)*

On trouve encore ici une limite de la gouvernementalité algorithmique. Celle-ci, par le *machine learning*, rend « directement possible la production d'hypothèse à partir des données elles-mêmes […]. Les normes semblent émerger directement du réel lui-même » (Rouvroy et Berns, 2013, p. 170). Comme le dit Cardon, les promoteurs du *big data* affirment pouvoir « chercher des corrélations sans se préoccuper d'avoir un modèle qui leur donne une explication. Les données massives et les mathématiques permettraient de faire l'économie des sciences de l'homme » (Cardon, 2015, p. 51). Là résiderait leur caractère a-normatif de « machines a-signifiantes, en abandonnant de la sorte l'ambition de donner de la signification aux

l'anormal statistique (Foucault, 2004). Ce qui diffère pourtant de la statistique traditionnelle, c'est l'application *en temps réel* : c'est la comparaison à l'instant t du comportement d'un individu à la norme « apprise » par l'algorithme et une définition évolutive de ce normal, au fur et à mesure de l'intégration et de l'analyse de nouvelles bandes de vidéosurveillance.

événements » (Rouvroy et Berns, 2013, p. 174). On voit bien que l'objectif pratique assigné à l'algorithme de détection oblige les acteurs à (re)trouver du sens, et à redonner de la signification à ce qui est détecté. D'où la complexification du calcul en croisant comportements et circonstances.

Pour parachever la recherche de signification, les chercheurs ont fait subir des tests pratiques à l'algorithme. Toute une série d'événements détectés par l'algorithme a été montrée à des utilisateurs finaux pour qu'ils confirment ou non l'intérêt de telles détections. En effet, « les données ne parlent qu'en fonction des questionnements et des intérêts de ceux qui les interrogent » (Cardon, 2015, p. 57). La détection algorithmique ne découvre rien, elle ne fait pas ici naître d'hypothèses sur ce qui serait suspect ou dangereux. La technique d'apprentissage a beau ne pas être « supervisée » dans un premier temps (l'algorithme regroupe des comportements qui se ressemblent, à partir de son analyse d'une quantité importante de bandes de vidéosurveillance, sans définition préalable de ce qu'il fallait détecter), il faut bien revenir à l'intérêt pratique des acteurs pour qu'ils soient d'une quelconque utilité.

Un autre test a consisté à comparer la détection algorithmique avec la détection humaine. Les chercheurs ont mobilisé des étudiants en sciences cognitives et leur ont fait noter ce qu'ils trouvaient d'anormal ou de suspect dans une série d'images de vidéosurveillance. Ils ont ensuite comparé avec ce que l'algorithme avait trouvé d'anormal (on retrouve dans ce test la confusion entre les deux sens de « normal »). Et ensuite, « *on essaie de faire aussi bien que les étudiants* ». Les résultats furent peu concluants :

> *Parce que les problèmes c'est qu'il y a énormément de données et qu'il y a peu de gens pour vraiment les regarder, et dire ce qui se passe dedans pour pouvoir comparer avec ce que fait l'ordinateur*
> *(Chercheur 1, INRIA)*

La comparaison n'a pas pu porter sur suffisamment de cas pour être jugée significative.

On voit bien ici comment les chercheurs tentent de trouver des règles qui reproduisent le travail d'enquête que les opérateurs de vidéosurveillance « non intelligente » effectuent lors d'une détection. L'approximation du travail des opérateurs vidéo par la règle procédurale basée sur la fréquence n'a cependant pas été jugée suffisamment robuste ; d'où les tentatives d'optimisation par des tests pratiques.

Notons pour conclure cette partie qu'à notre connaissance, cette technologie n'est pas utilisée actuellement dans le système de vidéosurveillance de la SNCF[27]. Des projets de recherche

[27] La seule utilisation de « caméra intelligente » dont nous avons connaissance concerne le terminal transmanche de l'Eurostar en gare du Nord à Paris. L' « intelligence » déployée est ici moindre que celle de l'algorithme du projet ANR dans la mesure où il s'agit seulement de détecter des intrusions. Il n'y a pas d'analyse comportementale, il s'agit simplement de déterminer une « zone interdite » et d'y détecter la présence d'individus.

similaires ont cependant été développés, avec des applications plus ciblées comme la prévention des suicides[28].

Conclusion

Le cas de la vidéosurveillance – exploitée ici par ses applications dans le transport ferroviaire en France – montre l'importance de ne pas céder au fonctionnalisme et déterminisme technologique dans l'étude des processus de surveillance. L'analyse de la technologie en usage (ici le travail de visionnage et d'extraction des images) démontre que les caméras ne surveillent pas par elles-mêmes. De même, l'attention à la fragilité technique permet de sortir d'une vision qui postule que les processus de surveillance remplissent automatiquement les buts pour lesquels ils sont mis en place. Le travail des opérateurs est primordial à l'effectivité de la surveillance, qui apparaît assez faible au vu des limites cognitives et matérielles mises en avant.

D'une manière plus générale, ces résultats permettent de défendre la thèse selon laquelle il n'y a pas d'automaticité des effets des pratiques de surveillance (comme peuvent souvent le penser une partie des « surveillance studies ») et ce quel que soit leur degré d'automatisation. La raison principale en est que dès la détection d'un élément considéré comme problématique, s'ouvre un nécessaire travail d'enquête (au sens de Dewey) pour décider si l'élément est bien suspect et quelles suites sont à lui donner. L'automatisation (qui dans le cas d'espèce peut aller jusqu'aux caméras dites "intelligentes") ne supprime pas ce travail d'enquête des opérateurs de vidéosurveillance, mais le déplace (que ce soit dans les règles organisationnelles ou dans les règles algorithmiques). Ce déplacement doit aussi conduire à un déplacement du travail de recherche et de la critique.

C'est ce que nous avons essayé de réaliser dans la troisième partie de ce papier, consacrée au projet de recherche VIDEO-ID. L'existence d'un tel projet de recherche est pour nous intéressante à plusieurs titres.
Tout d'abord, elle souligne que l'un des problèmes pratiques rencontrés par les acteurs chargés de surveiller est bien le trop-plein de données et de leur capacité à les traiter. D'où le développement de projet visant à automatiser encore plus le traitement des images de vidéosurveillance. Ensuite, elle montre concrètement ce que recouvre l'automaticité des détections. La stabilisation des règles algorithmiques peut s'apparenter au travail d'enquête que fournissent les acteurs. Tout l'enjeu pour les chercheurs de l'INRIA est de pouvoir reproduire l'enquête des opérateurs de vidéosurveillance : ce que repère l'algorithme est-il intéressant pour eux ? Ce faisant, dans notre cas, l'utilisation de la technique d'apprentissage non supervisé pour repérer des comportements suspects entérine un jugement social : ce qui n'est pas commun est anormal, ce qui est peu fréquent est suspect. Le caractère « intelligent » réside alors en réalité dans un certain rapport au temps, dans la recherche d'un gain de

[28] C'est le cas du projet de recherche RESTRAIL (Reduction of Suicides and Trespasses on RAILway property, 2011-2014), financé par le 7e programme-cadre pour la recherche de l'Union Européenne et portée par l'UIC. L'un des objectifs fut la détermination de comportements à risque annonciateur d'un acte suicidaire et potentiellement détectable par vidéosurveillance. Source : RESTRAIL [http://restrail.eu/, consulté le 26/05/2017].

réactivité. Avec ces « caméras intelligentes », il s'agit d'avoir une connaissance des besoins (ici des interventions de la Suge, des forces de l'ordre, etc.) en étant le plus proche possible de l'état réel des interactions sociales. Ainsi, une surveillance « intelligente » ne serait pas une surveillance qui voit tout, mais une surveillance qui voit ce qu'il faut voir au bon moment. Il s'agit de la recherche d'une connaissance immédiate, à t0. Le rêve du prédictif réside dans un dépassement de ce t0 : non pas intervenir au plus vite après la détection d'un événement, mais intervenir avant, dès ses prémisses. En ce sens, il faut fortement relativiser le sens d' « intelligent ». Dans le *machine learning*, il ne s'agit pas de reproduire le raisonnement humain, mais d'une « statistique des contextes » (Cardon, 2015, p. 60) : à partir d'une masse considérable de données, des modèles comportementaux sont définis et appliqués à des individus ou groupes d'individus dont le contexte a été jugé similaire à tel ou tel modèle[29]. Ainsi, si ces algorithmes sont prédictifs, c'est « parce qu'ils font constamment l'hypothèse que notre futur sera une reproduction de notre passé » (p.70). D'où un fort effet de clôture du réel sur lui-même avec ces algorithmes (Rouvroy et Berns, 2013). C'est particulièrement le cas avec ces « caméras intelligentes » qui font l'hypothèse que ce qui est peu fréquent statistiquement est suspect, renforçant ainsi les normes dominantes.

Ces résultats appellent à davantage d'études empiriques sur l'automatisation des pratiques de surveillance. Il nous semble que la mise en algorithme d'un nombre de plus en plus important de fonctions pourrait être saisie pour débattre collectivement des principes que les programmateurs tentent de transformer en règles procédurales. Ces procéduralisations informatiques sont autant de points critiques à saisir, dans la mesure où ils peuvent objectiver et rendre visibles des biais, discriminations ou inégalités jusque-là plus difficilement identifiables. Des recherches sont ainsi à mener sur le potentiel réflexif (et donc critique) de l'algorithmisation de la société.

[29] « L'enjeu n'est plus d'apprendre aux machines une grande théorie appliquée, mais de multiplier les petites théories en demandant à beaucoup de données contextuelles de sélectionner la ou les meilleures d'entre elles » (Cardon, 2015, p. 61).

Bibliographie

AKRICH M., 1987, « Comment décrire les objets techniques ? », *Techniques & Culture. Revue semestrielle d'anthropologie des techniques*, 9, p. 49-64.

AYRES I., 2007, *Super crunchers : Why thinking-by-numbers is the new way to be smart*, New York, Bantam Books.

BANNISTER J., FYFE N., KEARNS A., 1998, « Closed circuit television and the city », dans NORRIS C., ARMSTRONG G., MORAN J. (dirs.), *Surveillance, Closed Circuit Television, and Social Control*, Aldershot, Ashgate.

BETIN C., MARTINAIS E., RENARD M.-C., 2003, « Sécurité, vidéosurveillance et construction de la déviance : l'exemple du centre-ville de Lyon », *Déviance et Société*, *27*, 1, p. 3-24.

BIDET A., VATIN F., 2016, « Travailler, c'est produire. Activité, valeur, et ordre social », dans DUJARIER M.-A., GAUDART C., GILLET A., LENEL P. (dirs.), *L'activité en théories. Regards croisés sur le travail*, Toulouse, Octares Editions, p. 13-33.

BIDET A., 2006, « Le travail et sa sociologie au prisme de l'activité », dans BIDET A., BORZEIX A., PILLON T., ROT G., VATIN F. (dirs.), *Sociologie du travail et activité*, Toulouse, Octares Editions, p. 5-23.

BONNET F., 2012, « Contrôler des populations par l'espace ? Prévention situationnelle et vidéosurveillance dans les gares et les centres commerciaux », *Politix*, 97, p. 25-46.

BOUSSARD V., LORIOL M., CAROLY S., 2006, « Catégorisation des usagers et rhétorique professionnelle », *Sociologie du Travail*, *48*, 2, p. 209-225.

BURRELL J., 2016, « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, *3*, 1, p. 2053951715622512.

CAMERON A., KOLODINSKI E., MAY H., WILLIAMS N., 2008, *Measuring the effects of video surveillance on crime in Los Angeles*, Los Angeles, Californian Research Bureau, University of Southern California: School of Policy Planning and Development.

CARDON D., 2015, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Le Seuil.

CARDON D., 2018, « Le pouvoir des algorithmes », *Pouvoirs*, 1, p. 63–73.

CASTAGNINO F., 2017, *Les chemins de faire de la surveillance: une sociologie des dispositifs de sécurité et de sûreté ferroviaires en France*, PhD Thesis, Université Paris-Est.

CASTAGNINO F., 2018, « Critique des surveillances studies. Éléments pour une sociologie de la surveillance », *Déviance et Société*, *42*, 1, p. 9–40.

CASTRO D., 2016, « Data detractors are wrong: The rise of algorithms is a cause for hope and optimism », *Center for Data Innovation.*

COLEMAN R., SIM J., 2000, « 'You'll never walk alone': CCTV surveillance, order and neo-liberal rule in Liverpool city centre1 », *The British journal of sociology*, *51*, 4, p. 623–639.

DENIS J., PONTILLE D., 2015, « Material Ordering and the Care of Things », *Science, Technology & Human Values*, *40*, 3, p. 338-367.

DESROSIERES A., 1988, « Masses, individus, moyennes : la statistique sociale au XIXe siècle », *Hermès, La Revue*, *2*, 2, p. 41-66.

DEWEY J., 1967, *Logique. La théorie de l'enquête*, Paris, PUF.

DUBBELD L., 2005, « The role of technology in shaping CCTV surveillance practices », *Information, Communication & Society*, *8*, 1, p. 84-100.

FEDORCZAK C., BREMOND F., 2007, « VIDEO-ID. Identification de comportements et de personnes par la vidéo », *Projet ANR-CSOSG 2007*.
FOUCAULT M., 2004, *Sécurité, territoire, population. Cours au Collège de France (1977-1978)*, Paris, Gallimard/Seuil.
GILL M., SPRIGGS A., ALLEN J., ARGOMANIZ J., BRYAN J., JESSIMAN P., KARA D., KILWORTH J., LITTLE R., SWAIN D., WAPLES S., 2005, *The impact of CCTV: fourteen case studies*, London, Home Office.
GRAHAM S., THRIFT N., 2007, « Out of Order Understanding Repair and Maintenance », *Theory, Culture & Society*, *24*, 3, p. 1-25.
HELTEN F., FISCHER B., 2004, « Reactive attention: Video surveillance in Berlin shopping malls », *Surveillance & Society*, *2*, 2/3, p. 323-345.
HENKE C.R., 1999, « The mechanics of workplace order: toward a sociology of repair », *Berkeley Journal of Sociology*, *44*, p. 55–81.
KLAUSER F., NOVEMBER V., RUEGG J., 2006, « Surveillance et vigilance dans la sécurité routière: L'exemple de l'autoroute de contournement à Genève », dans ROUX J. (dir.), *Etre vigilant. L'opérativité discrète de la société du risque*, Saint-Etienne, Publications de l'Université de Saint-Etienne, p. 33-45.
KROENER I., NEYLAND D., 2012, « New technologies, security and surveillance », dans BALL K.S., HAGGERTY K.D., LYON D. (dirs.), *The Routledge Handbook of Surveillance Studies*, New York, Routledge, p. 141-148.
LE GOFF T., 2013, « Dans les «coulisses» du métier d'opérateur de vidéosurveillance », *Criminologie*, *46*, 2, p. 91–108.
LEMAIRE É., 2017, « La fabrique de la vidéo-preuve », *Champ pénal/Penal field*, Vol. XIV.
LEMAIRE E., 2019, *L'oeil sécuritaire: mythes et réalités de la vidéosurveillance*, La Découverte.
LEVY R., 1987, *Du suspect au coupable. Le travail de police judiciaire*, Genève, Méridiens Klincksiek.
MCCAHILL M., 1998, « Beyond Foucault: towards a contemporary theory of surveillance », dans NORRIS C., ARMSTRONG G., MORAN J. (dirs.), *Surveillance, closed circuit television and social control.*, Aldershot, Ashgate, p. 41–65.
MILGRAM A., 2013, « Why smart statistics are the key to fighting crime »,.
MUCCHIELLI L., 2016a, « À quoi sert la vidéosurveillance de l'espace public ? », *Deviance et Societe*, *Vol. 40*, 1, p. 25-50.
MUCCHIELLI L., 2016b, « De la vidéosurveillance à la vidéoverbalisation : usages réels et fantasmés d'une technologie moderne », *Archives de politique criminelle*, *n° 38*, 1, p. 249-264.
NORRIS C., ARMSTRONG G., 1999, *The maximum surveillance society: The rise of CCTV*, Oxford, Berg Publishers.
ORLIKOWSKI W.J., 1992, « The duality of technology: Rethinking the concept of technology in organizations », *Organization science*, *3*, 3, p. 398–427.
PATINO L., BENHADDA H., NEFZI N., BOULAY B., BREMOND F., THONNAT M., 2011, « Abnormal behavior detection in video protection systems », p. 12.
PEROUMAL F., 2009, « Le monde précaire et illégitime des agents de sécurité », *Actes de la recherche en sciences sociales*, *n° 175*, 5, p. 4-4.
ROUVROY A., BERNS T., 2013, « Gouvernementalité algorithmique et perspectives d'émancipation », *Réseaux*, *177*, 1, p. 163-196.

SMITH G.J.D., 2004, « Behind the Screens: Examining Constructions of Deviance and Informal Practices among CCTV Control Room Operators in the UK », *Surveillance & Society*, *2*, 2/3, p. 376-395.
SMITH G.J.D., 2012, « Surveillance work(ers) », dans LYON D., BALL K.S., HAGGERTY K. (dirs.), *Routledge Handbook of Surveillance Studies*, New York, Routledge, p. 107-115.
VATIN F., CAILLE A., FAVEREAU O., 2010, « Réflexions croisées sur la mesure et l'incertitude », *Revue du MAUSS*, 1, p. 83–109.
WALBY K., 2005, « How closed-circuit television surveillance organizes the social : an institutionnal ethnography », *Canadian Journal of Sociology*, 30, p. 189-214.
WELSH B.C., FARRINGTON D.P., 2003, « Effects of closed-circuit television on crime », *The Annals of the American Academy of Political and Social Science*, *587*, 1, p. 110–135.
WELSH B., FARRINGTON D.P., 2007, « Closed-circuit television surveillance and crime prevention: A systematic review », Report prepared for the Swedish National Council for Crime Prevention, Stockholm, Sweedish Council for Crime Prevention.
ZAIDENBERG S., BOULAY B., BREMOND F., 2012, « A generic framework for video understanding applied to group behavior recognition », *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, p. 136-142.