

École Nationale Supérieure Louis-Lumière



## **Mémoire de Master 2**

**D'une surveillance par l'image à une image surveillée :**  
Comment la photographie peut-elle éveiller nos consciences ?

Victor Poullain

Spécialité Photographie – Promotion 2016

Sous la direction de Mme Claire Bras

Enseignante à l'École Nationale Supérieure Louis-Lumière

Membres du jury

Mme Claire Bras, enseignante à l'ENS Louis-Lumière

Mme Véronique Figini, enseignante-chercheure à l'ENS Louis-Lumière

M. Franck Maindon, enseignant à l'ENS Louis-Lumière

M. Pascal Martin, maître de conférences HDR

École Nationale Supérieure Louis-Lumière



## **Mémoire de Master 2**

**D'une surveillance par l'image à une image surveillée :**  
Comment la photographie peut-elle éveiller nos consciences ?

Victor Poullain

Spécialité Photographie – Promotion 2016

Sous la direction de Mme Claire Bras

Enseignante à l'École Nationale Supérieure Louis-Lumière

Membres du jury

Mme Claire Bras, enseignante à l'ENS Louis-Lumière

Mme Véronique Figini, enseignante-chercheure à l'ENS Louis-Lumière

M. Franck Maindon, enseignant à l'ENS Louis-Lumière

M. Pascal Martin, maître de conférences HDR

## **Remerciements**

Merci à Claire Bras, pour la direction de ce mémoire et la pertinence de ses conseils.

Merci aux membres du jury, Véronique Figini, Franck Maindon et Pascal Martin, pour leur lecture attentive.

Merci à Florent Fajole, pour son appui bibliographique.

Merci également à Mathilde Galis et Liza Bodt, pour leur aide dans la réalisation de ma partie pratique.

Merci à Mohamed Bourouissa, pour ses discussions éclairées.

Et surtout merci à mes parents, ma grand-mère, ma sœur et mes amis, pour leur soutien sans faille.

## Résumé

La surveillance est omniprésente dans notre société, que ce soit par les caméras de vidéosurveillance visibles au quotidien ou les images qui en sont issues. Malgré leur apparente banalité, ces images singulières ont un réel impact sur le public à travers leur utilisation par les médias ou la fiction, dans leurs réponses aux fantasmes et aux angoisses.

Après une étude historique de la généralisation des systèmes de surveillance dans nos sociétés, ce mémoire a pour but de comprendre les fonctions des caméras et images de surveillance, la façon dont elles sont relayées dans les médias et leur appropriation dans un cadre artistique. D'abord support de surveillance, l'image est peu à peu devenue surveillée à travers notre utilisation de l'Internet et des réseaux sociaux. C'est à partir de ces différents points que cette recherche interroge les fonctions de la photographie vis-à-vis des questions sociétales posées par ces images.

**Mots-clés : surveillance ; image ; pouvoir ; panoptique ; caméra ; appropriation ; réseaux sociaux ; exhibition**

## **Abstract**

Through videosurveillance cameras or images we see everyday, surveillance is everywhere. Despite their commonness, these singular images have a real impact on the public in the media or in fiction, because of their answers to fantasies and fears.

After an historic study of the generalization of the surveillance in our societies, this report tries to understand the surveillance cameras and images fonctions through their relay in the media and their appropriation in artistic projects. At first surveillance support, the image became gradually watched through our use of the Internet and the social networks. In these points, this work interrogates the fonctions of photography towards these societal questions.

**Keywords : surveillance ; picture ; power ; panoptic ; camera ; appropriation ; social networks ; exhibition**

## Table des matières

Remerciements.....	3
Résumé.....	4
Abstract.....	5
Table des matières .....	6
Introduction.....	8
<b>I – Évolutions du modèle panoptique.....</b>	<b>10</b>
<b>1 – La théorie du panoptique .....</b>	<b>10</b>
a – L’origine de la théorie.....	10
b – Le panoptique de Michel Foucault.....	15
c – La société disciplinaire .....	17
<b>2 – Transposition à l’ère du numérique .....</b>	<b>19</b>
a – Caractérisation de la vidéosurveillance .....	19
b – Du panoptique par les caméras ... ..	21
c – ... à la surveillance individualisée .....	24
<b>3 – Résistance .....</b>	<b>27</b>
a – Remises en cause .....	27
b – Questions éthiques .....	29
c – Activismes .....	31
<b>II – La surveillance par l’image.....</b>	<b>36</b>
<b>1 – Usages politiques de l’image.....</b>	<b>36</b>
a – L’identité et la surveillance .....	36
b – L’image témoin.....	37
c – Le pouvoir de l’image .....	40
<b>2 – Esthétique de la surveillance.....</b>	<b>41</b>
a – Le point de vue .....	41
b – Les rendus d’image.....	45
c – La sérialité.....	48
<b>3 – Appropriation de la surveillance.....</b>	<b>52</b>
a – Utilisations de systèmes de surveillance dans l’art.....	52
b – Conception de dispositifs de prise de vue omnivoyants ....	56
c – Appropriations d’images de surveillance.....	60

<b>III – L’image surveillée.....</b>	<b>66</b>
<b>1 – L’acceptation de la surveillance .....</b>	<b>66</b>
a – L’obsession sécuritaire.....	66
b – La caméra totem .....	67
<b>2 – La société exhibitionniste .....</b>	<b>68</b>
a – L’exhibition personnelle.....	68
b – L’image surveillée .....	72
c – Une évolution du modèle panoptique.....	75
<b>3 – Quel rôle peut jouer la photographie ? .....</b>	<b>76</b>
a – Lien étroit entre la surveillance et la photographie .....	76
b – Entre distanciation et fantasmes.....	80
c – Partie pratique de mémoire .....	83
<b>Conclusion.....</b>	<b>85</b>
<b>Bibliographie.....</b>	<b>87</b>
<b>Table des illustrations.....</b>	<b>92</b>
<b>Table des annexes .....</b>	<b>94</b>
<b>Annexes.....</b>	<b>95</b>

## Introduction

Depuis la découverte, la maîtrise et la démocratisation de l'électricité dans nos sociétés du XVIIIème au XXème siècle, le monde a considérablement évolué et d'énormes progrès ont vu le jour notamment en matière de communication.

Tout va maintenant très vite. Notre monde est lancé dans cette course effrénée à l'instantané, déjà quasiment atteinte par la circulation des données numériques.

Nous nous appuyons quotidiennement sur les techniques informatiques pour répondre à nos besoins jusqu'à ce que nos vies soient peu à peu devenues régies par des ordinateurs. En effet, la plupart des informations nous concernant circulent au sein de multiples réseaux et sont enregistrées dans d'immenses banques de données. Les informations ainsi collectées à l'échelle planétaire peuvent s'avérer très intéressantes pour des raisons commerciales, politiques ou autres et susciter de nombreuses convoitises à l'insu des citoyens. Ceci peut constituer une forme de surveillance représentative du système qui s'est installée.

Il en existe ainsi plusieurs types, dont fait partie celle par l'image. Afin de répondre à des besoins sécuritaires, militaires ou simplement informatifs, une quantité astronomique de dispositifs de surveillance a été installée depuis quelques décennies aussi bien dans des espaces publics que privés. Ils fonctionnent comme des miradors : disposés en hauteur, ils enregistrent de manière passive ou active tout ce qui se passe sur un espace donné, tout en se faisant oublier. Tous ces systèmes étant reliés entre eux, ils constituent de véritables réseaux pouvant visuellement couvrir la totalité de certains espaces.

Maintenant plus que jamais, la surveillance suscite sous toutes ses formes de nombreux questionnements. Certains citoyens, activistes et artistes ont compris les enjeux liés aux libertés mises en péril par ces différents types de surveillance et se sont mobilisés de différentes manières afin que la population prenne conscience des dangers représentés pour leur vie privée.

Il paraît alors intéressant de se poser la question suivante : comment la photographie peut-elle éveiller nos consciences vis-à-vis de la surveillance dans notre société ?

Afin de comprendre l'origine des moyens de surveillance, la première partie de ce mémoire sera centrée sur la théorie du panoptique décrite par le philosophe français Michel Foucault. Cette dernière se révèle prophétique dans l'organisation de dispositifs adaptés au contrôle des individus dans l'espace public et privé et la répartition des points d'observation du pouvoir dans nos sociétés.

Dans un second temps, nous tenterons de comprendre comment fonctionnent ces images issues de vidéosurveillances, en analysant leurs caractéristiques esthétiques et leurs utilisations dans le cadre de projets artistiques.

Enfin, nous verrons en quoi avec l'arrivée de l'Internet, de nouvelles formes de surveillance sont apparues, passant ainsi d'une logique de surveillance par l'image à une image surveillée, pour déboucher sur la problématique du rôle de la photographie vis-à-vis de celle-ci.

# I – Évolutions du modèle panoptique

## 1 – La théorie du panoptique

Avant tout, il est important de comprendre l'origine de la théorie du panoptique afin de bien comprendre et cerner le raisonnement que Michel Foucault a développé dans le chapitre intitulé du même nom, tiré de son ouvrage *Surveiller et punir* paru en 1975. Cela nous permettra de comprendre plus tard comment ce modèle initialement architectural s'est peu à peu déplacé vers les dispositifs récents de captation d'image intégrés à l'environnement.

### a – L'origine de la théorie

D'abord, un strict quadrillage spatial : fermeture, bien entendu, de la ville et du « terroir », interdiction d'en sortir sous peine de la vie, mise à mort de tous les animaux errants ; découpage de la ville en quartiers distincts où on établit le pouvoir d'un intendant. Chaque rue est placée sous l'autorité d'un syndic ; il la surveille ; s'il la quittait, il serait puni de mort.<sup>1</sup>

Michel Foucault s'est d'abord appuyé sur des archives militaires de Vincennes datant de la fin du XVII<sup>ème</sup> siècle, qui à l'instar de nombreuses autres de la même époque en France, décrivaient les mesures à appliquer lorsque la peste se déclarait dans une ville.

Ces archives témoignent d'une organisation spatiale du pouvoir bien particulière, que le philosophe qualifie de « modèle compact du dispositif disciplinaire »<sup>2</sup>. En effet, les quartiers concernés par l'épidémie, habituellement ouverts à la circulation, étaient clos et découpés en secteurs. Chaque secteur était placé sous la garde de responsables. Des postes de surveillance étaient placés dans des endroits stratégiques telles que les portes de la ville ainsi qu'à chaque bout de rue afin de contrôler les entrées et sorties de chacun. Cette organisation permettait ainsi une gestion plus facile des zones contaminées : la surveillance était plus précise car établie dans un périmètre

---

1 *Archives militaires de Vincennes, A 1 516 91 sc. Pièce*, cité par Michel Foucault in *Surveiller et Punir*, Paris, Gallimard, 1975, p. 228

2 *Ibid.*, p. 230

de taille réduite et délimité ; les habitants (dont certains d'entre eux sont malades) ne pouvaient pas se déplacer librement afin d'éviter les contagions ; les informations étaient centralisées et circulaient selon un protocole établi.

De plus, les règles à suivre dans cette situation de crise étaient expliquées point par point et concernaient l'ensemble des acteurs, aussi bien les habitants que les employés pour l'organisation et la décontamination des quartiers. Chaque individu était analysé, classé en fonction de son état et tous ses déplacements étaient répertoriés. Le pouvoir s'appliquait donc à différentes échelles, du groupe d'habitants du quartier jusqu'à l'individu même.

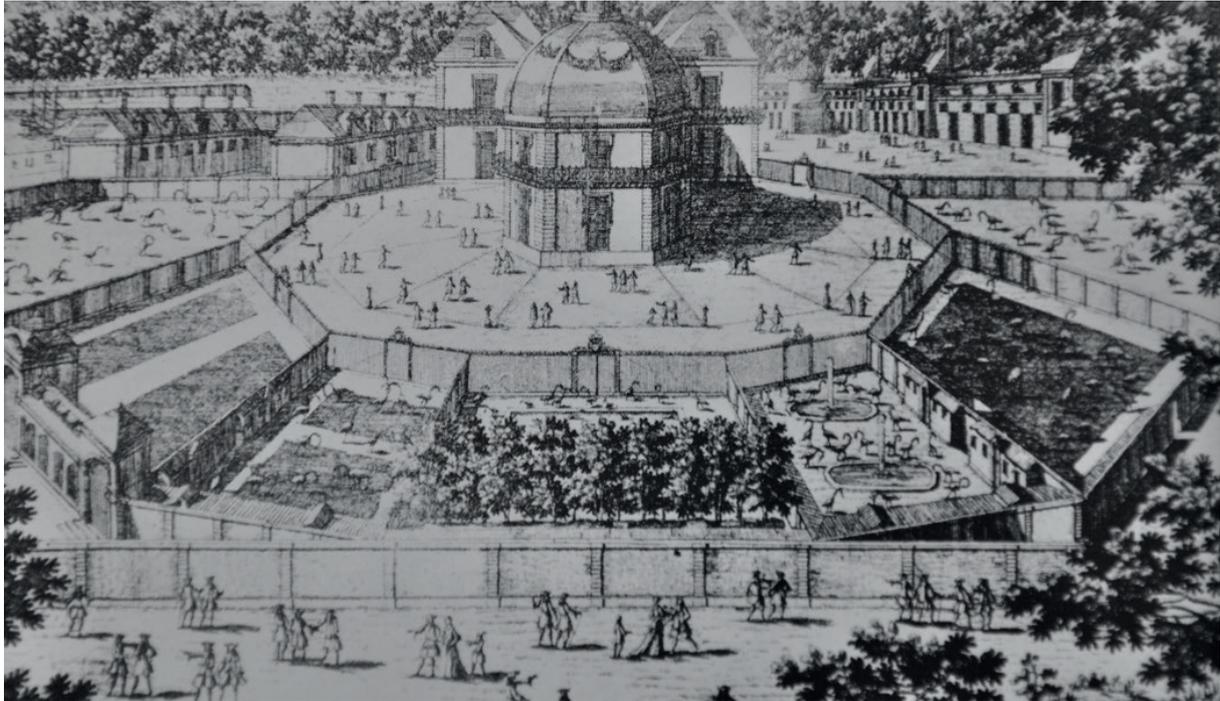
Le pouvoir s'exerçant sur ces espaces était établi à plusieurs degrés de hiérarchie et en devenait ainsi omniprésent. Cette organisation spatiale permettait une application optimale de ce dernier.

En se penchant sur ce type de répartition spatiale du pouvoir, Michel Foucault a ainsi pu déceler les origines de la réflexion du *Panopticon* du philosophe anglais Jeremy Bentham dans les plans de la Ménagerie Royale de Versailles conçue par l'architecte français Louis Le Vau en 1663, pour ensuite construire sa théorie du panoptique.

(...) au centre un pavillon octogonal qui, au premier étage, ne comportait qu'une seule pièce, le salon du roi ; tous les côtés s'ouvraient par de larges fenêtres, sur sept cages (le huitième côté est réservé à l'entrée) où étaient enfermées différentes espèces d'animaux.<sup>3</sup>

---

3 *Ibid.*, p. 237



**Figure 1**, AVELINE Claude, *Vue et perspective de la ménagerie de Versailles*, gravure, 1638-1715

L'architecture du pavillon central en fait le parfait point de vue sur l'ensemble de la ménagerie royale de Versailles en trois points.

D'abord, son emplacement central le situait à équidistance des différents enclos de la ménagerie construits autour. De plus, sa forme octogonale, s'approchant géométriquement d'un cercle, permettait d'y intégrer de nombreuses et larges ouvertures et ainsi obtenir un très large champ de vision sur l'ensemble de l'espace entourant le pavillon. Enfin, le salon du roi était établi au premier étage de ce pavillon, donc en hauteur, permettant ainsi de surplomber l'ensemble des enclos. Tous ces éléments permettaient donc d'observer toutes les espèces animales depuis le même point de vue.

Cette tour étant réservée au roi, son architecture ainsi que sa disposition, représentent parfaitement le pouvoir que ce dernier exerce sur son royaume : avec fermeté, hauteur et distance mais omniprésent.

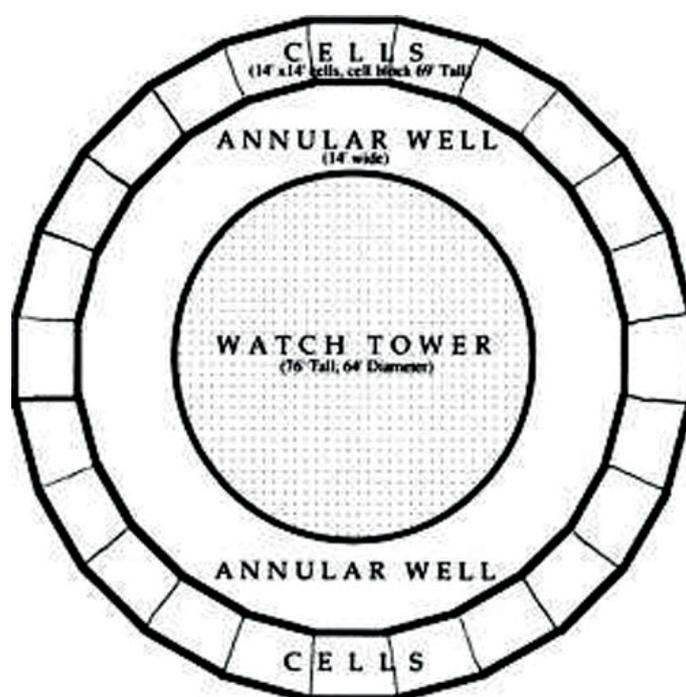
Il est ainsi intéressant de noter que l'enjeu architectural de la ménagerie Royale de Versailles de Le Vau s'oppose parfaitement aux besoins rencontrés à l'Antiquité. En

effet, à cette époque que Michel Foucault associe à « une civilisation du spectacle »<sup>4</sup>, de nombreux édifices servant à accueillir des fidèles dans les temples ou encore des spectateurs dans les théâtres, avaient dû être construits dans l'optique de rendre un sujet ou un groupe de sujets visibles par une multitude d'individus. Ces architectures peuvent être considérées comme représentatives de la cohésion des individus qui ne constituaient qu'un seul et même corps lors de ces différentes réunions.

À l'inverse, dans l'architecture de la ménagerie et dans le *Panopticon* de Bentham que nous développerons plus loin, l'objectif est au contraire de rendre visible une multitude de sujets par un seul individu, ou du moins depuis un même point de vue. Cette nouvelle problématique est représentative pour Michel Foucault de la société moderne qui vise à compartimenter et séparer les individus de l'État.

Les points clés de l'architecture de la ménagerie de Le Vau ont été réemployés et développés dans la conception du *Panopticon* de Jeremy Bentham.

Ce modèle d'architecture carcéral, élaboré par ce philosophe à la fin du XVIII<sup>ème</sup> siècle, est constitué d'une tour centrale, dans laquelle sont installés les surveillants, entourée par des cellules individuelles dans lesquelles sont enfermés les prisonniers.



**Figure 2**, BENTHAM Jeremy, *Plan schématique du Panopticon*, 1791

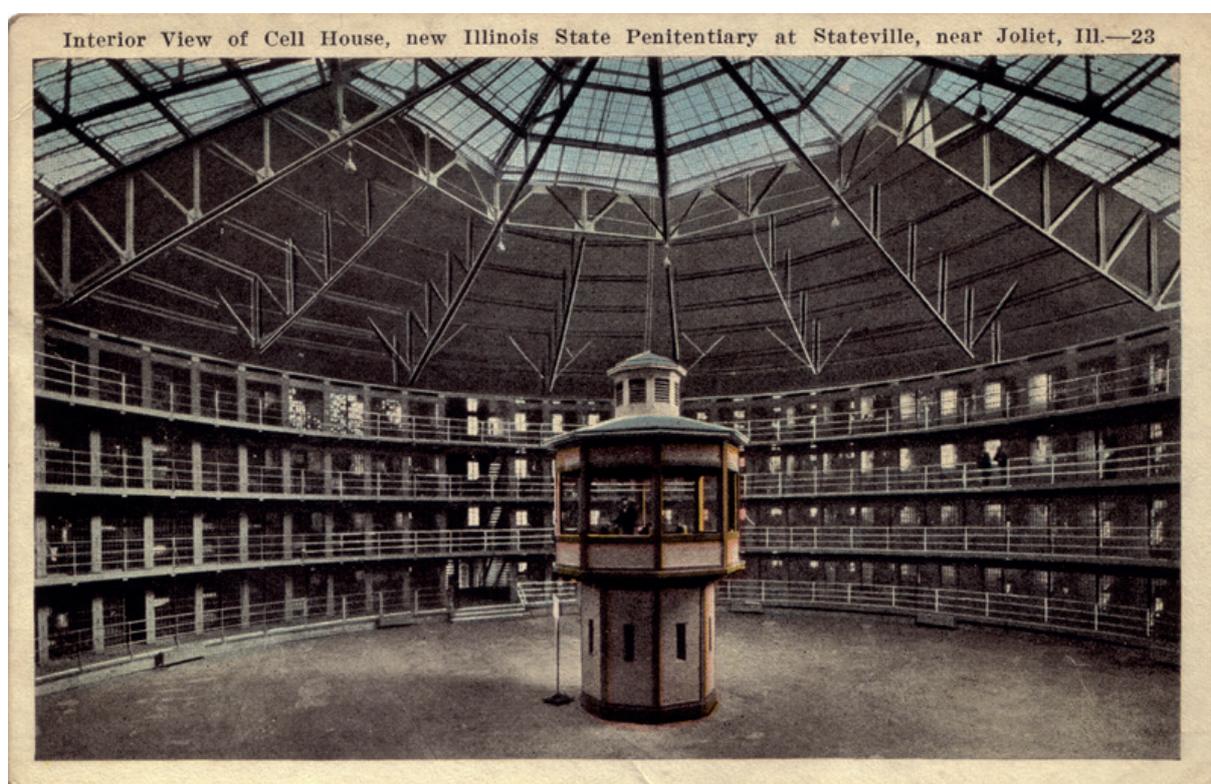
4 *Ibid.*, p. 252

Plusieurs spécificités architecturales ont été imaginées afin de faciliter et tenter de rendre infaillible la surveillance des détenus.

En effet, les cellules ne communiquent pas entre elles afin d'éviter tout contact entre prisonniers. Dans chacune d'entre elles se trouvent deux ouvertures : l'une donnant sur l'intérieur de la prison, l'autre sur l'extérieur. Ainsi, la lumière du jour entrant par l'extérieur, le prisonnier est très rapidement visible et observable en contre-jour. Il n'existe pas d'angle mort pour la surveillance, l'espace est lumineux et aménagé de manière à empêcher le détenu de se cacher ou préparer quelque chose à l'insu des surveillants.

Les ouvertures de la tour centrale, permettant aux surveillants d'observer les détenus, sont occultées par des persiennes. De plus, à l'aide d'un système de chicanes et de cloisons, l'intérieur de la tour est agencé de telle sorte qu'aucune lumière ne peut y rentrer par l'intérieur et ainsi trahir la présence ou l'absence du surveillant.

Grâce à ces divers aménagements, Bentham a ainsi réussi à rendre le pouvoir du surveillant visible mais invérifiable du point de vue du surveillé.



**Figure 3**, *Carte postale de l'intérieur du centre pénitencier de Stateville, photographie, 1920*

## **b – Le panoptique de Michel Foucault**

À partir du modèle d'architecture carcéral élaboré par Bentham, Michel Foucault a théorisé le concept de panoptique développé dans le chapitre éponyme tiré de son ouvrage *Surveiller et Punir*.

Le panoptique est une machine à dissocier le couple voir-être vu : dans l'anneau périphérique, on est totalement vu, sans jamais voir ; dans la tour centrale, on voit tout, sans être jamais vu.<sup>5</sup>

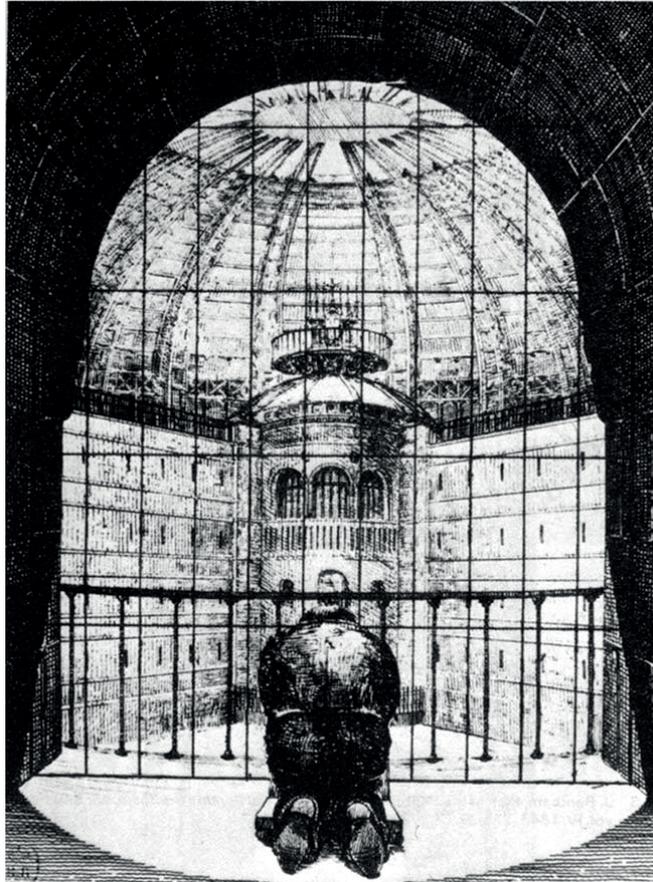
D'après les différentes spécificités de cette architecture énoncées précédemment, tout est fait pour que le détenu soit observable en permanence par son surveillant, sans qu'il puisse savoir s'il est réellement surveillé. Pour Foucault, c'est justement en ce point que réside toute la force du modèle panoptique. En effet, le système architectural induit chez le détenu, qui ne peut savoir si on l'observe, un sentiment de visibilité permanente auprès du surveillant. Le prisonnier est alors implicitement obligé de respecter les règles en permanence. Grâce à ce dispositif, le pouvoir est donc assuré automatiquement : que les surveillants soient présents ou non dans la tour centrale, le détenu se sentira observé et par conséquent dominé par le pouvoir. Le simple fait de se savoir possiblement observé l'empêchera de désobéir et de causer des troubles.

De plus, la présence centrale ainsi que la forme de cette tour de surveillance érigée devant tous ces détenus renvoie à la symbolique totémique, accroissant encore son pouvoir de soumission.

La gravure suivante, représentant un détenu dans sa cellule priant face à la tour centrale de surveillance, symbolise son assujettissement face au pouvoir exercé par le système panoptique dont il est prisonnier.

---

5 *Ibid.*, p. 235



**Figure 4**, HAROU-ROMAIN N., *Projet de pénitencier*, gravure, 1840

Michel Foucault va même jusqu'à considérer le panoptique comme une forme idéale de gestion du pouvoir. En effet, de par sa centralité, il permet d'effectuer des économies à différents niveaux :

- Matériel : Depuis la même tour centrale, il est possible de surveiller un grand nombre de détenus.
- Personnel : La surveillance s'effectuant depuis le même point de vue, un même surveillant peut observer plus de détenus. Par conséquent, pour le même nombre de détenus, moins de surveillants sont nécessaires.
- Temporel : L'ensemble de l'organisation architecturale facilitant la surveillance des détenus, le temps de réaction et d'intervention en cas de problème est extrêmement court.

De plus, le mécanisme panoptique assure une efficacité remarquable dans l'application du pouvoir. À travers l'omniprésence de la tour centrale et du sentiment de visibilité suscité chez le détenu, le pouvoir est assuré de manière préventive et continue.

C'est pourquoi il ne voit pas son utilisation réservée au monde carcéral. En qualifiant l'architecture de *modèle*, il sous-entend déjà que son principe a été repris dans différents types de cadres constituant nos sociétés devenues peu à peu disciplinaires. De plus, cette analyse des caractéristiques du dispositif architectural révèle deux points déterminant dans la surveillance par l'image que Foucault ne connaît pas encore : la disposition et la fonction symbolique de l'œil de la caméra.

### **c – La société disciplinaire**

C'est au temps moderne, à l'influence toujours croissante de l'État, à son intervention de jour en jour plus profonde dans tous les détails et toutes les relations de la vie sociale, qu'il était réservé d'en augmenter et d'en perfectionner les garanties, en utilisant et en dirigeant vers ce grand but la construction et la distribution d'édifices destinés à surveiller en même temps une grande multitude d'hommes.<sup>6</sup>

Le modèle panoptique est en effet transposable dans de nombreuses structures et applicable à chaque fois qu'une tâche ou des règles doivent être imposées à un groupe d'individus. Il trouve ainsi sa place dans les lieux de travail, écoles, institutions, ... Ces lieux ne nécessitant évidemment pas les mêmes besoins que les prisons, seulement l'organisation du dispositif est reprise afin de répondre à des besoins disciplinaires. Qu'il s'agisse du maître d'école ou du contremaître au travail, leur position est souvent centrale et en hauteur par rapport aux élèves ou ouvriers généralement placés en contrebas et organisés circulairement ou en quadrillage afin de faciliter la circulation entre ces derniers. Cela permet aux exécutants du pouvoir d'adopter un point de vue global et en plongée sur l'ensemble de l'espace dont ils sont responsables.

Bien que d'abord exploité à des fins disciplinaires, le panoptique peut aussi avoir une utilité et une praticité dans l'observation et l'expérimentation. En effet, son fonctionnement a par exemple été transposé dans les hôpitaux afin de faciliter l'observation des malades. Il peut aussi se révéler utile dans des cadres expérimentaux afin d'observer en même temps le déroulement de plusieurs expériences sans influencer leur résultat.

---

<sup>6</sup> Nicolaus Heinrich Julius in *Leçons sur les prisons*, Paris, F.G. Levrault, 1831, p. 384-386, cité par Michel Foucault in *Surveiller et Punir*, *op. cit.*, p. 252

Une analogie contemporaine au panoptique est par exemple constatable avec les salles équipées de glaces sans tain. Cette vitre possédant un taux de réflexion très élevé sépare deux pièces contiguës, l'une étant éclairée, l'autre plongée dans l'obscurité. Ainsi, la glace sans tain fonctionne comme un miroir pour les personnes situées à l'intérieur de la pièce éclairée (à cause de l'important taux de réflexion et de la lumière), tandis qu'elle fonctionne comme une vitre transparente pour celles situées dans la pièce sombre. Elle permet ainsi aux personnes se trouvant dans la pièce sombre de voir sans être vues. Ces vitres sont par exemple utilisées dans les salles d'interrogatoires de commissariat de police afin d'observer discrètement les réactions des interrogés ou encore dans des commerces afin d'étudier les comportements de consommation des clients dans un but marketing.

La généralisation des espaces dits *open-space* est un autre exemple d'évolution du modèle panoptique dans le monde du travail. Cet aménagement dépourvu de cloisons a connu un grand essor aux États-Unis, puis en Europe dans les années 1980. Le gain de place, la facilité de communication entre collègues étaient les principaux arguments en faveur de cette organisation. L'espace étant ouvert, les employés y travaillant sont naturellement appelés à s'observer : la surveillance mutuelle implicitement induite s'est trouvée être intéressante pour les employeurs, produisant un effet de pression sur l'ensemble des travailleurs.

Le modèle panoptique s'est ainsi étendu aux différents cadres dans lesquels son application pouvait avoir un intérêt en vue de l'application du pouvoir. Les évolutions sociétales et techniques l'ont conduit à se diviser et se répartir dans l'espace afin d'étendre son champ d'action. La vidéosurveillance en est sûrement la forme actuelle la plus répandue et la plus aboutie.

## **2 – Transposition à l'ère du numérique**

### **a – Caractérisation de la vidéosurveillance**

Afin de comprendre les systèmes de vidéosurveillance et de mieux cerner leur champ d'action, nous commencerons par analyser leur fonctionnement à travers leurs principales caractéristiques que l'on peut distinguer en trois points : le point de vue, la répartition spatiale et l'image automatisée.

#### ***Le point de vue***

La première caractéristique de ces dispositifs de surveillance est le point de vue constitué par les éléments d'articulation que sont les caméras. L'objectif de ces caméras étant de permettre la surveillance d'un espace, leur disposition doit ainsi généralement permettre de couvrir le plus large angle de champ. C'est pourquoi ces caméras sont la plupart du temps installées en hauteur pour surplomber le lieu, orientées vers l'espace à surveiller (souvent en plongée vers le sol) et équipées d'un objectif grand angle (ou à focale variable si la caméra est pilotable) afin de filmer l'ensemble de l'espace sans être gênées par un quelconque obstacle. Le point de vue pris par ces caméras domine l'espace. Situées au dessus de l'environnement qu'elles permettent de surveiller, elles sont difficiles à atteindre. Ces caméras symbolisent un pouvoir s'exécutant impartialement, sans relâche, inatteignable pour le citoyen et contre lequel il ne peut que difficilement lutter.

Le fait que les caméras de vidéosurveillance soient systématiquement placées de la sorte produit une harmonisation des images en terme de distance au sujet, d'angle de champ, de perspectives, ... Grâce à ces différents éléments, une image de vidéosurveillance est facilement reconnaissable. Nous pouvons même parler d'imagerie de la surveillance, présente dans l'imaginaire collectif et en grande partie répandue par les productions audiovisuelles telles que les films ou les séries. Nous reviendrons sur ce point dans la suite de ce mémoire.

### ***La répartition spatiale***

La seconde caractéristique des systèmes de vidéosurveillance est leur répartition spatiale. En effet, l'efficacité d'une installation de vidéosurveillance est souvent liée à sa capacité à couvrir l'intégralité de l'espace à surveiller, ce qu'Anne-Marie Duguet définit comme « l'omnivoyance »<sup>7</sup>. Par conséquent, deux paramètres sont à prendre en compte dans la couverture d'un lieu : le nombre de caméras ainsi que leurs emplacements.

Après une étude des besoins de surveillance du lieu, les caméras de vidéosurveillance installées forment ainsi un réseau, structure reliant différents périphériques permettant ainsi d'échanger des informations. Il en existe de toutes les tailles, du réseau de vidéosurveillance domestique au réseau couvrant l'espace public de mégapoles.

En fonction de l'échelle à laquelle nous les observons, ces différents réseaux de caméras de surveillance peuvent ainsi se superposer jusqu'à former un seul et même réseau, à la fois vaste et très fin. Nous pouvons prendre pour simple exemple le quotidien d'un citoyen : que ce soient les caméras présentes dans l'espace public lorsqu'il marche dans la rue, celles du réseau de transport en commun lorsqu'il prend le métropolitain ou encore celles du supermarché lorsqu'il fait ses courses, les dispositifs de surveillance sont installés dans de nombreux lieux de vie ou de passage. Cette personne est ainsi potentiellement observable à de nombreuses reprises à travers cette succession de réseaux de vidéosurveillance dont il s'avère difficile d'échapper. C'est pourquoi nous pouvons parler d'omniprésence de ces systèmes de surveillance dans notre quotidien.

### ***L'image automatisée***

Enfin, la dernière caractéristique importante que je souhaite aborder est le fait que ces systèmes de vidéosurveillance produisent des images en continu. Bien que cette continuité soit relative au modèle de caméra ainsi qu'aux paramètres d'enregistrement pouvant dans certains cas être bridés à une fréquence d'enregistrement de l'ordre d'une image par seconde (voire plusieurs) ne permettant ainsi pas une couverture temporelle totale dans l'enregistrement de l'espace, la captation d'images est automatisée.

---

<sup>7</sup> Anne-Marie Duguet, « Dispositifs » in *Revue Communications* n°48, Paris, Seuil, 1988, p. 230

Il n'y a pas d'interruption de la saisie, le balayage est constant, avec ou sans mouvement. Tout, à tout moment, est susceptible d'être interprété comme un événement, peut en être le signe.<sup>8</sup>

En effet, dans la vidéosurveillance, contrairement à la photographie ou même à la vidéographie et cinéma dits « conventionnels », le déclenchement n'existe plus. La notion de déclenchement induit à la fois une mise en route ainsi qu'un arrêt de la captation d'images, que cette durée soit de l'ordre de la fraction de seconde pour figer un mouvement en photographie, de la minute pour filmer une scène au cinéma ou de l'heure pour enregistrer une émission à la télévision. Dans la vidéosurveillance, ce déclenchement se voit remplacé par une simple mise en route du système, l'arrêt de l'enregistrement ne fait pas partie de la démarche de captation.

De plus, la notion de déclenchement implique une volonté dans l'acte d'enregistrement de cette durée de la part de l'opérateur. Or une fois la caméra de vidéosurveillance ainsi que le réseau installés, la captation d'images est continue et automatique, aucun acte n'est nécessaire pour la déclencher. C'est précisément en ce point que réside tout l'intérêt de ces dispositifs : la surveillance des espaces est ainsi assurée de manière permanente.

## **b – Du panoptique par les caméras ...**

Les différentes caractéristiques des systèmes de vidéosurveillance que nous venons d'énoncer sont issues du modèle panoptique. En effet, la disposition en hauteur des caméras provient de la même recherche de point de vue « omniscient » que les surveillants situés dans la tour centrale de la prison dont il est impossible d'échapper. De plus, l'automatisation et la continuité de la captation d'images des caméras rappellent aussi l'architecture de la tour centrale empêchant aux détenus de connaître la présence ou non des surveillants, instaurant ainsi un pouvoir de surveillance automatique.

Bon nombre de caméras sont d'ailleurs non-opérationnelles et sont installées seulement pour faire acte de présence et ainsi dissuader les éventuels malfaiteurs à l'aide d'un panneau les annonçant. Ces systèmes factices reposent sur la visibilité

---

8 *Ibid.*

de l'objet caméra induisant une éventuelle présence. Il s'agit là d'une synthèse contemporaine de la théorie de Foucault : le sentiment de permanence et d'invisibilité du regard surveillant se retrouve canalisé dans la caméra de surveillance en tant qu'objet, qu'elle soit effective ou non.

Ces dispositifs de surveillance peuvent donc être considérés comme une transposition du système panoptique au travers des caméras. Cette mutation progressive du modèle a pu avoir lieu grâce aux évolutions techniques notamment liées au domaine de l'informatique, permettant ainsi une application du pouvoir encore plus efficace.

La grande avancée réside dans la centralisation des informations. En effet, la surveillance ne se fait plus physiquement sur le lieu : elle est déportée. Grâce aux réseaux informatiques, chaque caméra retransmet en direct son flux vidéo en un point qui les centralise. Le dispositif peut être considéré comme encore plus efficace que le panoptique, dans le sens où il permet de faire de plus grandes économies de personnels, puisqu'une même personne peut alors surveiller différents espaces éventuellement éloignés depuis le même endroit. Il s'agit là d'une dissémination du point de vue démultiplié.

Certaines caméras de vidéosurveillance sont même maintenant connectées au réseau de l'Internet, dont le flux est parfois même laissé en accès libre volontairement ou non. En faisant quelques recherches, il est facile de s'y connecter et ainsi visionner leur flux vidéo en temps réel à partir de son ordinateur, peu importe leur localisation dans le monde.

C'est d'ailleurs ce sur quoi portait mon projet de majeure de 3ème année en section Photographie à l'École Nationale Supérieure Louis-Lumière, présenté lors de l'exposition collective *Point final*, qui se tint du 5 au 7 décembre 2015 à Éléphant Paname (Paris). Intitulée *Panoptismes*, mon installation se présentait sous la forme totémique trônant au centre d'un espace circulaire, dans laquelle un moniteur vidéo diffusait des flux de caméras de surveillance récupérés depuis l'Internet et diffusés en temps réel. Le spectateur pouvait choisir la vue de la caméra en naviguant dans la mosaïque à l'aide d'une interface se trouvant sur la face de l'armoire. Après une certaine durée d'interaction avec l'installation, mon propre dispositif de surveillance

installé sur le lieu d'exposition se déclenchait et le retour vidéo venait alors s'intégrer dans la mosaïque. Le spectateur en train d'observer des lieux et des personnes en temps réel à travers le moniteur se voyait lui-même observé. Il lui était alors impossible d'échapper à son image, le programme restant bloqué sur son propre retour vidéo affiché en plein écran.

Ce projet questionne l'utilisation des dispositifs de vidéosurveillance dans nos sociétés. En effet, les caméras de surveillance visibles dans cette installation donnent à voir les lieux caractéristiques de nos sociétés : lieux de travail, de consommation, de divertissement, d'éducation, ... pour montrer l'omniprésence de la surveillance. Ma volonté était de permettre au spectateur d'entrer en interaction avec l'installation et de lui offrir ainsi la possibilité de prendre un point de vue d'observateur. Mais en y intégrant mon propre dispositif de surveillance, ce dernier s'aperçoit être alors lui-même observé. Il s'agit d'une mise en abîme du voyeur observé par un autre voyeur, mettant en place un jeu de pouvoir à plusieurs échelles.



**Figure 5**, POUILLAIN Victor, *Panoptismes*, capture écran, 2015

Michel Foucault nous apprend d'ailleurs que le créateur de l'architecture panoptique avait déjà imaginé un système de cette ampleur à son époque : « Bentham rêve d'en faire un réseau de dispositifs qui seraient partout et toujours en éveil, parcourant la société sans lacune ni interruption »<sup>9</sup>. Jeremy Bentham était donc un visionnaire, il avait déjà compris ce vers quoi son système tendrait à évoluer pour le rendre encore plus efficace, près de deux siècles avant que la technique le permette !

Mais si Bentham avait réussi à imaginer comment se développerait son système de surveillance aussi longtemps à l'avance, est-il possible de concevoir comment évoluera ce domaine à l'avenir ?

### **c – ... à la surveillance individualisée**

Jeremy Bentham avait déjà imaginé à son époque que le système panoptique évoluerait avec le temps et tendrait vers une division et une répartition de plus en plus fine du pouvoir. Comme nous l'avons vu dans la partie précédente, les différentes avancées techniques ont permis l'installation et la répartition de nombreuses caméras de surveillance dans les espaces publics et privés et cette tendance de la surveillance à se diviser est toujours d'actualité.

Cette division de la surveillance s'effectue d'abord sur le corps surveillant. En effet, afin d'assurer une surveillance encore plus efficace, les surveillants sont plus nombreux et surtout de natures plus diverses. En témoigne le concept de *Neighborhood Watch* (*Surveillance de voisinage*), en place dans la plupart des pays anglo-saxons, consistant en l'organisation d'habitants d'un quartier en groupes de surveillance dans le but de prévenir la délinquance. Tous les citoyens, membres de ces associations, deviennent alors surveillants dans le but de protéger leur prochain. La force de ce dispositif réside justement dans l'ambivalence du rôle de chacun : chaque voisin étant à la fois surveillant ou surveillé, une véritable cohésion se met en place et permet une surveillance discrète et quasi permanente.

---

9 Michel Foucault in *Surveiller et Punir*, op. cit., p. 243

À l'ère du numérique, un dérivé de ce système s'est mis en place sur l'Internet. Dans son article intitulé « Délateurs en pantoufles » publié dans *Le Monde diplomatique* de décembre 2015, Ignacio Ramonet décrit le fonctionnement de l'entreprise *Internet Eyes*, lancée en 2009 au Royaume-Uni, proposant aux internautes de devenir surveillants à leurs heures perdues.

Assis dans leur fauteuil, les adhérents observent en direct à travers l'œil des caméras. S'ils détectent un vol, une agression, un comportement suspect, ils cliquent sur un bouton d'alerte. L'image se fige alors et il leur est possible de zoomer pour vérifier. Ensuite, le gérant du magasin reçoit un message avec l'image retenue. S'il considère cette alerte utile, l'internaute délateur se voit créditer de trois points. S'il estime que l'alerte était justifiée même s'il n'y a pas eu finalement d'infraction, l'internaute reçoit un point. En revanche, si le commerçant juge l'alerte injustifiée, le « surveillant » perd des points. À l'internaute espion qui aura détecté le plus de fraudes ou de vols à la fin du mois, *Internet Eyes* promet une récompense pouvant atteindre jusqu'à 1000 livres ...<sup>10</sup>

Tony Morgan, le créateur d'*Internet Eyes*, explique l'intérêt de son système : « Il y a plus de quatre millions de caméras de surveillance, mais seulement une sur mille est regardée. De cette manière, les caméras seront regardées vingt-quatre heures sur vingt-quatre. C'est la meilleure arme de prévention de délits jamais inventée. »<sup>11</sup>.

Ce site internet suscite plusieurs inquiétudes. Par son intermédiaire, il est donc possible à quiconque à travers le monde de devenir surveillant. La diffusion des flux de vidéosurveillance à travers le réseau de l'Internet et leur visionnage par ces surveillants questionne le droit à l'image des personnes filmées. De plus, le fonctionnement de ce site, pouvant s'apparenter à un jeu lucratif, révèle un potentiel danger en déconnectant l'internaute du réel et ainsi le déresponsabilisant de ses actes.

Cette division dans la surveillance se présente aussi dans les sujets surveillés. Son application est de plus en plus précise sur le territoire, dans le but de cibler l'individu qui constitue la pièce élémentaire de la société. En effet, celui qui auparavant (et encore aujourd'hui) n'était qu'un individu suspect au milieu d'une foule sera bientôt

---

10 Ignacio Ramonet in *L'Empire de la surveillance*, Paris, Galilée, 2015, p. 105

11 Tony Morgan, *The Telegraph*, cité par Ignacio Ramonet in *L'Empire de la surveillance*, *op. cit.*, pp. 105-106

automatiquement identifiable grâce à la technique de la reconnaissance faciale. Cette technique repose sur un programme capable de reconnaître les personnes filmées à partir de photographies de visages couplées à des informations comme le nom, l'âge ou l'adresse répertoriés dans une base de données. Comme le montre Nicolas Combalbert dans son reportage intitulé « Reconnaissance faciale : tous filmés, tous identifiés ? » diffusé dans l'émission *Envoyé Spécial*, ce dispositif de reconnaissance faciale a été installé dans la gare routière de Madrid. Si un individu recherché par la police pénètre dans l'enceinte de la gare et que son profil a préalablement été intégré à la base de données, le système de reconnaissance faciale lancera une alerte dès que le visage de ce dernier sera détecté.<sup>12</sup> Grâce à cette technique, la surveillance semble faire un grand pas en terme d'efficacité.



**Figure 6**, COMBALBERT Nicolas, *Reconnaissance faciale : tous filmés, tous identifiés ?*, capture d'écran, 2016

Mais cette technique de reconnaissance faciale ne constituerait-elle pas une menace en portant atteinte à notre droit de libre circulation dans l'espace public et

<sup>12</sup> Nicolas Combalbert, « Reconnaissance faciale : tous filmés, tous identifiés ? » in *Envoyé Spécial*, diffusé le 11 février 2016, France 2

à notre anonymat ? Pouvons-nous accepter de le sacrifier pour « garantir » notre sécurité ?

### **3 – Résistance**

#### **a – Remises en cause**

Nous avons vu précédemment en quoi les dispositifs de vidéosurveillance sont des outils permettant, en théorie, une surveillance étendue et précise. Mais dans la pratique, à quoi servent réellement ces caméras ?

D'après un rapport établi par le ministère de l'intérieur britannique en février 2005<sup>13</sup>, les systèmes de vidéosurveillance ne seraient pas aussi efficaces que ce que l'on nous ferait croire. D'après cette étude, leurs faiblesses reposent en trois points.

#### ***La mise en œuvre technique***

Le premier problème d'efficacité de ces systèmes de surveillance réside dans leur positionnement inintéressant. En effet, Noé Le Blanc explique dans son article intitulé « Sous l'œil myope des caméras », que leur installation souvent faite à la hâte « se fonde sur des statistiques sur la criminalité incomplètes ou trop générales ; d'où le positionnement de certaines caméras dans des zones où elles se révèlent à peu près inutiles »<sup>14</sup>.

De plus, l'exploitation des images peut se révéler impossible à cause d'un obstacle obstruant le champ de vision de la caméra ou encore par des conditions lumineuses particulières sur (ou sous) -exposant l'image.

#### ***Le facteur humain***

La vidéosurveillance étant un moyen technique, il est important de considérer que l'action de surveillance repose sur des humains dans les salles de contrôle. Or

---

13 Martin Gill et Angela Spriggs, « Assessing the impact of CCTV », *Home Office Research Study*, n°292, Londres, 2005

14 Noé Le Blanc, « Sous l'œil myope des caméras », in *Manière de voir – Le Monde diplomatique*, n°133, février-mars 2014, Paris, Le Monde diplomatique, pp. 48-51

d'une part, le nombre d'écrans dans ces salles ne correspond évidemment pas au nombre de caméras, ne permettant pas l'affichage de l'intégralité des caméras et d'autre part, les opérateurs ne peuvent surveiller correctement plus d'un écran.

De plus, une étude révèle que 86% des individus surveillés ont moins de 30 ans, que 93% sont de sexe masculin, et que les individus noirs ont deux fois plus de chances d'attirer l'attention des opérateurs que les individus blancs<sup>15</sup>. Avant même que le moindre délit ou crime ne soit commis, la surveillance des opérateurs est donc déjà ciblée à partir de préjugés et manque généralement de neutralité.

### ***La démesure des objectifs***

La troisième faiblesse de ces dispositifs réside dans la démesure des objectifs qui leur sont assignés. Les entreprises de surveillance ont en effet tendance à présenter leurs dispositifs comme des solutions miracles, suscitant de trop grandes attentes vis-à-vis des systèmes de surveillance. Or il ne s'agit que d'un moyen technique non-autonome, nécessitant une exploitation des images enregistrées.

Contrairement à ce que de nombreux discours prétendent, Claude-Marie Vadrot explique dans son ouvrage *La grande surveillance – Caméras, ADN, portables, Internet...* que les systèmes de vidéosurveillance n'ont aucune efficacité préventive, qu'il s'agisse de délinquance, de crime ou de terrorisme, même lorsque leur présence est signalée. Il s'appuie même sur des chiffres pour le prouver : « Depuis avril 2001, tous les bus de la ligne 255 de Saint-Denis sont équipés. Pendant les vingt-quatre mois précédant l'équipement : 26 délit violents constatés ; au cours des vingt-quatre mois suivants, jusqu'au printemps 2004 : 48. »<sup>16</sup>. En revanche, ces systèmes se révèlent être utiles après-événement pour identifier les auteurs et reconstituer la scène.

Tous ces facteurs relativisent donc l'efficacité de la vidéosurveillance, du moins par rapport à la façon dont les hommes politiques ont l'habitude de la présenter. Claude-Marie poursuit en ce sens :

---

15 Gary Armstrong et Clive Norris, « CCTV and the social structuring of surveillance », *Crime Prevention Studies*, vol. 10, New York, 1999

16 Claude-Marie Vadrot in *La grande surveillance – Caméras, ADN, portables, Internet...*, Paris, Seuil, 2007, p. 49

La vidéosurveillance ne répond pas à une authentique demande sociale mais illustre ou appuie un choix d'élus cherchant à faire valoir leur rôle dans la mise en place d'une politique de sécurité. Des politiques d'équipements sécuritaires essentiellement réservées aux centres-villes et aux rues commerçantes : les élus installent rarement des systèmes de vidéosurveillance dans les quartiers dits « sensibles », ce qui tendrait à prouver qu'ils croient bien plus à ses vertus électorales qu'à ses fonctions préventives.<sup>17</sup>

Alors à quoi servent toutes ces caméras de surveillance si elles sont si peu efficaces ? Si elles voient si mal, ne seraient-elles pas plus là pour être vues ?

Depuis les attentats du 11 septembre 2001, la sécurité est devenue un sujet pilier de la scène politique, profitant ainsi à un « business de la surveillance ».

## **b – Questions éthiques**

L'ampleur de l'usage de ces caméras de surveillance dans nos sociétés pose des questions sur l'atteinte aux libertés, ainsi que sur la protection de l'identité. Qu'a-t-on le droit de filmer dans un établissement commercial ou sur un lieu de travail ? Qui visionne les images enregistrées par les caméras ? Que deviennent ces images après leur exploitation ? Toutes ces questions sont d'autant plus pertinentes avec l'actuel essor de la reconnaissance faciale. Avec cette technique, la reconnaissance automatique de l'identité d'un individu passe par son visage : d'autant plus d'enjeux dans la protection de la vie privée ...

De plus, l'omniprésence des systèmes de surveillance instaure un climat d'oppression au quotidien, pouvant faire allusion à la société décrite par George Orwell dans *1984*, dans laquelle des « télécrans » sont disséminés à travers la ville afin de surveiller les moindres faits et gestes des citoyens.<sup>18</sup> Ces systèmes, placés entre de mauvaises mains, peuvent représenter un véritable danger en constituant les premiers pas vers l'instauration d'une dictature. C'est pourquoi il est important que les citoyens soient conscients des dangers ainsi que des droits dont ils disposent vis-à-vis de ces dispositifs.

---

17 *Ibid.*, p. 35

18 George Orwell in *1984*, Paris, Gallimard, 1950, p. 13

Malgré les textes de lois régissant la vidéoprotection (articles L251-1 à L255-1 du code de la sécurité intérieure<sup>19</sup>), il est difficile de répondre à toutes ces questions puisque dans la pratique, elles ne sont pas toujours respectées. Le domaine de la sécurité, et par conséquent de la vidéosurveillance, comportent des zones de flou.

Pour lutter contre les différents dangers liés aux mauvais usages de la vidéosurveillance, un organisme a été mis en place : la CNIL. La Commission nationale de l'informatique et des libertés, fondée en 1978, est une autorité administrative indépendante française réglementant l'usage de l'informatique en France afin de protéger les libertés individuelles et publiques. Par extension, la réglementation de la vidéosurveillance a été confiée à cet organisme. Cette commission a par exemple publié un article intitulé « Vidéosurveillance / vidéoprotection : les bonnes pratiques pour des systèmes plus respectueux de la vie privée »<sup>20</sup> sur son site internet, dans lequel sont fournies six fiches pratiques rappelant les règles et lois à respecter en fonction du lieu d'installation (voie publique, travail, établissements scolaires, commerces, immeubles d'habitation, logement). Il est d'ailleurs intéressant de définir la différence entre le terme général de *vidéosurveillance*, et celui de *vidéoprotection*, employé lorsque la première concerne le domaine public depuis la loi d'orientation et de programmation pour la performance de la sécurité intérieure (LOPPSI) du 14 mars 2011<sup>21</sup>. Le terme *protection* n'a pas été choisi au hasard : il se veut plus rassurant que l'angoissant *surveillance* et fait donc moins peur dans les discours.

Dans le document concernant la voie publique, en plus des textes de lois de référence, il est par exemple précisé que « seules les autorités publiques (les mairies notamment) peuvent filmer la voie publique », que les entreprises « peuvent seulement filmer les abords immédiats de leurs bâtiments et installation (...) dans les lieux susceptibles d'être exposés à des actes de terrorisme », ou encore que « la

---

19 Legifrance, « Code de la sécurité intérieure », [En ligne], mis en ligne le 12 mars 2012. URL : <https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000025505406&idSectionTA=LEGISCTA000025508190&cidTexte=LEGITEXT000025503132&dateTexte=20120618> (consulté le 8 mars 2016)

20 CNIL, « Vidéosurveillance/vidéoprotection : les bonnes pratiques pour des systèmes plus respectueux de la vie privée », [En ligne], mis en ligne le 21 juin 2012. URL : <https://www.cnil.fr/fr/videosurveillance-vidéoprotection-les-bonnes-pratiques-pour-des-systemes-plus-respectueux-de-la-vie> (consulté le 8 mars 2016)

21 Le Monde, « Quand la « vidéoprotection » remplace la « vidéosurveillance », [En ligne], mis en ligne le 16 février 2010. URL : [http://www.lemonde.fr/societe/article/2010/02/16/quand-la-vidéoprotection-remplace-la-vidéosurveillance\\_1306876\\_3224.html](http://www.lemonde.fr/societe/article/2010/02/16/quand-la-vidéoprotection-remplace-la-vidéosurveillance_1306876_3224.html) (consulté le 8 mars 2016)

conservation des images ne doit pas excéder un mois »<sup>22</sup>. Toutes ces informations concernant la réglementation en rapport à la vidéosurveillance sont à connaître pour quiconque souhaitant s'équiper, afin de respecter la vie privée d'autrui et éviter tout litige.

## c – Activismes

Tandis que la CNIL sert à contrôler et réglementer les usages des systèmes de vidéosurveillance, des associations regroupant des citoyens militants se sont mises en place afin de prendre des mesures plus radicales. C'est le cas du collectif *Souriez vous êtes filmés*<sup>23</sup>, fondé en 1995, dénonçant les menaces que représente la vidéosurveillance pour les libertés publiques et individuelles et s'étant « donnée comme but le retrait des caméras de vidéo-surveillance ». Leur site internet leur sert de plateforme d'échange en diffusant des informations concernant leurs activités et en recevant des propositions d'articles. Leurs activités s'apparentent à de l'activisme, à grands coups de campagnes de masquage de caméras en France et dans le monde entier et de recours juridiques pour la suppression de caméras de surveillance.

Une autre forme d'activisme peut se présenter à travers l'art. En effet, certains artistes s'engagent pour la protection de l'identité en développant une démarche artistique. C'est par exemple le cas d'Adam Harvey avec ses créations textiles intitulées *Stealth wear*<sup>24</sup>. Il s'est servi de matières isolantes pour créer toute une série de vêtements, empêchant les émanations thermiques corporelles d'être observables. Comme nous pouvons le voir sur l'illustration suivante, les parties du corps habillées par le vêtement ressortent très denses par rapport aux parties dénudées, devenant visibles sous l'objectif de la caméra thermique. Avec cette collection de vêtements, l'artiste a exprimé une volonté d'échapper à la vision des caméras de surveillance thermiques, de disparaître dans son environnement tel un caméléon.

---

22 CNIL, « Vidéosurveillance-vidéoprotection - Sur la voie publique », [En ligne], mis en ligne en octobre 2015. URL : [https://www.cnil.fr/sites/default/files/typo/document/CNIL\\_Video\\_voie\\_publique.pdf](https://www.cnil.fr/sites/default/files/typo/document/CNIL_Video_voie_publique.pdf) (consulté le 8 mars 2016)

23 Souriez vous êtes filmés, [En ligne]. URL : <http://souriez.info/> (consulté le 8 mars 2016)

24 Adam Harvey, *Stealth wear*, [En ligne]. URL : <https://ahprojects.com/projects/stealth-wear/> (consulté le 8 mars 2016)



**Figure 7**, HARVEY Adam, *Stealth wear*, photographies, 2013

Le même artiste a conçu un autre projet pour échapper à la reconnaissance faciale. Intitulé *CV Dazzle*<sup>25</sup>, Adam Harvey a réalisé une série de photographies dans lesquelles il a élaboré des maquillages ainsi que des coiffures empêchant les logiciels de reconnaissance faciale de détecter le visage de l'individu. Comme nous pouvons le voir sur la figure suivante, en « déstructurant » le visage à l'aide d'un maquillage créant d'importants contrastes sur le visage, l'algorithme n'est alors plus capable de considérer ce qu'il voit comme un visage. De plus, l'artiste partage les différentes techniques à employer pour se camoufler de la reconnaissance faciale sur son site internet.

---

25 Adam Harvey, *CV Dazzle*, [En ligne]. URL : <https://ahprojects.com/projects/cv-dazzle/> (consulté le 8 mars 2016)



**Figure 8**, HARVEY Adam, *CV Dazzle*, photographies, 2010

Mais cette démarche d'activisme artistique peut aller encore plus loin, comme nous le montre le projet de Leonardo Selvaggio intitulé *URME*<sup>26</sup>. L'artiste a en effet modélisé son propre visage en 3D, puis diffusé le modèle du masque sur l'Internet afin que les internautes puissent l'imprimer à l'aide d'une imprimante 3D. Il encourage ainsi les citoyens du monde entier désireux de conserver leur anonymat à arborer son masque lorsqu'ils se déplacent dans des lieux publics équipés de systèmes de surveillance à reconnaissance faciale. Le programme considérera alors que la personne identifiée est Leonardo Selvaggio puisqu'il reconnaîtra son visage. L'artiste souhaite ainsi brouiller le système, perturber la mécanique en insérant un caillou dans l'engrenage. En livrant son visage, il va même jusqu'à sacrifier son identité pour protéger son prochain.

<sup>26</sup> Leonardo Selvaggio, *URME*, [En ligne]. URL : <http://leoselvaggio.com/urmesurveillance/> (consulté le 8 mars 2016)



**Figure 9**, SELVAGGIO Leonardo, *URME*, photographies, 2014

Ces trois démarches visent à perturber l'identification des personnes par les systèmes de vidéosurveillance de différentes manières. En effet, dans son projet *Stealth wear*, Adam Harvey cherche à faire disparaître le corps (et par conséquent, le visage) de la vision de la caméra, tandis que dans *CV Dazzle*, il cherche à le dénaturer en trompant la codification du schéma d'identification d'un visage. La troisième manière est celle de Leonardo Selvaggio qui propose de remplacer le visage en mettant le sien en avant. On constate donc que ces trois propositions de perturbation de la surveillance sont conçues pour le surveillé et tournées vers l'observateur caché derrière la caméra. Elles fonctionnent comme des filtres cherchant à masquer le champ de vision du dispositif de surveillance et par extension échapper au regard du surveillant.

Cependant, ces trois projets constituent plus des expériences artistiques que des propositions d'utilisation quotidienne, étant peu pratiques à mettre en œuvre. Cela constitue d'ailleurs leur limite. De plus, est-ce une solution responsable que de chercher à se cacher de celui qui se cache ? Cela démontre en tout cas que peu de

solutions s'offrent aux citoyens souhaitant s'affranchir individuellement du regard de la vidéosurveillance.

Mais ne serait-ce pas plutôt dans le partage et par conséquent, dans la prise de conscience que résiderait l'intérêt de ces projets ? La solution ne se trouverait-elle donc pas à une échelle plus collective ?

Ailleurs, certains artistes utilisent des caméras de vidéosurveillance dans leurs projets, passant ainsi du statut d'observé au statut d'observateur. Mais que peut apporter ce changement de point de vue à leur démarche artistique et dans quel but ?

## **II – La surveillance par l’image**

### **1 – Usages politiques de l’image**

#### **a – L’identité et la surveillance**

La question de l’identité est centrale dans l’imagerie de surveillance. Étant filmés dans de très nombreux lieux dans lesquels nous nous rendons et depuis l’arrivée des caméras à reconnaissance faciale, l’anonymat est plus que jamais remis en question. La protection de l’identité constitue la première cause de défense des activistes évoqués précédemment. De plus, en présentant plusieurs techniques de camouflage du corps et du visage à travers leurs créations, les artistes militants proposent différentes manières de protéger notre identité face à la vidéosurveillance. Mais en quoi ces caméras portent-elles atteinte à notre identité ?

Le principal reproche fait à ces caméras de vidéosurveillance réside dans leur caractère intrusif. Même en étant conscient de la présence de la vidéosurveillance dans notre quotidien, il reste surprenant de découvrir la quantité de caméras installées, constituant un regard permanent sur chacun d’entre nous. Contrairement aux images que nous avons l’habitude de produire (photographies, vidéos, ...), celles produites par ces caméras ne nous appartiennent pas : réalisées à notre insu, nous ne pourrions, la plupart du temps, jamais les voir. Cette dépossession entraîne une forme de frustration et un sentiment de suspicion d’atteinte à notre vie privée.

Mais grâce à l’omniprésence et la continuité de l’enregistrement des images, ces caméras de vidéosurveillance permettent d’enregistrer des preuves visuelles de crimes ou délits, et idéalement l’identification de leurs auteurs. Leurs images servent alors de support dans une enquête policière afin de repérer les suspects. Ce fut le cas avec le photogramme extrait de la vidéosurveillance de l’aéroport de Zaventem lors des attentats de Bruxelles du 22 mars 2016 et largement diffusé dans les médias. Dès le soir des attentats, les deux individus à gauche de l’image avaient été identifiés comme kamikazes mais le troisième suspect ne pouvait l’être et s’est alors trouvé

surnommé « l'homme au chapeau », à cause de son bob lui masquant partiellement le visage. La diffusion de l'image avait justement servi d'appel à témoin dans le but d'identifier l'inconnu.



**Figure 10**, Police belge, *Vidéosurveillance de l'aéroport de Zaventem lors des attentats de Bruxelles du 22 mars 2016*, photogramme, 2016

Mais ne serait-ce pas restreindre le pouvoir de l'image en ne considérant que l'utilité sécuritaire de la diffusion de cette dernière ?

### **b – L'image témoin**

En témoigne l'exemple précédent, depuis quelques décennies, les images de vidéosurveillance sont régulièrement exploitées par la police, puis relayées par les médias suite à des événements majeurs tels que des attaques terroristes ou des catastrophes naturelles. Lorsque l'événement est imprévisible, ces dernières s'avèrent plus efficaces que les médiums dits « classiques » (photographie, vidéo, ...), nécessitant l'acte du déclenchement et par conséquent la présence d'un auteur. De plus, la continuité de la captation et le point de vue rendraient donc ces images de vidéosurveillance plus neutres que celles issues de tout autre dispositif de prise de vue.

Le photogramme visible ci-dessous est issu de la vidéosurveillance du restaurant *Casa Nostra*, touché par les attentats de Paris du 13 novembre 2015. Ces images font partie des seules à témoigner de l'intégralité d'une des attaques terroristes et à avoir été relayées dans les médias. Nous pouvons donc parler ici d'images témoins.

L'achat des droits d'exploitation de la vidéo, ainsi que des photogrammes qui en étaient issus, par le journal britannique *Daily Mail* avait d'ailleurs fait polémique. En effet, la vente avait été jugée inappropriée et le gérant du restaurant avait été largement critiqué d'avoir voulu profiter financièrement de la tragédie. Cette polémique prouve bien que ces images sont loin d'être banales et sont devenues les traces visuelles d'un événement historique majeur.

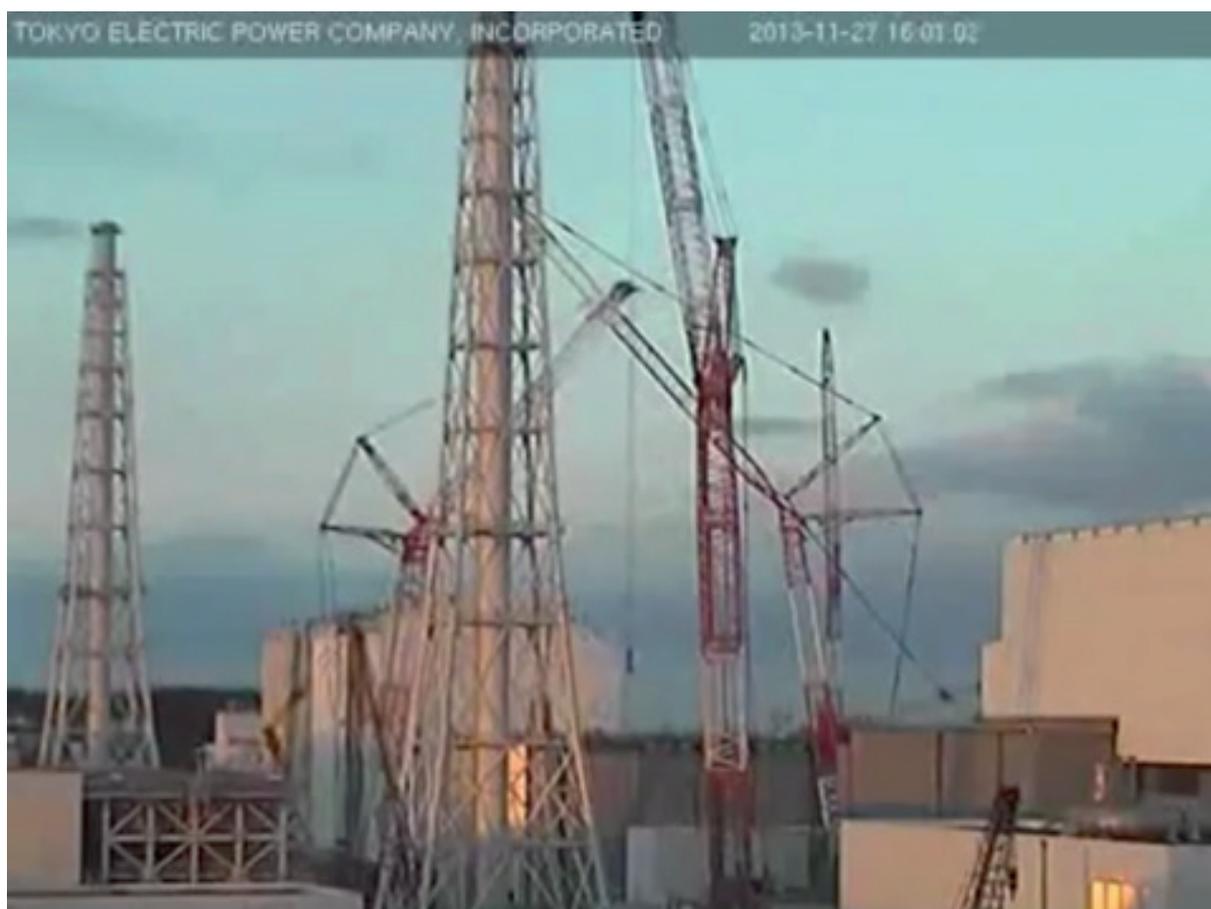


**Figure 11**, Daily Mail, *Vidéosurveillance du restaurant Casa Nostra lors des attentats de Paris du 13 novembre 2015*, photogramme, 2015

Cet exemple démontre que ces images tirées d'un enregistrement de vidéosurveillance peuvent devenir les preuves du passé. Mais celles-ci peuvent aussi constituer les preuves d'un présent grâce à l'Internet et le streaming.

Cette forme de témoignage a intéressé la société *TEPCO* (*Tokyo Electric Power Company*) suite à une catastrophe historique. En effet, suite au séisme suivi d'un tsunami ayant touché la moitié Nord du Japon le 11 mars 2011, entraînant l'accident nucléaire

de la centrale de Fukushima, *TEPCO* a choisi de diffuser en continu sur l'Internet des images du lieu par l'intermédiaire de quatre caméras<sup>27</sup>. En plus des raisons techniques évidentes liées aux radiations empêchant une véritable documentation photographique ou vidéo « traditionnelle », la société a ainsi cherché à faire preuve de transparence vis-à-vis de la population mondiale. Ces images issues de plusieurs points de vue sont accessibles en temps réel, depuis le monde entier. Tout est montré ... du moins ce qui se passe dans le cadre.



**Figure 12**, TEPCO, *Webcam diffusant en direct depuis la centrale nucléaire de Fukushima*, photogramme, 2013

Mais comment des images, paraissant si banales et dans lesquelles rien ne se passe, peuvent avoir autant d'impact auprès du public ?

<sup>27</sup> TEPCO, *Fukuichi Live camera*, [En ligne], mis en ligne le 9 janvier 2014. URL : <http://www.tepco.co.jp/en/nu/f1-np/camera/index-e.html> (consulté le 21 avril 2016)

## c – Le pouvoir de l'image

En étudiant l'imagination des foules, nous avons vu qu'elles sont impressionnées surtout par des images. Si l'on ne dispose pas toujours de ces images, il est possible de les évoquer par l'emploi judicieux des mots et des formules. Maniés avec art, ils possèdent vraiment la puissance mystérieuse que leur attribuaient jadis les adeptes de la magie. Ils provoquent dans l'âme des multitudes les plus formidables tempêtes, et savent aussi les calmer. On élèverait une pyramide plus haute que celle du vieux Khéops avec les seuls ossements des victimes de la puissance des mots et des formules.<sup>28</sup>

En 1895, Gustave Le Bon avait déjà compris que les mots, et en particulier les images, possédaient un véritable pouvoir dans leur diffusion auprès du peuple. Les images étant universellement compréhensibles, car s'affranchissant de la barrière de la langue, elles constituent une forme de vecteur de la pensée des plus efficaces. Avec les progrès techniques réalisés en terme de diffusion d'images avec la télévision puis l'Internet, leur impact n'est qu'amplifié.

Si la diffusion d'images issues de vidéosurveillance telles que celles d'attentats possède un tel pouvoir, c'est avant tout pour leur attache au réel. Par l'aspect commun des lieux dans lesquels ces événements se déroulent, il est facile de se projeter dans ces images habituellement de l'ordre du banal, se transformant alors en images de l'extra-ordinaire. Afin d'assouvir nos besoins de représentation de l'événement, leur visionnage devient nécessaire, entraînant alors une certaine peur mêlée à une forme de fascination sur le spectateur.

De plus, au-delà de l'utilité policière de la diffusion de l'image des trois suspects de l'attentat de l'aéroport de Bruxelles évoquée précédemment, cette dernière présente une utilité politique en véhiculant un certain nombre de messages auprès du public. Dans un climat tendu post-attentat, la diffusion de cette image permet aux autorités belges de montrer aux citoyens du monde entier qu'ils maîtrisent la situation en ayant des pistes pour l'enquête le jour même. De plus, en dévoilant l'identité des auteurs, l'image de vidéosurveillance permet à la population de se rassurer et de mieux cerner l'événement, en assouvissant le besoin de mettre un visage sur ces criminels.

---

28 Gustave Le Bon in *Psychologie des foules*, Paris, Presses Universitaires de France, 1895, pp. 59-60

## 2 – Esthétique de la surveillance

Différentes caractéristiques participent à la constitution d'une esthétique de la surveillance. Grâce à la récurrence de ces éléments, les images issues de ces caméras de surveillance sont très rapidement identifiables. Mais quels sont ces codes esthétiques de l'image de surveillance et comment sont-ils utilisés ?

### a – Le point de vue

La première caractéristique de l'imagerie de surveillance est son point de vue. Cette notion est premièrement déterminée par l'emplacement et l'orientation du dispositif de prise de vue. Comme nous l'avons déjà évoqué dans la première partie, la position des caméras de vidéosurveillance est systématiquement en hauteur afin de surplomber l'espace sans qu'aucun obstacle ne vienne obstruer le champ de vision et leur orientation est très généralement dirigée vers le sol. Cette position et cette orientation systématiques des appareils de prises de vue homogénéisent les cadrages et les perspectives des images issues de la vidéosurveillance.

Mais la notion de point de vue est aussi relative à l'emplacement et l'orientation du sujet photographié ou filmé. En effet, si l'emplacement des caméras est qualifié d'« en hauteur », cette localisation se réfère à la position de l'individu au sol, défini par son échelle ainsi que celles de ses activités quotidiennes. Le regard de l'humain étant plutôt porté à sa hauteur ou à celle de ses activités, les caméras de surveillance situées beaucoup plus haut se trouvent alors souvent en dehors de son champ visuel. De plus, les dimensions des caméras se réduisant de plus en plus, les systèmes de vidéosurveillance ont tendance à être moins visibles et s'intègrent aux paysages urbains.

L'œuvre de l'artiste ukrainienne Zhanna Kadyrova, intitulée *Invisible Forms, Security Cameras*<sup>29</sup>, présentée sur la photographie ci-dessous au *Cent-Quatre* à Paris en novembre 2015, s'intéresse justement à la place physique qu'occupent ces caméras de surveillance et à la façon dont elles s'intègrent dans l'architecture urbaine.

---

29 Zhanna Kadyrova, *Invisible Forms*, [En ligne]. URL : <http://www.kadyrova.com/#!invisible-forms/c87n> (consulté le 21 avril 2016)

En effet, cette sculpture en béton représente deux angles de champ de deux caméras de surveillance se croisant. L'artiste matérialise l'invisible afin de mettre en évidence la présence des caméras ainsi que leur couverture de l'espace, alors que ces dernières ont pourtant tendance à disparaître dans le paysage urbain et se faire oublier. Les volumes représentés sont d'ailleurs affinés par rapport à la réalité du champ couvert par les caméras. Pourvues de focales courtes formant des pyramides dont la base est bien plus large, la sculpture aurait alors été trop encombrante si le véritable champ avait été matérialisé. Mais la puissance symbolique de l'œuvre n'a pas besoin de cette objectivité. Zhanna Kadyrova cherche à susciter une prise de conscience de l'étendue et de l'omniprésence de la surveillance au quotidien et du danger que celle-ci représenterait si l'on venait à l'oublier.



**Figure 13**, KADYROVA Zhanna, *Invisible Forms, Security Cameras*, photographie, 2015

Ce point de vue en hauteur et en plongée, depuis lequel les caméras de surveillance captent leurs images, produit inévitablement un rapport de force et de domination du voyant sur le vu, du surveillant sur le surveillé.

Dans sa série *Sites of technology*, le photographe américain Lewis Baltz s'est intéressé entre 1989 et 1991 aux sociétés et centres de recherches spécialisés dans les « hautes technologies ». En plus d'y avoir photographié les bureaux et laboratoires à la propreté clinique et vides de présence humaine, renvoyant alors ces lieux au domaine de l'irréel, il a aussi photographié les moniteurs permettant la surveillance de ces véritables forteresses. Photographiée en plongée, la personne semble être observée à son insu, ne pouvant échapper à l'objectif de la caméra. Ses moindres faits et gestes sont capturés, elle semble comme prise au piège. Le point de vue en hauteur domine le sujet. De plus, cette photographie s'intègre parfaitement au reste de la série en reflétant un autre aspect de ces lieux ultra-sécurisés.



**Figure 14**, BALTZ Lewis, *Sites of technology*, photographie, 1989-1991

En adoptant un point de vue en hauteur, la prise de vue aérienne permet une surveillance encore plus totale, récemment facilitée par des appareils tels que les drones. Ces aéronefs dépourvus de pilote sont utilisés dans des missions militaires de reconnaissance de territoires étrangers ou de bombardements. Depuis quelques années, des modèles grand public ont été largement commercialisés, démocratisant

ainsi la prise de vue photographique et vidéo aérienne.

Suite aux frappes d'un drone militaire américain ayant tué des civils lors de rassemblements collectifs au Pakistan et au Yemen, Tomas van Houtryve a décidé de réaliser sa série *Blue Sky Days*<sup>30</sup>. Il s'est servi d'un drone grand-public afin de photographier depuis le ciel les mêmes événements, cette fois-ci aux États-Unis. Le point de vue en hauteur et orienté vers le sol écrase toutes les perspectives : le drone domine les espaces qu'il survole. De plus, les conditions de lumière produisent des ombres contrastées et nettes à cette distance réduisant l'espace à une surface. Les silhouettes allongées et plaquées au sol semblent prises au piège de cette image.



**Figure 15,** VAN HOUTRYVE Tomas, *Blue Sky Days*, photographie, 2015

30 Tomas Van Houtryve, *Blue Sky Days*, [En ligne]. URL : <http://tomasvh.com/works/blue-sky-days/> (consulté le 21 avril 2016)

## **b – Les rendus d’image**

À travers les différentes techniques de captation des caméras de vidéosurveillance, les images qui en sont issues présentent plusieurs types de rendus, faisant ainsi leurs caractéristiques propres. Nous en analyserons trois variables : les inscriptions, les textures et les couleurs.

### ***Les inscriptions***

Les premières caractéristiques des images de vidéosurveillance en terme de rendu sont les inscriptions que celles-ci peuvent contenir. Intégrées aux images pour des raisons pratiques, les inscriptions renseignent différentes informations pour l’observateur.

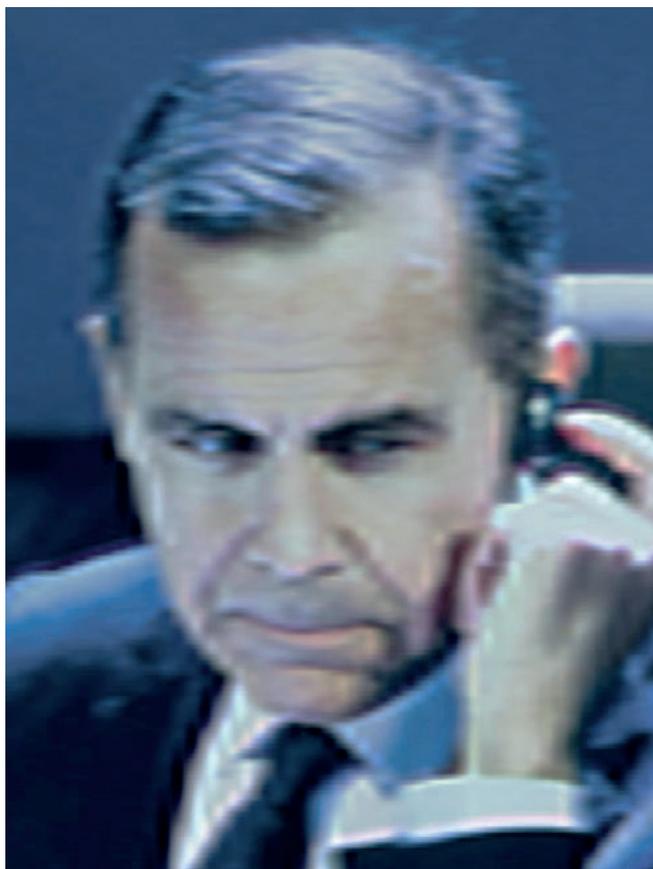
Sur la photographie de Lewis Baltz visible précédemment, nous pouvons voir que la date, l’heure et probablement le numéro d’identification de la caméra étaient inscrits sur le moniteur photographié. Ces informations rattachent cette image au réel et par conséquent, participent à son authentification en prouvant que l’action photographiée à travers le moniteur s’est bien déroulée.

### ***Les textures***

Les textures des images influent aussi sur leurs rendus en fonction des techniques de prise de vue des caméras de surveillance (vidéo ou numérique) et de leur ancienneté. Les caméras vidéo présentent généralement des images assez molles, à faible définition (comme le montre la photographie de Lewis Baltz ci-dessus), tandis que les images provenant de caméras numériques ont tendance à accentuer le rendu « net » voire pixelisé et comportent très souvent des aberrations optiques et colorimétriques, très visibles lorsque des recadrages sont effectués. Les images provenant de caméras de surveillance ne sont pas qualitatives par rapport aux critères habituels, les fabricants se souciant plutôt de leur résistance et de leur fiabilité.

Le photographe Daniel Mayrit s’est intéressé à ces textures en collectant les portraits des cent personnalités les plus puissantes de Londres dans une série intitulée

*They haven't seen their faces*<sup>31</sup>, constituée de fausses images de vidéosurveillance. Suite à une distribution de tracts sur lesquels figuraient les portraits, tirés de vidéosurveillances, de jeunes suspectés d'avoir participé aux émeutes de Londres de 2011, Daniel Mayrit s'est questionné sur le pouvoir de ces images. En imitant le rendu pixelisé (voire compressé), les couleurs désaturées et les cadrages des visages très serrés, le photographe simule l'effet produit par le grossissement d'images extraites de vidéosurveillances, créant ainsi une ambiguïté concernant le caractère suspect des personnes ainsi représentées.



**Figure 16**, MAYRIT Daniel, *They haven't seen their faces*, photographie, 2015

À travers l'utilisation d'images sorties de leur contexte et à leur rendu caractéristique, ce projet nous montre à quel point notre lecture de l'image est conditionnée et nous questionne justement sur notre rapport à celle-ci dans nos sociétés.

---

31 Daniel Mayrit, *They haven't seen their faces*, [En ligne]. URL : <http://www.danielmayrit.com/> (consulté le 21 avril 2016)

## **Les couleurs**

En plus d'influer sur les textures, la technique de prise de vue rend différents types de couleurs et participe ainsi à une caractérisation des images issues de caméras de surveillance. Comme nous pouvons l'observer sur la photographie d'Adam Harvey analysée en première partie (figure 8), ces couleurs peuvent aller de dégradés violet-jaune pour des caméras thermiques en fonction de l'intensité thermique perçue par la caméra, jusque des tons verts pour d'autres caméras infra-rouge.

La photographie suivante est issue de la série *Nacht*<sup>32</sup>, réalisée par Thomas Ruff en 1992. Pour ce projet, il s'est promené dans sa ville de Düsseldorf et a photographié des paysages urbains banals comme des immeubles, des ponts ou encore des parkings. Mais ces images n'en demeurent pas moins intrigantes. Le photographe s'est servi d'un appareil équipé d'un système infra-rouge permettant la prise de vue de nuit, donnant cette fameuse coloration verte aux images produites ainsi que ce halo autour des sources lumineuses. Cette technique de prise de vue infrarouge étant aussi employée par les militaires, le rendu des photographies de Thomas Ruff fait justement écho aux images issues des opérations américaines pendant la première guerre en Irak en 1991, largement relayées par les télévisions du monde entier de l'époque. Le rendu de ces images, dans lesquelles rien de spectaculaire ne se passe, suscite une curieuse inquiétude chez le spectateur qui se met alors à chercher instinctivement un élément suspect dans un recoin du cadre. La texture, associée à l'imaginaire collectif de la guerre, véhicule ainsi une forme de peur dans l'image.

---

32 Stefan Gronert, *L'école de photographie de Düsseldorf*, Paris, Hazan, 2009, p. 49



**Figure 17**, RUFF Thomas, *Nacht*, photographie, 1992

Les caméras de surveillance étant conçues à des fins utilitaires et non esthétiques, les images sont parfois considérées comme « sales ». En effet, leur forte présence de bruits, leurs couleurs mal équilibrées (désaturées ou sursaturées), leurs aberrations chromatiques et optiques liées aux objectifs de mauvaise qualité des caméras sont autant de paramètres habituellement chassés des photographes et vidéastes car en dehors des codes de représentation classique d'une « bonne » image. Mais ici, tous ces éléments participent à la véracité de l'image de surveillance et font partie intégrante de cette esthétique.

### **c – La sérialité**

La force des caméras de surveillance réside dans leur répartition spatiale et dans leurs échanges à travers les réseaux, souvent représentés par l'image du mur de moniteurs de vidéosurveillance symbolisant l'omniprésence de ces caméras. La

multiplicité est donc une caractéristique fondamentale de la vidéosurveillance, véhiculée dans leurs images symbolisant l'herméticité et la complexité de ces systèmes.

La question de la sérialité est très importante dans le projet intitulé *Spirit is a bone*<sup>33</sup>, du duo d'artiste Broomberg & Chanarin. Ils se sont servis d'un système de caméras à reconnaissance faciale développé en Russie pour la sécurité publique ainsi que la surveillance des frontières. Ce dispositif équipé de quatre objectifs produit des modèles en trois dimensions des visages des passants à leur insu. Ils ont ainsi réalisé une grande quantité de portraits et les ont exposés sous forme de mosaïque recouvrant l'un des murs du Musée de la Photographie d'Anvers, lors de leur exposition *Everything Was Beautiful and Nothing Hurt* en 2014. À travers ce dispositif, le visage est quantifié pour ne devenir qu'une image résultante de données numériques.

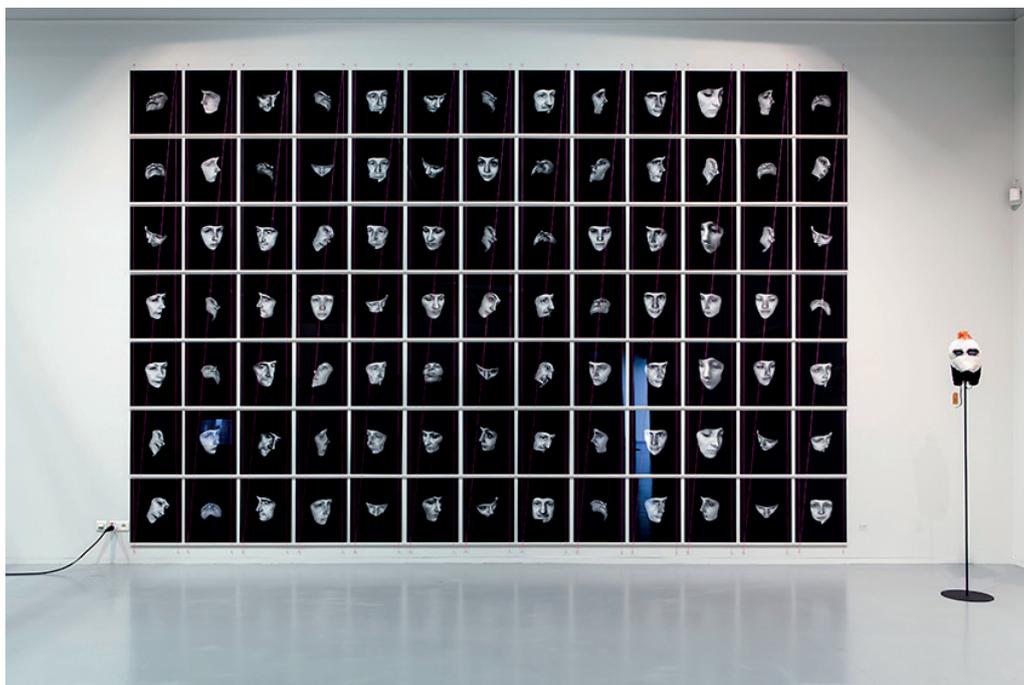


**Figure 18**, Broomberg & Chanarin, *Spirit is a bone*, photographie, 2014

---

33 Broomberg & Chanarin, *Spirit is a bone*, [En ligne]. URL : <http://www.broombergchanarin.com/root/two-eyes-above-a-nose-above-a-mouth/> (consulté le 21 avril 2016)

Ci dessous, la photographie de l'exposition de ce projet au musée de la photographie d'Anvers dévoile l'intégralité des portraits. Juxtaposés, les tirages vont même jusqu'à recouvrir le mur. Cette masse ainsi formée par la quantité de visages enregistrés fait écho au motif de matrice des bases de données servant aux répertiages.



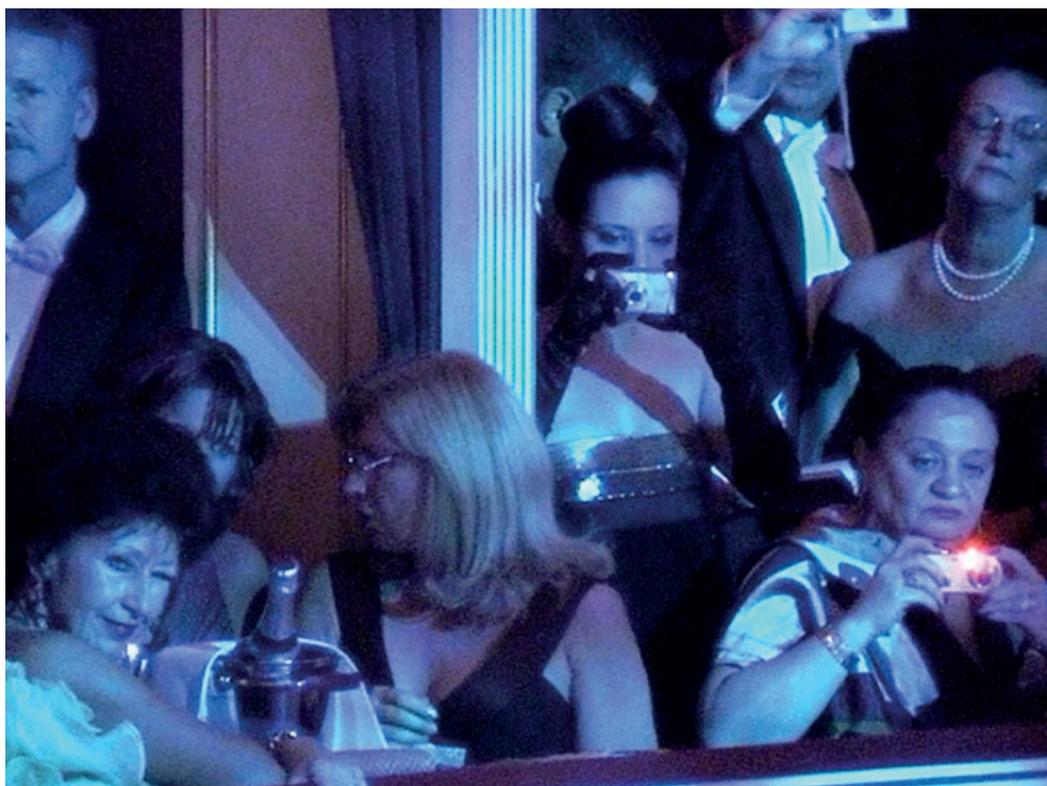
**Figure 19**, VERDIJK Lukas et FAASSEN Vesna, *Vue d'exposition de Spirit is a bone*, photographie, 2014

Ce même motif de matrice se retrouve dans les panoramas du *Surveillance Panorama Project*<sup>34</sup> du photographe suisse Jules Spinatsch. Composé de dix-mille huit images capturées par deux caméras de vidéosurveillance de l'opéra de Vienne, *No.4 Vienna MMIX* englobe l'espace. L'artiste en a pris le contrôle pour photographier l'ensemble du lieu durant toute une nuit (précisément de 20h32 à 5h17), au rythme d'une photographie prise toutes les trois secondes. Il les a ensuite assemblées chronologiquement pour constituer le panorama visible en annexe. Le photographe explique dans un article publié sur le site du *British Journal of Photography* qu'en assemblant ce panorama, il a souhaité réaliser un portrait de la haute société viennoise présente dans cet opéra, composée ce soir-là de « la moitié du gouvernement autrichien,

<sup>34</sup> Jules Spinatsch, *Surveillance Panorama Project*, [En ligne]. URL : <http://jules-spinatsch.ch/?p=76> (consulté le 21 avril 2016)

tous les grands industriels et quelques célébrités »<sup>35</sup>. L'objet qui en résulte devient une forme d'étude sociologique. Jules Spinatsch est intéressé par la neutralité offerte par l'automatisme de ces caméras de surveillance dont les cadrages et les compositions sont contrôlés par un programme informatique. L'homme n'intervenant pas dans la réalisation de ces images, aucune distinction n'est faite entre ces individus, les plaçant tous sur le même plan.

En plus du panorama visible en annexe, le photographe présente indépendamment certaines images le constituant afin de montrer des détails qu'il souhaite mettre en évidence. Il met ainsi en tension le caractère systématique de la caméra de surveillance et le regard d'un observateur. En venant ainsi se focaliser sur ces détails, la démarche de Jules Spinatsch est proprement photographique. Il fait ainsi la distinction entre le regard et la vision.



**Figure 20**, SPINATSCH Jules, *Détails de Surveillance Panorama Project – No.4 Vienna MMIX*, photographie, 2009

35 Tom Seymour, « A secret surveillance project of the Vienna Opera Ball », *British Journal of Photography*, [En ligne], mis en ligne le 6 mai 2015. URL : <http://www.bjp-online.com/2015/05/jules-spinatsch-vienna-opera-ball-surveillance/> (consulté le 21 avril 2016)

La sérialité de ces deux projets reflète le fonctionnement des dispositifs de surveillance. D'abord à travers le répertoriage d'identités de cette collecte virtuelle de visages, puis dans cette vision englobante de l'espace de l'Opéra de Vienne, la quantité d'images témoigne de l'omniprésence et de la constance d'enregistrement de ces caméras. Ces séries cherchent ainsi à submerger et noyer le spectateur sous cette masse photographique, métaphore de cette surveillance oppressante.

Comme nous l'avons vu à travers ces précédents exemples, certains artistes utilisent des dispositifs de surveillance dans la réalisation de leurs projets. Mais que cherchent-ils en s'appropriant ces appareils techniques à priori non destinés à un usage artistique ?

### **3 – Appropriation de la surveillance**

#### **a – Utilisations de systèmes de surveillance dans l'art**

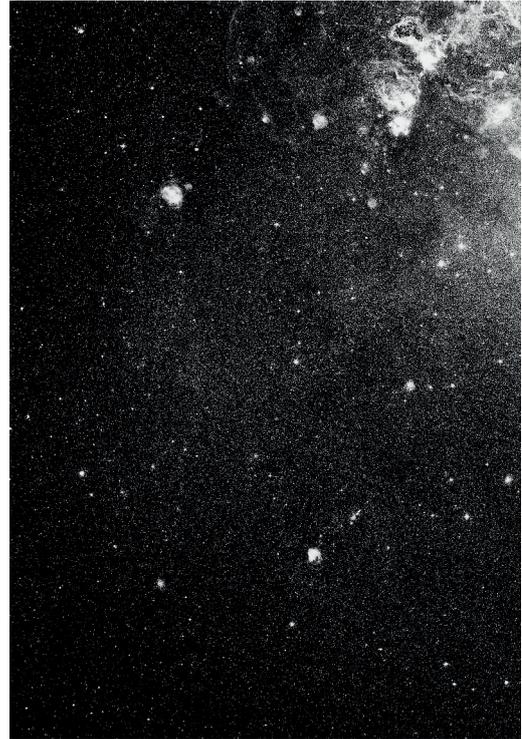
Ce réemploi de techniques de prise de vue initialement conçues pour un usage scientifique et/ou militaire n'est pas inédit dans l'histoire de la photographie et de l'art : il a même toujours existé ! En effet, comme l'explique Quentin Bajac dans *La photographie, du daguerréotype au numérique*, cette dernière est devenue « le principal instrument d'enregistrement du visible pour les cercles scientifiques »<sup>36</sup> dès les années 1880. Nous pouvons prendre l'exemple de la photographie astronomique, qui suscitait déjà une grande fascination lors des premières cartographies du ciel entreprises par l'Observatoire de Paris en 1889, et employée un siècle plus tard par le photographe allemand Thomas Ruff dans son projet artistique *Sterne*.

---

<sup>36</sup> Quentin Bajac in *La photographie, du daguerréotype au numérique*, Paris, Gallimard, 2010, p. 138-139



**Figure 21**, HENRY Paul, *Constellation du cygne*, photographie, tirage sur papier albuminé, 1885



**Figure 22**, RUFF Thomas, *Sterne*, photographie, 1989-92

Ces dispositifs photographiques servant initialement un but scientifique, permettant d'enregistrer « ce que nous ne pouvons pas voir naturellement ou correctement »<sup>37</sup>, ont ainsi été réemployés dans un autre contexte afin de servir un propos artistique.

L'utilisation par des artistes de ces systèmes de surveillance constitue une forme d'appropriation, dans la continuité des *Ready-made*<sup>38</sup> de Marcel Duchamp, et leur permet ainsi d'étendre leur champ d'expression. Comme nous le verrons par la suite, l'usage de ces caméras est généralement lié à une dimension imaginaire ou politique et reflète des images du monde dans lequel nous vivons.

---

37 Michel Frizot in *Nouvelle histoire de la photographie*, Paris, Adam Biro, 1994, p. 274

38 Définition du *Dictionnaire de l'art moderne et contemporain* : « Nom donné par Marcel Duchamp, à partir de 1915, aux objets « tout faits » qu'il choisit et signe depuis 1913 et dont le premier exemple est la roue de bicyclette fixée sur un tabouret. »

## La position du spectateur

Dan Graham est l'un des premiers à avoir employé des caméras de surveillance dans ses installations<sup>39</sup>. Pionniers de l'art vidéographique, il a conçu une série d'une dizaine d'installations en 1974 intitulée *Time Delay Room*. Dans l'une d'elle, comme le montre le schéma ci-dessous, deux caméras de surveillance filment les spectateurs se trouvant dans deux pièces ouvertes l'une sur l'autre. Dans chaque pièce sont disposés deux moniteurs : l'un diffuse les images de la caméra de l'autre pièce en direct ; l'autre diffuse les mêmes images avec un délai de 8 secondes. Le spectateur possède ainsi deux rôles : spectateur (il se voit ...) et acteur (... observé). À travers ses images en direct et du passé immédiat, l'artiste lui permet alors de se questionner sur sa propre perception et sa position vis-à-vis du dispositif.

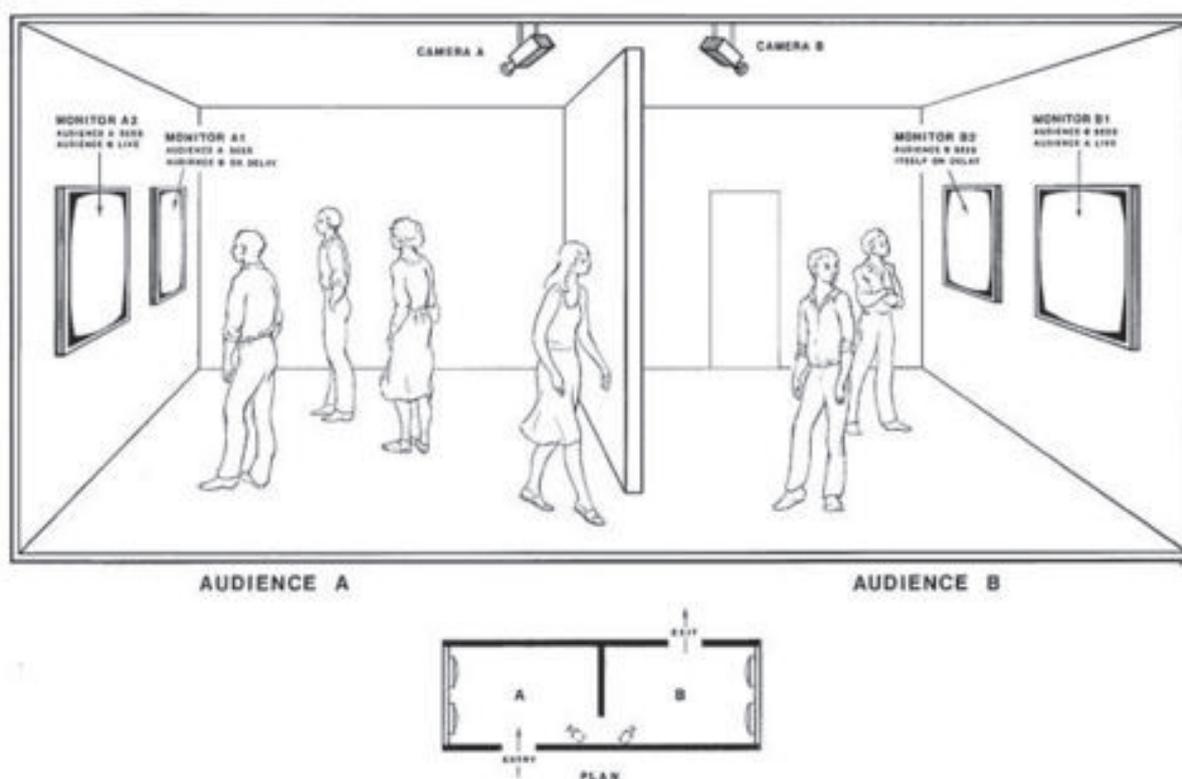


Figure 23, GRAHAM Dan, *Time Delay Room*, dessin, 1974

39 Michael Rush in *Video Art*, Londres, Thames & Hudson, 2003, pp. 27-30

### ***La position de l'auteur***

La série de panoramas *Surveillance Panorama Project*<sup>40</sup> de Jules Spinatsch, que nous avons évoquée dans la partie précédente, questionne la position de l'auteur dans sa démarche créative à travers l'appropriation de systèmes de vidéosurveillance. En effet, c'est grâce à un programme informatique pilotant les deux caméras de surveillance de l'opéra de Vienne que l'artiste a pu réaliser les dix-mille huit photographies constituant le panorama *No.4 Vienna MMIX*. Le cadrage de chaque image n'a pas directement été réalisé par l'artiste mais résulte donc de commandes automatisées liées à des calculs informatiques. Dans un premier temps, le rôle de l'auteur reste au stade de conception de son œuvre. C'est ensuite qu'il effectuera sa sélection d'images pour mettre en avant son propos. L'automatisme du procédé permet une plus grande précision dans la réalisation des images et délivre ainsi une certaine neutralité aux images en les plaçant toutes au même plan. Cette dépossession de la prise de vue de l'auteur interroge sa position au sein de sa démarche et plus largement la relation de l'homme par rapport à la machine et l'automatisation.

### ***L'acte de surveillance***

La vidéo de Walid Ra'ad intitulée *I only wish I could weep*<sup>41</sup> a été montée à partir de plans réalisés par un certain *opérateur #17*, agent des renseignements secrets libanais assigné à la surveillance de la Corniche, lieu de promenade de Beyrouth où se rencontrent des agents de services secrets étrangers. Seulement, durant plusieurs années, depuis son camion de surveillance, cet opérateur aurait préféré filmer le coucher du soleil plutôt que les cibles. Démasqué, il aurait tout de même pu récupérer les bandes vidéos avant de se faire renvoyer de son poste, afin de les confier à Walid Ra'ad pour monter et diffuser la vidéo. Le contraste entre la nature austère de la prise de vue et la poésie du sujet éphémère questionne l'acte même de surveillance en le tournant en dérision. À travers le contexte fictif de cette œuvre et en prenant cette caméra en main, l'artiste détourne le point de vue du surveillant en offrant une vision poétique de l'acte de surveillance.

---

40 Jules Spinatsch, *Surveillance Panorama Project*, [En ligne]. URL : <http://jules-spinatsch.ch/?p=76> (consulté le 21 avril 2016)

41 Le Bal in *Le Bal paper : Images à charge – La construction de la preuve par l'image*, Paris, Le Bal, 2015, p. 14



**Figure 24**, RA'AD Walid, *I only wish I could weep*, photogramme, 2002

En insérant ces dispositifs de surveillance dans le champ artistique, ces artistes remettent en cause l'utilisation de ces outils ainsi que nos rapports avec ceux-ci. Certains artistes poursuivent leurs expérimentations en tentant de créer leurs propres appareils de prise de vue omnivoyants.

### **b – Conception de dispositifs de prise de vue omnivoyants**

L'expérimentation de certains artistes est poussée encore plus loin en s'affranchissant des systèmes de surveillance déjà existants pour tenter de créer leurs propres dispositifs « omnivoyants ». Comme pour la vidéosurveillance publique, l'objectif de ces différents appareils est de tout voir. Ces artistes cherchent ainsi à couvrir visuellement au mieux l'espace, jusqu'à l'englober. Les différents appareils présentent ainsi des analogies avec les appareils de surveillance en exploitant certaines de leurs caractéristiques et en les complétant.

## ***Le mouvement***

La première caractéristique de couverture de l'espace est le mouvement de balayage. Pour *La Région centrale*, réalisé en 1971, Michael Snow a conçu avec l'aide d'un ingénieur un système piloté par un programme informatique, permettant une rotation de la caméra sur trois axes. L'équipe de tournage s'est installée en pleine nature québécoise afin de filmer le paysage dans tous les sens. Le film expérimental qui en résulte rend dans sa durée une vision à 360° de l'espace entourant la caméra, constituant une première forme de point de vue omnivoyant, bien que limité par le cadre.



**Figure 25,** WIELAND Joyce, *Michael Snow et le dispositif de La Région centrale*, photographie, 1971

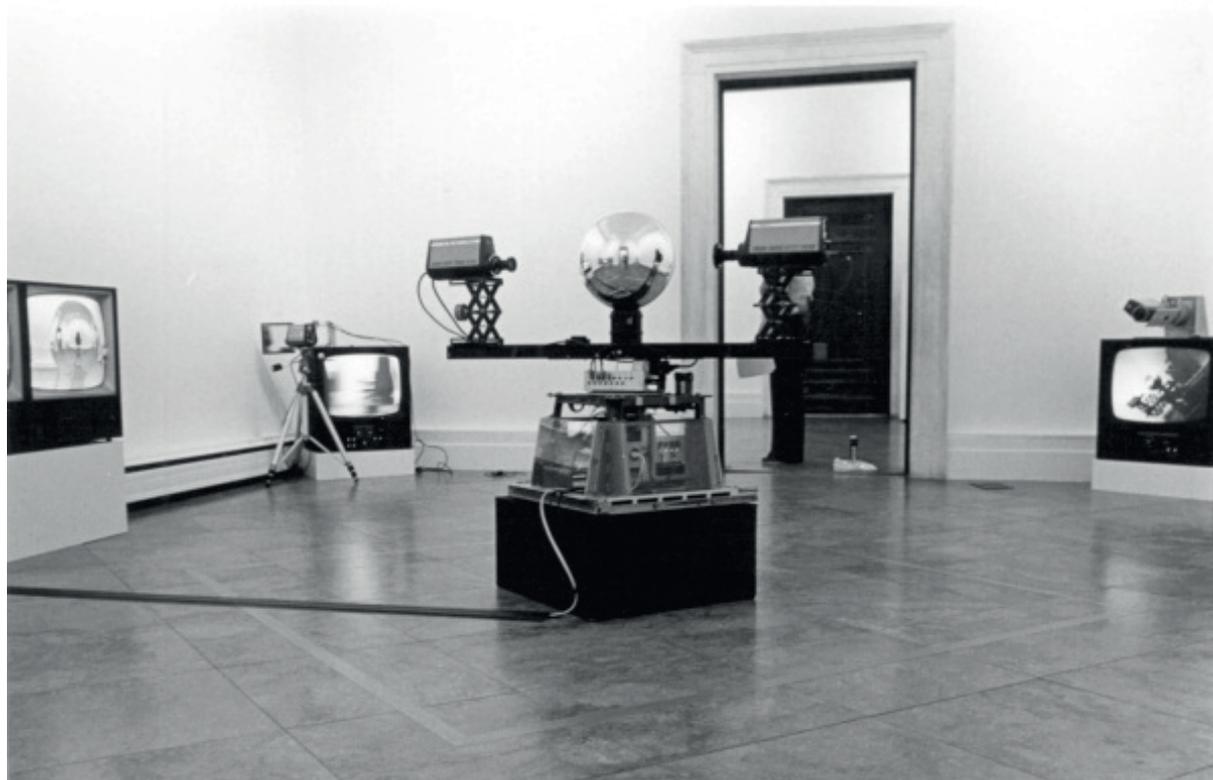
## ***La sphéricité***

À la contrainte du cadre restreignant ainsi la vision de l'observateur, Steina Vasulka a apporté la solution de la sphéricité dans l'une de ses installations intitulée *Allvision*<sup>42</sup>, faisant partie de sa série *Vision machine* de 1976. De chaque côté d'un

---

42 Françoise Parfait in *Video : un art contemporain*, Paris, Éditions du Regard, 2001, p. 279

miroir sphérique, l'artiste a disposé deux caméras vidéo sur un axe métallique en lente rotation. Chaque caméra filme dans la sphère l'une des deux moitiés du reflet de la pièce dans laquelle l'installation se trouve, et diffuse les images qui en résultent en temps réel sur deux moniteurs. L'espace est ainsi capté et retranscrit dans sa globalité.



**Figure 26**, VASULKA Steina, *Allvision*, photographie, 1976

Nous retrouvons une extension du schéma sphérique dans les photographies à effet « petite planète » (traduction de *little planet*) de Nicolas Burtey. Cet effet, expliqué dans son mémoire<sup>43</sup>, est réalisé à partir de plusieurs photographies réalisées depuis le même point de vue, assemblées pour constituer un panorama ensuite « projeté » à l'aide d'un logiciel de traitement d'image. La sphéricité de l'image qui en résulte donne une impression de paysage englobé, encore plus forte que pour des photographies panoramiques classiques : l'image semble tout montrer.

---

43 Nicolas Burtey in *La représentation de l'espace et du temps dans la photographie panoramique*, Paris, ENS Louis-Lumière, 2007



**Figure 27**, BURTEY Nicolas, *Petite planète*, photographie, 2007

De plus, l'arrivée sur le marché de la caméra *Kodak SP360*<sup>44</sup> ou du support *Omni* permettant l'assemblage de caméras de la marque *GoPro*<sup>45</sup>, a lancé une démocratisation de la prise de vue de photos et vidéos à 360°. Le véritable intérêt de ces images étant d'être affichées dans les casques de réalité virtuelle tels que l'*Oculus Rift*<sup>46</sup> afin de procurer aux spectateurs une sensation d'immersion alors toute nouvelle.

Mais ces techniques constituent-elles pour autant une forme d'objectivité ?

Dans les dispositifs de télésurveillance, et dans les *Vision machine* (1978) de Steina Vasulka, et *De La* (1969-1971), le dispositif de captation inventé par le cinéaste Michael Snow pour réaliser le film à effets centrifuges *La Région Centrale* (1969-1971), les caméras captent des images, mais ne « regardent » pas.<sup>47</sup>

44 Kodak, *SP360*, [En ligne]. URL : <http://kodakpixpro.com/Europe/fr/cameras/actioncam/sp360.php> (consulté le 24 avril 2016)

45 Go Pro, *Omni VR*, [En ligne]. URL : <http://vr.gopro.com/> (consulté le 24 avril 2016)

46 Oculus Rift, *Rift*, [En ligne]. URL : <https://www.oculus.com/en-us/rift/> (consulté le 24 avril 2016)

47 Nicolas Thély in *L'image sous le régime de la surveillance*, Paris, Université Paris 1 – Panthéon-Sorbonne, 2001, pp. 37-38

Nicolas Thély souligne très justement que cette quête d'omnivoyance entreprise avec ces différents dispositifs automatisés va à l'encontre de la notion de cadrage et par conséquent de choix, liée au regard.

De plus, lorsque Bernard Comment écrit que « tout point de vue est partiel et partial : ce qu'il offre à la visibilité implique aussi le choix de reléguer certains éléments à l'extérieur – soit ce qui est au-delà de la ligne d'horizon, soit ce qui, à l'intérieur de cet horizon, se trouve dérobé par tel ou tel élément qui ferait écran »<sup>48</sup>, il prouve une nouvelle fois que l'omnivoyance est bien une utopie. Par définition, une image est tout sauf objective puisqu'elle résulte forcément de choix (point de vue, cadrage, exposition, ...) de la part de son auteur.

L'affranchissement du dispositif de prise de vue par l'auteur ne permettrait-il pas alors l'émergence de nouveaux regards sur le sujet ?

### **c – Appropriations d'images de surveillance**

Certains artistes ont choisi de s'affranchir des moyens techniques de prises de vue de surveillance pour n'en exploiter que l'image après les avoir récupérés. Ils n'en sont donc pas auteurs mais le deviennent en se les appropriant. Cette action peut aussi bien passer par un traitement apposé sur l'image (cadrage, retouche, ...) que par une mise en relation avec d'autres images au sein d'une série ou d'un montage. L'image est leur matériau de création manipulable. Mais que peuvent apporter ces modes d'appropriation à leur démarche artistique ?

#### ***Suggérer une narration***

Considérée comme « proche du réel », le caractère neutre de la vidéosurveillance peut servir de matière première à la construction d'une narration. Le réalisateur allemand Michael Klier a réalisé *Der Riese* à partir de bandes de vidéosurveillances provenant de lieux banals (aéroport, rue, hôpital, ...). Sa seule intervention dans ce film est celle du montage. Les images sont reconnaissables par leurs caractéristiques déjà énoncées telles que le point de vue en plongée, les mouvements mécaniques ou encore le grand angle. Rien de spectaculaire à l'image : le point de vue est distant,

---

48 Bernard Comment in *Le 19ème siècle des panoramas*, Paris, Adam Biro, 1993, p. 58

les plans s'étendent sur la durée, aucune action n'est extra-ordinaire. Françoise Parfait qualifie même ces images de « sans qualités »<sup>49</sup>. Dans cet enchaînement de séquences du réel, non-mises en scène, le réalisateur suggère une narration que le spectateur interprète comme il le souhaite. En effet, dans un premier temps démunie, il est naturel pour lui de s'attacher au moindre élément reconnaissable afin que chacun fabrique sa propre histoire, en s'impliquant dans la constitution de l'œuvre.



**Figure 28**, KLIER Michael, *Der Riese*, photogramme, 1982-83

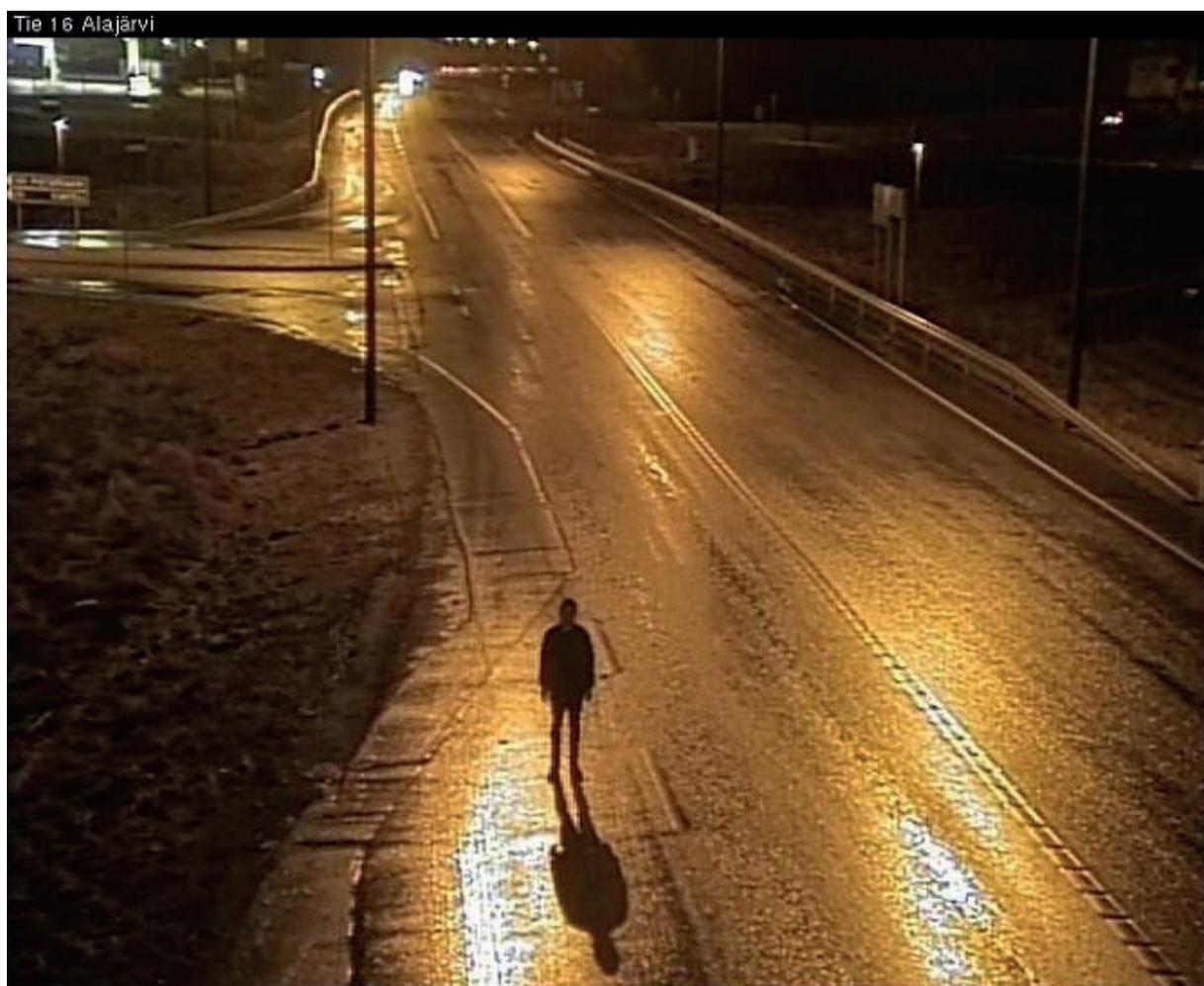
### ***Apporter de nouveaux regards***

Depuis l'explosion de l'Internet il y a un vingtaine d'années, la toile constitue un espace de création pour de nombreux artistes, grâce au flux continu de contenus, d'images et de techniques mis à disposition. Pour son projet intitulé *Weather Camera*

---

49 Françoise Parfait in *Video : un art contemporain*, op. cit., p. 281

*Selfportraits*<sup>50</sup>, l'artiste Tatu Gustafsson a réalisé une série d'autoportraits en utilisant des caméras de surveillance connectées à l'Internet. Il a ainsi sillonné la Finlande en ayant préalablement localisé ces caméras, généralement situées dans des lieux transitoires tels que des axes routiers ou des parkings, puis a posé devant pour enfin récupérer son portrait en ligne. Sa démarche peut être considérée comme hybride, se situant entre la photographie et la performance artistique : il se met en scène tout en s'appropriant les caméras trouvées afin de se photographier. En s'affichant ainsi face à elles et en récupérant son image, l'artiste tisse un lien entre l'espace public réel et celui virtuel de l'Internet. Ces images apportent un nouveau regard sur ce phénomène de *non-lieux*<sup>51</sup> et sur la relation qu'entretient l'homme avec son environnement.



**Figure 29**, GUSTAFSSON Tatu, *Weather Camera Selfportraits*, photographie, 2012

50 Tatu Gustafsson, *Weather Camera Selfportraits*, [En ligne]. URL : <http://www.tatugustafsson.com/Weather-Camera-Self-portraits-2012> (consulté le 20 avril 2016)

51 Dominique Baqué in *Photographie plasticienne, l'extrême contemporain*, Paris, Éditions du Regard, 2004, pp. 119-139

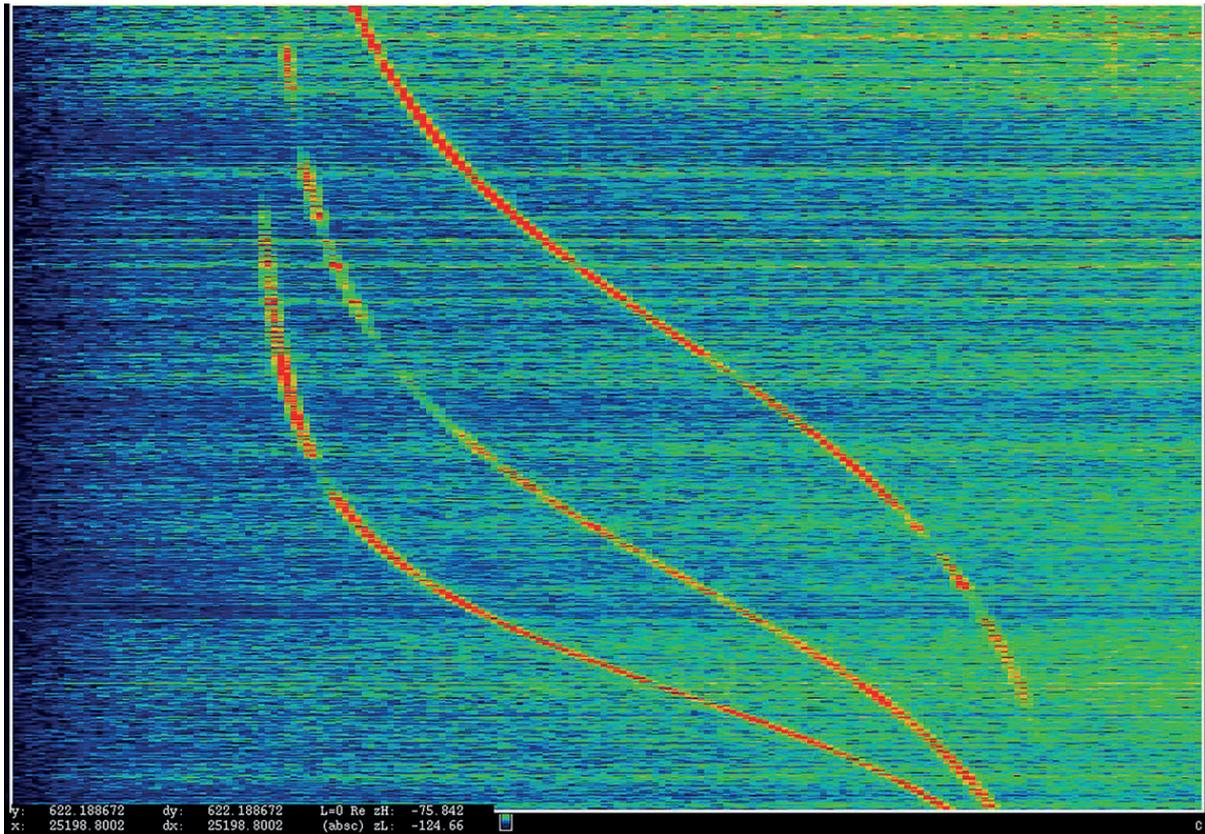
Par la nature de ces lieux banals représentés sur ces images provenant du réseau commun de l'Internet, leur appropriation par l'artiste renvoie à une proximité avec le spectateur, qui peut ainsi s'y projeter. Mais à l'inverse, l'appropriation et l'exhibition d'images extra-ordinaires par l'artiste peut constituer une rupture et renverser les rapports du pouvoir.

### ***Renverser la surveillance***

Le contrôle de la relation surveillants/surveillés est essentiel pour son fonctionnement. En effet, la surveillance est efficace tant que son mécanisme reste inaccessible de ceux sur lesquels elle s'exerce. La démarche de la réalisatrice et journaliste américaine Laura Poitras est justement de la mettre en lumière. Après son film *Citizenfour* sur la révélation de l'affaire des écoutes de la NSA (*National Security Agency*) par le lanceur d'alerte Edward Snowden, elle a constitué un projet intitulé *Astro Noise*<sup>52</sup> exposé en janvier 2016 au *Whitney Museum of American Art* de New-York, constitué à partir de documents visuels fournis par l'ex-collaborateur de la NSA. Ces fichiers images et graphiques sont issus d'opérations d'observation de drones militaires, de satellites commandées par les services de renseignements américains. Malgré le manque de lisibilité de ces images, leur rapprochement et l'explication de leur provenance témoignent du pouvoir dont ces organismes de surveillance disposent et l'exercent. En s'appropriant et en exposant ces documents confidentiels dans un musée, Laura Poitras dévoile au public les rouages de la mécanique d'une société de surveillance et amorce ainsi son renversement.

---

52 Whitney Museum of American Art, *Laura Poitras: Astro Noise*, [En ligne]. URL : <http://whitney.org/Exhibitions/LauraPoitras> (consulté le 6 mai 2016)



**Figure 30**, POITRAS Laura, *ANARCHIST : Data Feed with Doppler Tracks from a Satellite (Intercepted May 27, 2009)*, photographie, 2016

Très peu de temps après son émergence, la vidéosurveillance a donc été utilisée à des fins artistiques selon les différents modes d'appropriation détaillés aux travers des exemples analysés. De par leur point de vue et leurs automatismes, les caméras de surveillance constituent des outils singuliers dans leur captation de l'espace, en cherchant à l'englober. De même, la « banalité » et la richesse des textures des images qui en sont issues en font une matière première formidable, permettant la construction de projets allant de l'expérience au militantisme. L'image de vidéosurveillance ne laisse pas indifférent et suscite généralement chez le spectateur une peur fascinante. En s'en appropriant, ces artistes la dépossèdent du pouvoir et inversent les tendances.

En écho à leur forte présence dans l'actualité, la surveillance et la sécurité restent des thématiques qui imprègnent notre société. Les questionnements de ces sujets par les artistes au travers de leurs projets sont essentiels pour que les citoyens ne tombent pas dans une quelconque soumission vis-à-vis du pouvoir débouchant sur une dictature.

Les formes de surveillance évoluent de paire avec les progrès techniques, plus particulièrement depuis le développement et la démocratisation des utilisations de l'Internet. Avec ces progrès, notre production et consommation des images ont aussi évolué, impliquant de nouveaux rapports entre surveillants et surveillés et renversant la logique de fonctionnement, passant d'une surveillance par l'image à une image surveillée.

### **III – L’image surveillée**

#### **1 – L’acceptation de la surveillance**

##### **a – L’obsession sécuritaire**

La visibilité de la surveillance passe aujourd’hui essentiellement par les médias à travers l’actualité. Suite aux attentats terroristes ayant frappés Paris en 2015 et Bruxelles en 2016, les sites d’information et les journaux télévisés se sont beaucoup appuyés sur des images de vidéosurveillance, constituant des preuves visuelles de ces catastrophes. Par ailleurs, passée une certaine heure, la grande majorité des programmes télévisés est constituée de reportages à sensations ou de séries policières. Dans le contexte actuel, où la sécurité occupe une place importante dans les discours politiques, les citoyens et téléspectateurs baignent dans une forme d’obsession sécuritaire, à la fois fascinés et apeurés par les images qu’on leur propose.

Cette omniprésence de la surveillance à l’écran passe évidemment aussi par la fiction. En effet, dans de nombreux films et séries d’action ou policiers, cette vidéosurveillance est généralement utilisée comme un outil permettant l’avancée et éventuellement le dénouement de l’intrigue. Une sorte de solution miracle, bien souvent idéalisée par rapport à ses capacités réelles. Cette récupération fictionnelle participe à la fabrication d’un imaginaire collectif de la surveillance, composé par exemple des salles de contrôle équipées de murs de moniteurs de surveillance permettant aux vigiles de couvrir la ville entière ou encore de l’irréaliste grossissement de l’image permettant d’accéder à des détails invisibles à l’œil nu.



**Figure 31**, DE PALMA Brian, *Snake Eyes*, photogramme, 1998

Ces représentations fictionnelles de la surveillance alimentent des fantasmes d'omnivoyance, en décalage avec les réelles possibilités de cette dernière. Dans ce contexte, la frontière entre le fictif et le réel est assez floue, son utilisation fictionnelle ayant tendance à déteindre sur la perception de la réalité. Le risque de cette confusion est une mauvaise évaluation de la véritable fonction et des possibilités de la vidéosurveillance ainsi qu'une confiance trop importante accordée à ce système, renforçant par ailleurs l'obsession sécuritaire, forme de paranoïa injustifiée.

### **b – La caméra totem**

La visibilité de la surveillance passe aussi par la caméra en tant qu'objet. Grâce aux progrès techniques, les fabricants de caméras de surveillance annoncent de nouveaux modèles équipés de capteurs toujours plus sensibles, permettant l'enregistrement d'images dans des conditions lumineuses très faibles, et toujours plus résolus afin de percevoir encore plus de détails depuis le même point de vue, et ainsi réduire le nombre de caméras à installer. En 2016, *Sony* propose par exemple à la

vente la *SNC-VB770*, avec un capteur d'une très haute sensibilité de ISO 409 600, enregistrant des images en 4K (définition de 4096 x 2160) sous un éclairage de 0,004 lux minimum<sup>53</sup>. Les caméras de surveillance sont ainsi de plus en plus performantes pour un encombrement de plus en plus réduit, leur permettant de se fondre dans le paysage urbain.

Mais paradoxalement, de plus en plus de panneaux les annoncent aux entrées de magasins et aux abords des communes, mettant d'ailleurs en avant le terme de *vidéoprotection* plus rassurant. De plus, entre leur relais par les médias et leur utilisation fictionnelle, leurs images sont très visibles. Mais cette visibilité ambiguë ne dissimulerait-elle pas une autre utilité conférée à ces caméras ?

Les caméras de surveillance jouent évidemment un rôle à travers leur simple présence dans les espaces dans lesquelles elles sont installées. En effet, leur visibilité étant assurée par l'usage des panneaux d'avertissement, l'objet caméra, signifiant la présence d'une autorité, sert à dissuader ou à rassurer. Elles peuvent alors être assimilées à des totems, ces mats situés dans des espaces collectifs représentant des esprits protecteurs dans certaines communautés. Cette fonction de caméra-objet se révèle clairement dans l'usage de caméras factices, dont le fonctionnement repose seulement sur leur présence apparente dans l'environnement surveillé.

## **2 – La société exhibitionniste**

### **a – L'exhibition personnelle**

Cette nouvelle forme d'exhibition et de monstration à travers l'image s'est fortement développée à partir du début du XXIème siècle, notamment avec l'émergence de l'émission de télé-réalité *Big Brother* en Hollande dès 1999, s'étant ensuite exportée dans le monde entier comme *Loft Story* en France en 2001.

Le concept de ces émissions repose en effet sur l'exhibition du quotidien de « célébrités d'un jour », enfermées dans une maison dans laquelle une batterie de

---

53 Sony, *SNC-VB770*, [En ligne]. URL : <http://www.sony.fr/pro/product/video-security-ip-cameras-fixed/snc-vb770/overview/> (consulté le 3 mai 2016)

caméras est installée, filmant ainsi les moindres faits et gestes des participants. Un montage de tous ces enregistrements est ensuite diffusé quotidiennement, invitant ainsi les téléspectateurs à devenir les voyeurs de ces prisonniers volontaires. Il est d'ailleurs intéressant de se pencher sur le terme de *télé réalité* induisant une réalité transposée, voir complètement figurée par la télévision puisqu'il est bien connu que les agissements de chaque participant sont dictés par la boîte de production de l'émission. Les participants y jouent donc un rôle d'exhibition au sein d'un grand spectacle auquel le téléspectateur prend du plaisir à assister.

In the reality shows Big Brother in Germany, Loft in France, and Taxi Orange in Austria, staged by TV stations for the mass audience, the panoptic principle « everything must be seen and all must be shown, » is put into effect more than ever before as a model for becoming immunized to the society of the future. Observation is not a menace ; observation is entertaining.<sup>54</sup>

Peter Weibel rapproche ainsi ces émissions de télé réalité au système panoptique de Foucault. Cette nouvelle forme panoptique dans laquelle les prisonniers sont remplacés par les participants et les surveillants par les téléspectateurs, n'est plus utilisée dans un objectif de soumission à un pouvoir autoritaire mais de divertissement. L'historien d'art Jonathan Crary avance même dans son ouvrage intitulé *Suspensions of Perception*<sup>55</sup> que *la société du spectacle*<sup>56</sup> de Guy Debord et *la société de surveillance*<sup>57</sup> de Foucault ont ainsi fusionné.

Mais en donnant à voir la vie d'autres personnes aux téléspectateurs, ces programmes ayant vocation à les distraire, leur insufflent aussi un sentiment de pouvoir. Ces émissions ne serviraient-elles pas alors à détourner leur attention de leur propre soumission au système ?

---

54 Peter Weibel, « Pleasure and the Panoptic Principle » in *CTRL [space] : Rhetorics of surveillance from Bentham to Big Brother*, Karlsruhe, ZKM Center for Art and Media, 2002, p. 215

55 Jonathan Crary, *Suspensions of Perception : Attention, Spectacle, and Modern Culture*, Cambridge (Mass.), MIT Press, 2001, pp. 73-76

56 Guy Debord in *La Société du Spectacle*, Paris, Gallimard, 1967

57 Michel Foucault in *Surveiller et punir*, op. cit.



**Figure 32**, Endemol France, *Loft Story*, photogramme, 2001

À la même époque, avec la démocratisation de l'Internet et des webcams permettant le partage d'images réalisées chez soi en temps réel, une nouvelle forme d'exhibition à distance a vu le jour, cette fois-ci individualisée. Dans son article « Le règne de la délation optique », le philosophe Paul Virilio évoquait à propos des webcams « une nouvelle sorte de télé-vision, non plus chargée d'informer ou de divertir la masse des téléspectateurs, mais d'exposer, d'envahir l'espace domestique des particuliers »<sup>58</sup>. Les webcams constituent selon lui une évolution de la vidéosurveillance que nous développerons ensuite.

Ce désir d'exhibition par l'image a continué à évoluer et s'est largement démocratisé avec les applications telles qu'*Instagram* et *Snapchat*, permettant de (se) photographier/filmer avec son téléphone portable et de partager le contenu dans

---

58 Paul Virilio, « Le règne de la délation optique » in *Le Monde diplomatique*, Paris, août 1998, p. 20

la foulée ou même en direct avec *Periscope*. À l'instar de célébrités comme Kim Kardashian, popularisées grâce à leur utilisation parfois outrancière de ces réseaux, tout le monde s'expose à travers ses images partagées.



**Figure 33**, KARDASHIAN Kim, *Selfie*, photographie, 2014

Pour trop de gens, tout est désormais images, écrans ; y compris la vie des autres. Cela transforme une part grandissante de notre existence (ou de celle des autres) en télé-réalité permanente et familiarise la proximité de plus en plus fusionnée avec l'image électronique ; au point de l'accepter et d'intégrer l'omniprésence de l'objet-caméra et du reflet-écran.<sup>59</sup>

Comme l'écrit Claude-Marie Vadrot, cette nouvelle forme d'exhibition entraîne des changements dans nos rapports sociaux passant désormais souvent par l'image numérique. Cette exhibition personnelle par leur intermédiaire est volontaire et même désirée. Cette soumission au regard d'autrui répond et alimente une forme de voyeurisme. Ce couple voyeurisme-exhibitionnisme est lié à la notion de scopophilie définie par le psychanalyste Sigmund Freud, établissant l'existence d'une pulsion du regard pouvant être active dans le cas du voyeur ou passive dans le cas de l'exhibitionniste.<sup>60</sup> Il s'agit donc d'un besoin auquel le partage de son image et l'observation de celle des autres constitue l'assouvissement.

Mais la nature des images et la continuité de leur partage peut aussi constituer un support d'une nouvelle forme de surveillance invisible et par conséquent plus efficace.

## **b – L'image surveillée**

L'individu, tout à son plaisir d'évoluer dans un univers technologique, ne se préoccupe pas de savoir, et encore moins de comprendre, que des machines gèrent son quotidien. Que chacun de ses faits et gestes est enregistré, filtré, analysé, éventuellement surveillé. Que, loin de le libérer de ses entraves physiques, l'informatique communicante constitue sans doute le plus formidable outil de surveillance et de contrôle que l'homme ait jamais mis au point.<sup>61</sup>

Depuis la démocratisation de l'Internet et de notre utilisation effrénée, notamment au travers des réseaux sociaux, une nouvelle forme de surveillance est apparue. Cette surveillance numérique se base sur les traces digitales que nous échangeons ou que

---

59 Claude-Marie Vadrot in *La grande surveillance – Caméras, ADN, portables, Internet...*, op. cit., p. 29

60 Nicolas Thély in *L'image sous le régime de la surveillance*, op. cit., p. 88

61 Jean Guisnel, dans sa préface au livre de Reg Whitaker, *Tous fliqués ! La vie privée sous surveillance*, tr. fr. C. Frappier, Paris, Denoël, 2001, p. 11

nous laissons à chaque connexion, notamment avec les téléphones portables que nous portons sur nous en permanence. Les textes, images ou localisations géographiques constituent ainsi de véritables sources d'informations.

Les études disponibles sur les nouvelles pratiques communicationnelles témoignent d'une extension sans précédent de leurs applications utilitaires. En associant la dimension visuelle aux données échangées, l'image permet de fournir des indications de situation (arrivée ou présence dans un lieu, usage d'un moyen de transport ...), des vérifications d'apparence (essai d'un vêtement, résultat d'une coupe de cheveux, aspect physique...), mais aussi d'innombrables informations pratiques, comme l'achat d'une marchandise, la composition d'un plat, l'état d'un bâtiment, etc., que la photographie permet d'enregistrer ou de transmettre plus rapidement qu'un message écrit.<sup>62</sup>

André Gunthert relève ainsi le pouvoir informatif de l'image dans son utilisation socio-numérique. Lorsqu'elles sont associées à d'autres données (géolocalisation, identification de personnes ou de sociétés, *hashtags*, ...), ces dernières se révèlent en effet être très informatives à deux échelles de collecte : d'une part, par les autres usagers des réseaux sociaux tels que *Google*, *Facebook* ou *Instagram*, sur lesquels les images sont postées, mais aussi par ces mêmes réseaux.

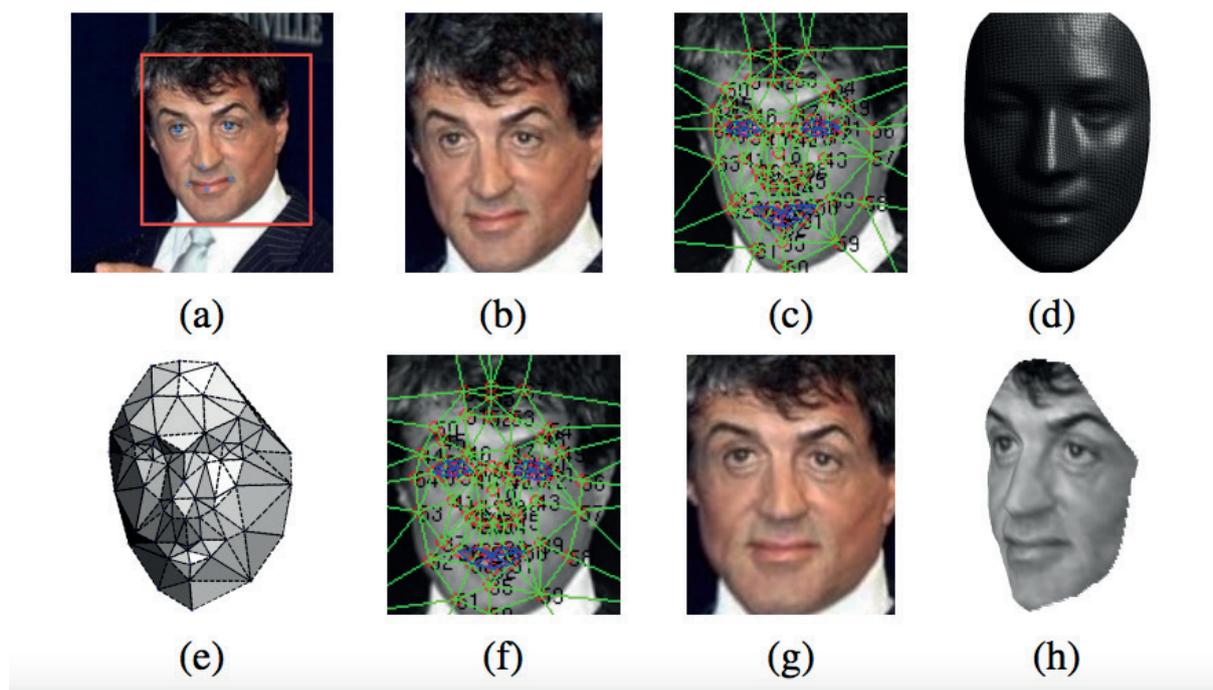
En plus d'accumuler des masses de données, ces sociétés conçoivent des programmes pour les traiter et les « comprendre ». Une équipe de chercheurs de *Facebook* a par exemple mis au point en 2014 l'algorithme de reconnaissance faciale *DeepFace* (expliqué en annexe) afin d'associer automatiquement le visage des personnes figurant sur les photos à leur compte pour les identifier automatiquement.<sup>63</sup> L'« automation de la perception »<sup>64</sup> présagée par Paul Virilio en 1988 semble être d'actualité.

---

62 André Gunthert in *L'image partagée. La photographie numérique*, Paris, Textuel, 2015, p. 144

63 Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf, « DeepFace : Closing the Gap to Human-Level Performance in Face Verification », *Facebook*, mis en ligne le 24 juin 2014. URL : <https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/> (consulté le 30 avril 2016)

64 Paul Virilio in *La machine de vision*, Paris, Galilée, 1988, p. 125



**Figure 34**, TAIGMAN Yaniv, YANG Ming, RANZATO Marc'Aurelio, WOLF Lior, *Schéma explicatif de l'algorithme DeepFace*, schéma, 2014

À travers cette collecte de données et ces algorithmes, l'objectif des géants de l'Internet est de nous connaître. En recoupant toutes ces informations, ils parviennent en effet à établir des profils associés à nos goûts et nos habitudes afin d'anticiper nos désirs et nous donner envie. L'utilisation de cette collecte est principalement commerciale, en proposant de la publicité de plus en plus ciblée, mais elle peut aussi servir à des fins sécuritaires en y laissant l'accès aux agences de renseignements comme l'explique Julian Assange, fondateur du site *WikiLeaks*<sup>65</sup> dans son ouvrage *When Google Met WikiLeaks*<sup>66</sup>.

Cette surveillance individualisée provient d'une évolution du modèle panoptique de Foucault, ayant ainsi conservé certaines de ses caractéristiques mais s'en démarquant singulièrement à d'autres niveaux.

65 « WikiLeaks est une organisation non-gouvernementale fondée par Julian Assange en 2006 dont l'objectif est de publier des documents ainsi que des analyses politiques et sociales à l'échelle du monde. Sa raison d'être est de donner une audience aux lanceurs d'alertes et aux fuites d'information, tout en protégeant ses sources. » (source : Wikipédia, *Wikileaks*, [En ligne]. URL : <https://fr.wikipedia.org/wiki/WikiLeaks> (consulté le 17 mai 2016))

66 Julian Assange in *When Google Met WikiLeaks*, New York, Or Books, 2014

## c – Une évolution du modèle panoptique

Nous pouvons considérer cette surveillance numérique à deux échelles comme une évolution du modèle panoptique, bien que cette dernière présente des divergences capitales. En effet, les rôles et les positions des surveillants et des surveillés clairement définis dans l'architecture panoptique carcérale, avec la tour centrale pour les premiers et le bâtiment circulaire pour les seconds, ne le sont plus dans cette nouvelle forme de surveillance entre usagers de réseaux sociaux.

Nous remarquons ici que les sites en question vont bien au-delà de la simple « évocation » du projet de panopticon et reconduisent littéralement l'assujettissement « panoptique » du prisonnier placé sous le regard d'un contrôle invisible, chaque internaute étant alors convié à occuper la position démiurgique du gardien-voyeur, tout en échappant au regard des personnes qu'il observe.<sup>67</sup>

Cette surveillance numérique s'étant largement diffusée à travers notre utilisation quotidienne des réseaux sociaux, les positions de voyants et de vus sont beaucoup plus dispersées et permutable. La logique du « être vu sans savoir qui voit » remplace le « ne pas savoir si je suis vu », instaurant une nouvelle omniprésence du regard.

Une autre divergence réside dans le sens de cette exhibition. Contrairement à l'architecture panoptique permettant au surveillant d'observer tous les prisonniers depuis un même point de vue, l'exhibition numérique expose une minorité à une majorité de personnes. Michel Foucault affiliait d'ailleurs cette exposition au domaine du spectacle. « Toute la vie des sociétés dans lesquelles règnent les conditions modernes de production s'annonce comme une immense accumulation de spectacles. »<sup>68</sup> écrivait déjà Guy Debors en 1967. Cette caractéristique semble symptomatique de ce phénomène d'exhibition.

Enfin, l'autre grande divergence par rapport au panoptique de Foucault est qu'à la différence des prisonniers enfermés contre leur gré, les internautes désirent leur

---

67 Jérôme Glicenstein, « Le Paysage panoptique d'Internet. Remarques à partir de Jeremy Bentham », in *Revue d'Esthétique* n°39, Paris, Jean-Michel Place, 2001, pp. 97-115

68 Guy Debord in *La Société du Spectacle*, op. cit., p. 15

propre exhibition.

Ce que nous ne pourrions accepter pour des prisonniers devient un but - et un spectacle - suivi par des foules entières. Désormais, ils s'agira d'exhiber plus la vie de la communauté et de faire spectacle de dysfonctionnements, qui dans les sociétés de type antérieur, auraient été soigneusement dissimulés.<sup>69</sup>

À travers cette exhibition personnelle par l'image, nous participons à notre propre surveillance et nous en devenons même responsables. Bernard Harcourt poursuit son raisonnement en expliquant dans un entretien radiophonique que « la chose que le modèle panoptique n'arrive pas à comprendre est que nous nous lançons dans ce monde numérique avec tout notre désir, nos fantasmes, toute notre jouissance. »<sup>70</sup> Obnubilés par notre quête du *like* et du *share*, nous prenons en effet plaisir à partager nos contenus et à alimenter cette collecte de renseignements. Ne s'agirait-il pas là du véritable danger de soumission de cette nouvelle forme de surveillance ?

### **3 – Quel rôle peut jouer la photographie ?**

#### **a – Lien étroit entre la surveillance et la photographie**

Photography was conceived around 1800, at about the same time then as Jeremy Bentham's panopticon. This, and the photograph's rapid incorporation into modernity's disciplinary apparatus, has led a number of scholars to examine how « photography served to introduce the panopticon principle into daily life »<sup>71</sup>. (...) So it's fair to say that a panoptic principle has always been inscribed at the very heart of photography's operation in our culture. (...) However a brief survey of photography's history reveals that this particular guilty pleasure has always been right there at the center of the medium, motivating a significant and surprisingly varied body of surveillance-type images.<sup>72</sup>

---

69 Jérôme Glicenstein, « Le Paysage panoptique d'Internet. Remarques à partir de Jeremy Bentham », in *Revue d'Esthétique* n°39, op. cit.

70 Bernard Harcourt, « Foucault, la société punitive et l'Amérique avec Bernard Harcourt », in *Hors-champs*, diffusé le 23 mai 2016, France culture

71 Allan Sekula, « Reading and archive: Photography between Labour and Capital », in *The Photography Reader*, Londres, Routledge, 2002, pp. 443-451

72 Geoffrey Batchen, « Guilty Pleasures » in *CTRL [space] : Rhetorics of surveillance from Bentham to Big Brother*, Karlsruhe, ZKM Center for Art and Media, 2002, pp. 447-448

Comme l'explique Geoffrey Batchen, la photographie et la surveillance ont toujours été intimement liées. En effet, la capture d'images à l'insu du sujet existe depuis l'invention de la photographie et a explosé à partir des années 1880 avec la réduction de l'encombrement des appareils photographiques et la sensibilisation des surfaces sensibles, permettant les prises de vues à main levée. Il est intéressant de noter l'apparition à la même époque des appareils dits *détectives*, grâce à leur viseur permettant à l'opérateur de ne pas donner l'impression de regarder la scène qu'il photographie et ne pas susciter d'attention particulière, renvoyant à l'imaginaire du détective.<sup>73</sup>

**MARION'S "PARCEL" DETECTIVE CAMERA (Patent).**



**FOR PLATES 4½ × 3½.**

**T**HIS Camera is made box-shape and neatly covered with brown linen-lined paper, and tied with string like an ordinary parcel, of which it has the exact appearance. The object is to disguise its real use, and to permit a Photograph to be taken without raising the slightest suspicion. It is so constructed that it is always in focus for an object 6 feet or more distant; after slipping the plates in from the holder, which may be done in daylight, it is only necessary to touch a spring to obtain a sharp Negative; the Plate is then returned to its holder. These Plate-holders are made of india-rubber, impervious to light and atmosphere, and, being flexible and thin, are light, and will pack in a very small compass. The Camera is fitted with a good double combination quick-acting Lens.

<b>PARCEL CAMERA, complete</b>	£3 10 0
Extra India-Rubber Plate-Holders	per dozen 1 0 0
Above Camera covered in Leather instead of Paper, with Leather Straps and Handle	4 4 0

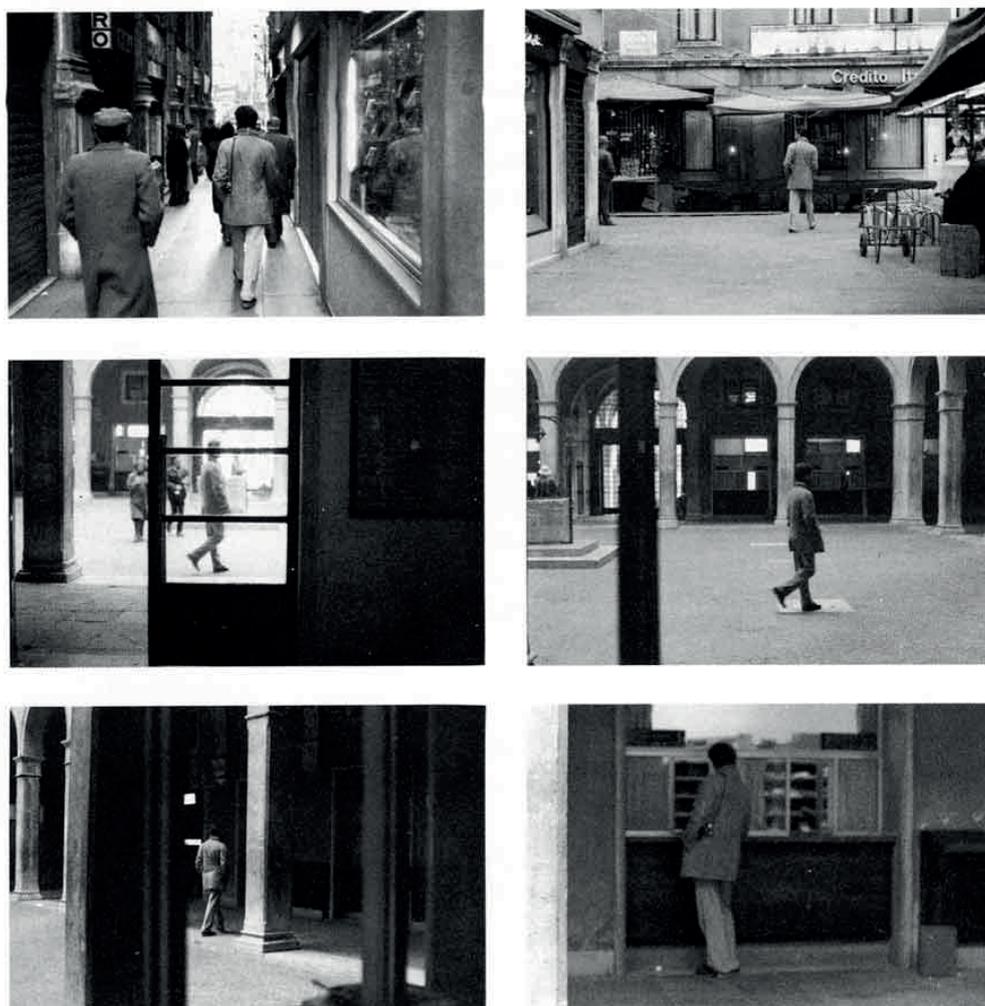
**MARION & CO., 22 & 23 Soho Square, London, W.**

Figure 35, Marion & Co., Marions's « Parcel » detective camera, annonce, 1890

<sup>73</sup> Quentin Bajac in *La photographie, du daguerréotype au numérique*, Paris, Gallimard, 2010, p. 131

Cette possibilité offerte par ces types d'appareils, d'observer et d'enregistrer l'image d'un autre sans être vu, a permis de nouveaux rapports entre photographes et sujets dans les images.

Cette même relation est questionnée dans la série photographique *Suite vénitienne*<sup>74</sup> de Sophie Calle. Après avoir rencontré un inconnu qu'elle suivait dans les rues de Paris, l'artiste plasticienne a décidé de le suivre à nouveau jusqu'à Venise et le photographier à son insu durant tout son séjour. À travers son point de vue distant, Sophie Calle capture discrètement les fragments de la vie d'un homme. Dans la continuité de cet espionnage, ses photographies ponctuent de manière régulière les actions de cet individu. Isolées, ces images semblent muettes. Mais une fois assemblées, elles reconstituent une histoire entre ces deux inconnus.



**Figure 36,** CALLE Sophie, *Suite vénitienne*, photographie, 1980

74 Sophie Calle, « Suite vénitienne » in *M'as-tu vue*, Paris, Centre Pompidou – Xavier Barral, 2003, pp. 85-95

Dans la continuité de cette démarche, Sophie Calle est invitée en 1991 par une banque américaine à la conception d'un projet à partir d'extraits de vidéosurveillances de distributeurs de billets. En 2003, l'artiste présente *Cash machine* et explique sa vision de ces images.

Observer à nouveau, sur l'écran, ces milliers de visages. Les photographier. Récapitulons : Si trois cassettes enregistrent chacune huit jours et autant de nuits, ça donne 108 000 images par heure x 24 heures x 8 jours x 3 cassettes. Environ 62 millions deux cent huit milles vues. Ça en représentait des photographies sans histoire. Je comprenais pourquoi cela faisait presque dix ans que je les regardais défiler.<sup>75</sup>



**Figure 37**, CALLE Sophie, *Cash machine*, photographie, 2003

Noyée par le flux continu et sans histoire de ces extraits, Sophie Calle a ainsi mis plus de dix ans à les appréhender, avant de les disséquer et les concevoir comme des suites de photographies. La faible fréquence d'images découpe et suspend la continuité

75 Sophie Calle, « Unfinished » in *M'as-tu vue*, op. cit., p. 419

des vidéos. L'œil, toujours à l'affût du fameux instant décisif, se fixe sur le moindre événement déclencheur se réduisant à une de ces photographies « insignifiantes », qui nous permettent « de voir des choses qui ont été photographiées presque accidentellement et au passage »<sup>76</sup>, comme l'explique Joanna Zylińska. À travers ces différents éléments, les images de vidéosurveillance ont une dimension photographique.

Par son lien étroit avec la surveillance, le médium photographique confirme la légitimité et la pertinence de son point de vue pour la questionner. Mais que peut-il apporter ?

## **b – Entre distanciation et fantasmes**

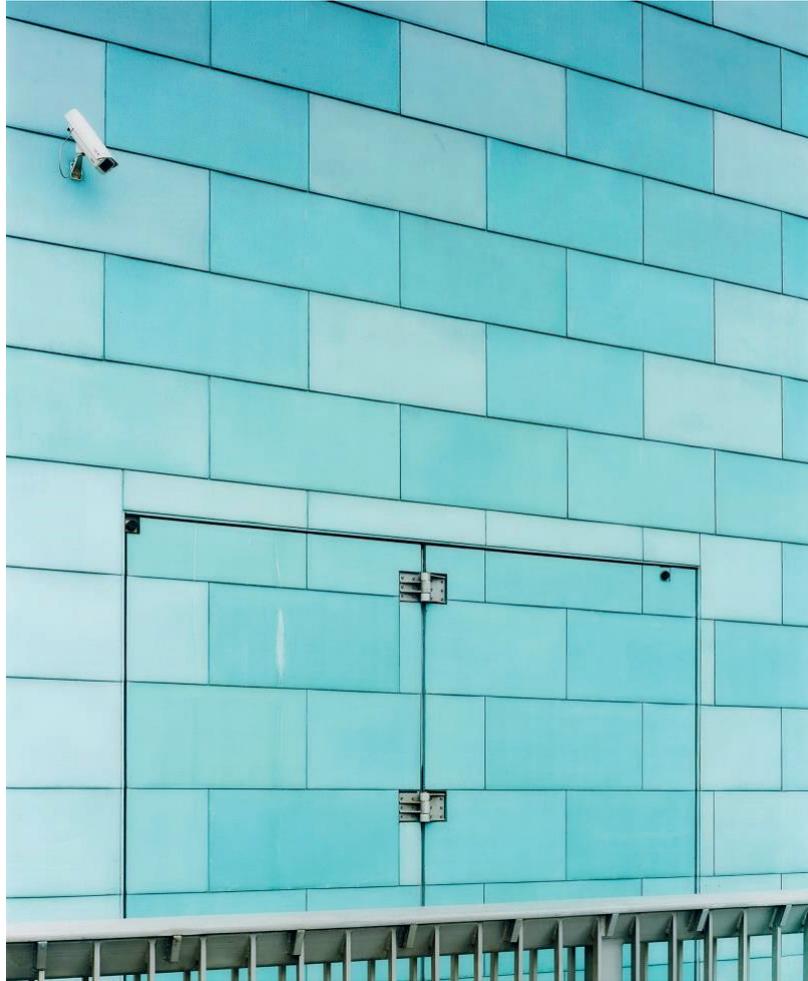
Par ses caractéristiques qui lui sont propres, la photographie peut apporter de nouveaux regards sur ces questions de surveillance. Contrairement aux projets précédemment étudiés, dans lesquels les artistes s'approprient la vidéosurveillance par les caméras ou les images, la photographie peut aussi permettre une prise de recul vis-à-vis de cette dernière, en s'en extrayant, rendant ainsi sa représentation possible comme objet.

Le projet *City TV*<sup>77</sup> du photographe allemand Frank Thiel constitue une documentation photographique de la présence des caméras de vidéosurveillance dans l'espace public. Réalisées à Berlin, ces photographies auraient pu être prises dans n'importe quelle capitale du monde entier, tant les logiques de fonctionnement des dispositifs se sont globalisées. La répétition des cadrages centrés sur les caméras intensifie la sensation de leur visibilité et de leur omniprésence. L'appareil photographique fonctionne comme un miroir renvoyant l'image de ces caméras. De plus, le point de vue de côté et la distance qui les sépare, laissent voir la place qu'elles occupent dans leur environnement et apportent une certaine stabilité aux images, une dimension descriptive.

---

76 Joanna Zylińska, « Le monde comme appareil photographique : notes sur une photographie non humaine » in *Drone : l'image automatisée*, Montréal, Le Mois de la Photo à Montréal – Kerber Verlag, 2013, p. 166

77 Frank Thiel, « City Tv », in *CTRL [space] : Rhetorics of surveillance from Bentham to Big Brother*, op. cit., pp. 136-139



**Figure 38**, THIEL Frank, *City TV*, photographie, 1997-99

À travers la distanciation permise par la photographie dans ce projet, une vision documentaire est portée sur la surveillance, amorçant une forme d'étude historique, d'inventaire des moyens mis en place dans l'environnement quotidien.

Mais la photographie permet aussi de se détacher du réel. En témoigne une image extraite du projet *I never had hair on my body or head*<sup>78</sup> du danois Niels Bonde. Dans cette installation recréant une chambre, se trouvent divers objets dont plusieurs peluches, sur lesquelles des caméras ont été implantées à la place d'un de leurs yeux. La photographie des trois jouets, visible à la page suivante, a été réalisée en parallèle de cette installation, les objectifs semblant avoir été ajoutés en post-production. La retouche grossière, à la limite de la peinture, détache cette photographie du réel. Par

---

78 Niels Bonde, « I never had hair on my body or head », in *CTRL [space] : Rhetorics of surveillance from Bentham to Big Brother*, op. cit., pp. 510-513

effet de contraste avec leur innocence originelle, cette modification absurde provoque un effet de suspicion immédiat chez le spectateur et renvoie à la société décrite par George Orwell où des caméras seraient même infiltrées dans des objets du quotidien pour surveiller les citoyens.<sup>79</sup>



**Figure 39**, BONDE Niels, *I never had hair on my body or head*, photographie, 1995-2001

Au-delà de représenter le réel, la photographie permet donc aussi de s'en affranchir en offrant une vision des peurs et fantasmes tirés d'imaginaires individuels ou collectifs. Les différentes possibilités offertes par le médium photographique constituent ainsi autant de pistes d'approche s'offrant à moi, en tant que photographe, dans la réalisation de ma partie pratique de mémoire.

---

79 George Orwell in 1984, op. cit.

## c – Partie pratique de mémoire

Dans le cadre de ma partie pratique de mémoire, le médium photographique m'est apparu comme évident afin d'appliquer et prolonger mon raisonnement entrepris tout au long de cette recherche. J'ai souhaité dans ce projet questionner les différentes formes que la surveillance a prise dans notre société. En effet, une première trame visible de cette surveillance est représentée dans notre quotidien par les caméras en tant qu'objet, les panneaux les annonçant ainsi que les images qui en sont issues et relayées par les médias. La visibilité constitue ainsi son principal atout auprès des citoyens. Une seconde trame, cette fois-ci virtuelle et très peu visible, est constituée par la collecte des traces digitales que nous laissons après utilisations de l'Internet. Cette surveillance individualisée que nous alimentons nous-même est par conséquent plus efficace que la première.

Mon projet s'est donc axé sur la mise en tension de ces deux principaux types de surveillance. Suite aux différents travaux photographiques de Frank Thiel ou Niels Bonde réalisés sur les dispositifs de surveillance et leur place dans notre société, j'ai souhaité poursuivre ce raisonnement en m'intéressant au pouvoir de l'objet-caméra et de ses paradoxes. J'ai ainsi réalisé une série photographique de dispositifs de dissuasion. Ces caméras de surveillance factices et ces panneaux d'avertissement sont vides de mécanisme : leur fonction est uniquement basée sur leur pouvoir de dissuasion transmis par le réalisme de leur apparence. Une fois installée, il est en effet très compliqué de différencier une caméra factice d'une opérationnelle. Ce réalisme est justement poussé jusqu'au clignotement de cette LED rouge signifiant le fonctionnement de la caméra. C'est précisément dans ce point rouge que le pouvoir de cette caméra se trouve canalisé, imposant ainsi à l'observé une domination du regard, véhiculée jusque dans l'image.

Nous ne pouvons dire tautologiquement *Je vois ce que je vois* que si nous dénonçons à l'image le pouvoir d'imposer sa visualité comme une ouverture, une perte – fût-elle momentanée – pratiquée dans l'espace de notre certitude visible à son égard.

Et c'est bien de là que l'image se rend capable de nous regarder.<sup>80</sup>

En analogie aux propos de Georges Didi-Huberman, l'image de ces faux dispositifs de surveillance s'impose à nous et nous regarde. Dans son ouvrage *La machine de vision*, Paul Virilio identifie ce phénomène comme une « inversion de la perception », qu'il associe particulièrement à la photographie publicitaire.

Si le cliché photographique publicitaire ébauche avec l'image phatique une inversion radicale des rapports de dépendance entre ce qui perçoit et ce qui est perçu, illustrant à merveille la phrase de Paul Klee, *maintenant les objets m'aperçoivent*, c'est qu'elle n'est déjà plus exactement une mémoire courte, le souvenir photographique d'un passé plus ou moins lointain, mais bien *une volonté*, la volonté, là aussi, d'engager l'avenir et non plus seulement de représenter le passé.<sup>81</sup>

La photographie instaure ainsi un rapport entre l'image et son spectateur. L'irréalité de l'image transmise par la lumière inquiétante, la présence de la LED rouge ainsi que l'importante dimension des tirages ne font qu'appuyer la domination de ce regard qui se pose sur le spectateur, élevant ces caméras au rang d'objets de culte, de totems.

Mais la nature factice des objets photographiés entraîne inévitablement une lecture en deux temps : d'abord une fascinante oppression suscitée par l'importante apparence du dispositif ; puis une décrédibilisation entraînée par la compréhension du mécanisme reposant sur son apparence et sa visibilité. Par cette appréhension du fonctionnement de cette forme de surveillance, l'émergence d'une réflexion chez le spectateur constitue ainsi l'objectif de ce projet.

---

80 Georges Didi-Huberman in *Ce que nous voyons, ce qui nous regarde*, Paris, Les éditions de minuit, 1992, pp. 75-76

81 Paul Virilio in *La machine de vision*, op. cit., p. 135

## Conclusion

Ce qui m’effraie, ce n’est pas l’arrivée d’une société de surveillance, mais que nous la vivions déjà sans nous en soucier.<sup>82</sup>

La surveillance est en effet omniprésente dans notre société. D’abord par la vidéosurveillance visible par l’intermédiaire des caméras disséminées dans l’espace public et privé ainsi que par leurs images relayées dans les médias. Malgré la fonction totémique de ces caméras, celles-ci se révèlent très peu efficaces dans leur fonctionnement : l’image seule ne suffit plus à assurer la surveillance. En témoigne l’émergence des caméras à reconnaissance faciale, l’image a maintenant besoin d’être analysée par des algorithmes pour être efficace. Une seconde forme de surveillance informatisée fonctionne ainsi à partir du contenu créé par notre utilisation de l’Internet, ayant ainsi renversé le modèle de surveillance par l’image par celui de l’image surveillée. Notre désir d’exhibition, notamment au travers des réseaux sociaux, constitue ainsi le moteur même de cette nouvelle forme de contrôle. Cette mutation des techniques de surveillance témoigne du remplacement du pouvoir par le regard par un pouvoir de l’informatique.

Il est cependant important de relativiser le portée de cette surveillance et de comprendre qu’elle comporte des limites. Basée sur des informations disséminées et plus ou moins volontairement partagées, elle n’établit qu’un profil partiel d’un individu, une image. L’être humain est beaucoup plus complexe et aucune forme de surveillance actuelle n’est capable de le cerner dans son intégralité.

Malgré ces limites certaines, il reste essentiel d’être conscient de sa présence et de ses potentiels dangers. La mise en lumière de cette surveillance constitue le cœur de la plupart de ces exemples de projets artistiques, à travers les différents moyens d’appropriations étudiés dans cette recherche.

N’est-ce pas d’ailleurs le rôle de l’art que de questionner les usages et les fantasmes de notre société en permettant l’expression de regards singuliers ? L’art

---

82 Zygmunt Bauman cité par Ignacio Ramonet in *L’Empire de la surveillance*, op. cit., p. 29

de la photographie doit nous apprendre à regarder : non pas en cherchant à nous conditionner, mais justement en nous poussant à réfléchir sur les images et leurs enjeux politiques, vis-à-vis des libertés et de l'émancipation de l'individu. La photographie cherche, entre autres, à nous ouvrir les yeux et éveiller nos consciences par rapport aux dangers des pouvoirs de notre société et chacun de ces artistes étudiés dans cette recherche le prouve à sa manière par l'intermédiaire de son œuvre. Au-delà de la question de la surveillance, ce mémoire constitue donc une réflexion globale sur nos rapports à l'image.

# Bibliographie

## Ouvrages

### *Dictionnaire :*

- DUROZOI Gérard, *Dictionnaire de l'art moderne et contemporain*, Paris, Hazan, 1992, 676 p.

### *Histoire de la photographie :*

- BAJAC Quentin, *La photographie, du daguerréotype au numérique*, Paris, Gallimard, 2010, 384 p.
- COMMENT Bernard, *Le 19ème siècle des panoramas*, Paris, Adam Biro, 1993, 127 p.
- FRIZOT Michel, *Nouvelle histoire de la photographie*, Paris, Adam Biro, 1994, 775 p.
- GRONERT Stefan, *L'école de photographie de Düsseldorf*, Paris, Hazan, 2009, 320 p.

### *Théorie de la photographie :*

- BAQUÉ Dominique, *Photographie plasticienne, l'extrême contemporain*, Paris, Éditions du Regard, 2004, 288 p.
- GUNTHERT André, *L'image partagée. La photographie numérique*, Paris, Textuel, 2015, 176 p.

### *Théorie de l'art :*

- CRARY Jonathan, *Suspensions of Perception : Attention, Spectacle, and Modern Culture*, Cambridge (Mass.), MIT Press, 2001, 397 p.
- DIDI-HUBERMAN Georges, *Ce que nous voyons, ce qui nous regarde*, Paris, Les éditions de minuit, 1992, 209 p.
- PARFAIT Françoise, *Video : un art contemporain*, Paris, Éditions du Regard, 2001, 368 p.
- RUSH Michael, *Video Art*, Londres, Thames & Hudson, 2003, 256 p.

#### *Catalogues d'exposition :*

- CALLE Sophie, *M'as-tu vue*, Paris, Centre Pompidou – Xavier Barral, 2003, 443 p.
- Le Bal, *Le Bal paper : Images à charge – La construction de la preuve par l'image*, Paris, Le Bal, 2015, 24 p.
- LEVIN Thomas Y., FROHNE Ursula, WEIBEL Peter, *CTRL [space] : Rhetorics of surveillance from Bentham to Big Brother*, Karlsruhe, ZKM Center for Art and Media, 2002, 655 p.
- WOMBELL Paul, *Drone : L'image automatisée*, Montréal, Le Mois de la Photo à Montréal – Kerber Verlag, 2013, 232 p.

#### *Essais philosophiques :*

- DEBORD Guy, *La Société du Spectacle*, Paris, Gallimard, 1967, 208 p.
- FOUCAULT Michel, *Surveiller et punir*, Paris, Gallimard, 1975, 360 p.
- LE BON Gustave, *Psychologie des foules*, Paris, Presses Universitaires de France, 1895, 130 p.
- VIRILIO Paul, *La machine de vision*, Paris, Galilée, 1988, 170 p.

#### *La surveillance dans notre société :*

- ASSANGE Julian, *When Google Met Wikileaks*, New York, Or Books, 2014, 223 p.
- RAMONET Ignacio, *L'Empire de la surveillance*, Paris, Galilée, 2015, 208 p.
- VADROT Claude-Marie, *La grande surveillance – Caméras, ADN, portables, Internet...*, Paris, Seuil, 2007, 252 p.

#### *Romans d'anticipation :*

- DAMASIO Alain, *La zone du dehors*, Paris, Cylibris, 1999, 494 p.
- ORWELL George, *1984*, Paris, Gallimard, 1950, 428 p.

#### *Mémoire et thèse :*

- BURTEY Nicolas, *La représentation de l'espace et du temps dans la photographie panoramique*, Mémoire de fin d'études en photographie, option traitement des images (sous la direction de M. MARTIN Pascal), Paris, École Nationale Supérieure Louis-Lumière, 2007, 120 p.

- THELY Nicolas, *L'image sous le régime de la surveillance*, Thèse de doctorat en Art et archéologie (sous la direction de Mme DUGUET Anne-Marie), Paris, Université Paris 1 – Panthéon-Sorbonne, 2001, 704 p.

### **Articles de revues**

#### *Études sociologiques :*

- ARMSTRONG Gary et NORRIS Clive, « CCTV and the social structuring of surveillance », in *Crime Prevention Studies*, vol. 10, New York, 1999
- GILL Martin et SPRIGGS Angela, « Assessing the impact of CCTV », in *Home Office Research Study*, n°292, Londres, 2005
- LE BLANC Noé, « Sous l'œil myope des caméras », in *Manière de voir – Le Monde diplomatique*, n°133, février-mars 2014, Paris, Le Monde diplomatique, pp. 48-51

#### *Théorie de l'image :*

- DUGUET Anne-Marie, « Dispositifs », in Vidéo, *Revue Communications*, n°48, Paris, Seuil, 1988
- BAUME Benoît, DELCOURT Maxime, FLITI Gwénaëlle, GUNTHERT André, HÉMON Jacques, KARSENTY Éric et ROCCO Hélène, « Images automatisées, la disparition des auteurs » in *Fisheye magazine*, n°16, janvier-février 2016, Paris, Fisheye, pp. 24-43
- GLICENSTEIN Jérôme, « Le Paysage panoptique d'Internet. Remarques à partir de Jeremy Bentham », in *Revue d'Esthétique*, n°39, Paris, Jean-Michel Place, 2001
- VIRILIO Paul, « Le règne de la délation optique » in *Le Monde diplomatique*, Paris, août 1998, p. 20

### **Reportage audiovisuel**

- COMBALBERT Nicolas, « Reconnaissance faciale : tous filmés, tous identifiés ? », in *Envoyé Spécial*, diffusé le 11 février 2016, France 2

## Entretien radiophonique

- HARCOURT Bernard, « Foucault, la société punitive et l'Amérique avec Bernard Harcourt », in *Hors-champs*, diffusé le 23 mai 2016, France culture

## Sources internet

### *Techniques :*

- TAIGMAN Yaniv, YANG Ming, RANZATO Marc'Aurelio, WOLF Lior, « DeepFace : Closing the Gap to Human-Level Performance in Face Verification », in *Facebook*, [En ligne], mis en ligne le 24 juin 2014. URL : <https://research.facebook.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/> (consulté le 30 avril 2016)

### *Protection contre la vidéosurveillance :*

- CNIL, « Vidéosurveillance / vidéoprotection : les bonnes pratiques pour des systèmes plus respectueux de la vie privée », [En ligne], mis en ligne le 21 juin 2012. URL : <http://www.cnil.fr/institution/actualite/article/article/videosurveillance-vidioprotection-les-bonnes-pratiques-pour-des-systemes-plus-respectueux-de/> (consulté le 07 mars 2016)
- Collectif « Souriez vous êtes filmés », [En ligne]. URL : <http://souriez.info/> (consulté le 07 mars 2016)
- « Quand la « vidéoprotection » remplace la « vidéosurveillance », in *Le Monde*, mis en ligne le 16 février 2010. URL : [http://www.lemonde.fr/societe/article/2010/02/16/quand-la-vidioprotection-remplace-la-videosurveillance\\_1306876\\_3224.html](http://www.lemonde.fr/societe/article/2010/02/16/quand-la-vidioprotection-remplace-la-videosurveillance_1306876_3224.html) (consulté le 8 mars 2016)

### *La vidéosurveillance dans l'art :*

- BARBIEUX Jean-Marc, « Surveillance : Dénier de faciès », in *Tracks, Arte*, [En ligne], mis en ligne le 13 octobre 2014. URL : <http://tracks.arte.tv/fr/surveillance-deni-de-facies/> (consulté le 01 novembre 2014)

- BENN Molly, « La surveillance, une autre réalité », in *Our Age Is Thirteen*, [En ligne], mis en ligne le 7 mars 2015. URL : <http://www.oai13.com/non-classe/la-surveillance-une-autre-realite/> (consulté le 10 mars 2015)
- SEYMOUR Tom, « A secret surveillance project of the Vienna Opera Ball », in *British Journal of Photography*, [En ligne], mis en ligne le 6 mai 2015. URL : <http://www.bjp-online.com/2015/05/jules-spinatsch-vienna-opera-ball-surveillance/> (consulté le 07 mars 2016)

## Table des illustrations

<b>Figure 1</b> , AVELINE Claude, <i>Vue et perspective de la ménagerie de Versailles</i> , gravure, 1638-1715 .....	12
<b>Figure 2</b> , BENTHAM Jeremy, <i>Plan schématique du Panopticon</i> , 1791.....	13
<b>Figure 3</b> , <i>Carte postale de l'intérieur du centre pénitencier de Stateville</i> , photographie, 1920.....	14
<b>Figure 4</b> , HAROU-ROMAIN N., <i>Projet de pénitencier</i> , gravure, 1840 .....	16
<b>Figure 5</b> , POUILLAIN Victor, <i>Panoptismes</i> , capture écran, 2015 .....	23
<b>Figure 6</b> , COMBALBERT Nicolas, <i>Reconnaissance faciale : tous filmés, tous identifiés ?</i> , capture d'écran, 2016.....	26
<b>Figure 7</b> , HARVEY Adam, <i>Stealth wear</i> , photographies, 2013 .....	32
<b>Figure 8</b> , HARVEY Adam, <i>CV Dazzle</i> , photographies, 2010.....	33
<b>Figure 9</b> , SELVAGGIO Leonardo, <i>URME</i> , photographies, 2014 .....	34
<b>Figure 10</b> , Police belge, <i>Vidéosurveillance de l'aéroport de Zaventem lors des attentats de Bruxelles du 22 mars 2016</i> , photogramme, 2016 .....	37
<b>Figure 11</b> , Daily Mail, <i>Vidéosurveillance du restaurant Casa Nostra lors des attentats de Paris du 13 novembre 2015</i> , photogramme, 2015 .....	38
<b>Figure 12</b> , TEPCO, <i>Webcam diffusant en direct depuis la centrale nucléaire de Fukushima</i> , photogramme, 2013.....	39
<b>Figure 13</b> , KADYROVA Zhanna, <i>Invisible Forms, Security Cameras</i> , photographie, 2015.....	42
<b>Figure 14</b> , BALTZ Lewis, <i>Sites of technology</i> , photographie, 1989-1991 .....	43
<b>Figure 15</b> , VAN HOUTRYVE Tomas, <i>Blue Sky Days</i> , photographie, 2015.....	44
<b>Figure 16</b> , MAYRIT Daniel, <i>They haven't seen their faces</i> , photographie, 2015 .....	46
<b>Figure 17</b> , RUFF Thomas, <i>Nacht</i> , photographie, 1992.....	48
<b>Figure 18</b> , Broomberg & Chanarin, <i>Spirit is a bone</i> , photographie, 2014 .....	49
<b>Figure 19</b> , VERDIJK Lukas et FAASSEN Vesna, <i>Vue d'exposition de Spirit is a bone</i> , photographie, 2014.....	50
<b>Figure 20</b> , SPINATSCH Jules, <i>Détails de Surveillance Panorama Project – No.4 Vienna MMIX</i> , photographie, 2009 .....	51
<b>Figure 21</b> , HENRY Paul, <i>Constellation du cygne</i> , photographie, tirage sur papier albuminé, 1885 .....	53
<b>Figure 22</b> , RUFF Thomas, <i>Sterne</i> , photographie, 1989-92 .....	53
<b>Figure 23</b> , GRAHAM Dan, <i>Time Delay Room</i> , dessin, 1974 .....	54
<b>Figure 24</b> , RA'AD Walid, <i>I only wish I could weep</i> , photogramme, 2002.....	56
<b>Figure 25</b> , WIELAND Joyce, <i>Michael Snow et le dispositif de La Région centrale</i> , photographie, 1971.....	57

<b>Figure 26</b> , VASULKA Steina, <i>Allvision</i> , photographie, 1976 .....	58
<b>Figure 27</b> , BURTEY Nicolas, <i>Petite planète</i> , photographie, 2007 .....	59
<b>Figure 28</b> , KLIER Michael, <i>Der Riese</i> , photogramme, 1982-83 .....	61
<b>Figure 29</b> , GUSTAFSSON Tatu, <i>Weather Camera Selfportraits</i> , photographie, 2012.....	62
<b>Figure 30</b> , POITRAS Laura, <i>ANARCHIST : Data Feed with Doppler Tracks from a Satellite (Intercepted May 27, 2009)</i> , photographie, 2016 .....	64
<b>Figure 31</b> , DE PALMA Brian, <i>Snake Eyes</i> , photogramme, 1998 .....	67
<b>Figure 32</b> , Endemol France, <i>Loft Story</i> , photogramme, 2001.....	70
<b>Figure 33</b> , KARDASHIAN Kim, <i>Selfie</i> , photographie, 2014 .....	71
<b>Figure 34</b> , TAIGMAN Yaniv, YANG Ming, RANZATO Marc'Aurelio, WOLF Lior, <i>Schéma explicatif de l'algorithme DeepFace</i> , schéma, 2014 .....	74
<b>Figure 35</b> , Marion & Co., <i>Marions's « Parcel » detective camera</i> , annonce, 1890...	77
<b>Figure 36</b> , CALLE Sophie, <i>Suite vénitienne</i> , photographie, 1980 .....	78
<b>Figure 37</b> , CALLE Sophie, <i>Cash machine</i> , photographie, 2003 .....	79
<b>Figure 38</b> , THIEL Frank, <i>City TV</i> , photographie, 1997-99 .....	81
<b>Figure 39</b> , BONDE Niels, <i>I never had hair on my body or head</i> , photographie, 1995- 2001.....	82

## Table des annexes

<b>Annexe 1, <i>Surveillance Panorama Project – No.4 Vienna MMIX</i>, Jules Spinatsch..</b>	<b>95</b>
<b>Annexe 2, <i>DeepFace : Closing the Gap to Human-Level Performance in Face Verification</i>, Yaniv Taigman, Ming Yang, Marc’Aurelio, Lior Wolf.....</b>	<b>97</b>
<b>Annexe 3, <i>Totems - Partie pratique de mémoire</i>, Victor Poullain .....</b>	<b>105</b>

**Annexe 1, *Surveillance Panorama Project – No.4 Vienna MMIX*, Jules Spinatsch**



SPINATSCH Jules, *Surveillance Panorama Project – No.4 Vienna MMIX (partie 1/2)*,  
photographie, 2009



SPINATSCH Jules, *Surveillance Panorama Project – No.4 Vienna MMIX (partie 2/2)*,  
photographie, 2009

# Annexe 2, DeepFace : Closing the Gap to Human-Level Performance in Face Verification, Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf

## DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman

Ming Yang

Marc'Aurelio Ranzato

Lior Wolf

Facebook AI Research  
Menlo Park, CA, USA

{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University  
Tel Aviv, Israel

wolf@cs.tau.ac.il

### Abstract

*In modern face recognition, the conventional pipeline consists of four stages: detect  $\Rightarrow$  align  $\Rightarrow$  represent  $\Rightarrow$  classify. We revisit both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Our method reaches an accuracy of 97.35% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 27%, closely approaching human-level performance.*

### 1. Introduction

Face recognition in unconstrained images is at the forefront of the algorithmic perception revolution. The social and cultural implications of face recognition technologies are far reaching, yet the current performance gap in this domain between machines and the human visual system serves as a buffer from having to deal with these implications.

We present a system (*DeepFace*) that has closed the majority of the remaining gap in the most popular benchmark in unconstrained face recognition, and is now at the brink of human level accuracy. It is trained on a large dataset of faces acquired from a population vastly different than the one used to construct the evaluation benchmarks, and it is able to outperform existing systems with only very minimal adaptation. Moreover, the system produces an extremely compact face representation, in sheer contrast to the shift

toward tens of thousands of appearance features in other recent systems [5, 7, 2].

The proposed system differs from the majority of contributions in the field in that it uses the deep learning (DL) framework [3, 21] in lieu of well engineered features. DL is especially suitable for dealing with large training sets, with many recent successes in diverse domains such as vision, speech and language modeling. Specifically with faces, the success of the learned net in capturing facial appearance in a robust manner is highly dependent on a very rapid 3D alignment step. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level. It is therefore possible to learn from the raw pixel RGB values, without any need to apply several layers of convolutions as is done in many other networks [19, 21].

In summary, we make the following contributions : (i) The development of an effective deep neural net (DNN) architecture and learning method that leverage a very large labeled dataset of faces in order to obtain a face representation that generalizes well to other datasets; (ii) An effective facial alignment system based on explicit 3D modeling of faces; and (iii) Advance the state of the art significantly in (1) the Labeled Faces in the Wild benchmark (*LFW*) [18], reaching near human-performance; and (2) the YouTube Faces dataset (*YTF*) [30], decreasing the error rate there by more than 50%.

#### 1.1. Related Work

**Big data and deep learning** In recent years, a large number of photos have been crawled by search engines, and uploaded to social networks, which include a variety of unconstrained material, such as objects, faces and scenes.

This large volume of data and the increase in computational resources have enabled the use of more powerful statistical models. These models have drastically improved the robustness of vision systems to several important variations, such as non-rigid deformations, clutter, occlusion and illumination, all problems that are at the core of many computer vision applications. While conventional machine

learning methods such as Support Vector Machines, Principal Component Analysis and Linear Discriminant Analysis, have limited capacity to leverage large volumes of data, deep neural networks have shown better scaling properties.

Recently, there has been a surge of interest in neural networks [19, 21]. In particular, deep and large networks have exhibited impressive results once: (1) they have been applied to large amounts of training data and (2) scalable computation resources such as thousands of CPU cores [11] and/or GPU’s [19] have become available. Most notably, Krizhevsky et al. [19] showed that very large and deep convolutional networks [21] trained by standard back-propagation [25] can achieve excellent recognition accuracy when trained on a large dataset.

**Face recognition state of the art** Face recognition error rates have decreased over the last twenty years by three orders of magnitude [12] when recognizing frontal faces in still images taken in consistently controlled (constrained) environments. Many vendors deploy sophisticated systems for the application of border-control and smart biometric identification. However, these systems have shown to be sensitive to various factors, such as lighting, expression, occlusion and aging, that substantially deteriorate their performance in recognizing people in such unconstrained settings.

Most current face verification methods use hand-crafted features. Moreover, these features are often combined to improve performance, even in the earliest *LFW* contributions. The systems that currently lead the performance charts employ tens of thousands of image descriptors [5, 7, 2]. In contrast, our method is applied directly to RGB pixel values, producing a very compact yet sparse descriptor.

Deep neural nets have also been applied in the past to face detection [24], face alignment [27] and face verification [8, 16]. In the unconstrained domain, Huang et al. [16] used as input LBP features and they showed improvement when combining with traditional methods. In our method we use raw images as our underlying representation, and to emphasize the contribution of our work, we avoid combining our features with engineered descriptors. We also provide a new architecture, that pushes further the limit of what is achievable with these networks by incorporating 3D alignment, customizing the architecture for aligned inputs, scaling the network by almost two order of magnitudes and demonstrating a simple knowledge transfer method once the network has been trained on a very large labeled dataset.

Metric learning methods are used heavily in face verification, often coupled with task-specific objectives [26, 29, 6]. Currently, the most successful system that uses a large data set of labeled faces [5] employs a clever transfer learning technique which adapts a Joint Bayesian model [6] learned on a dataset containing 99,773 images from 2,995 different subjects, to the *LFW* image domain. Here, in order

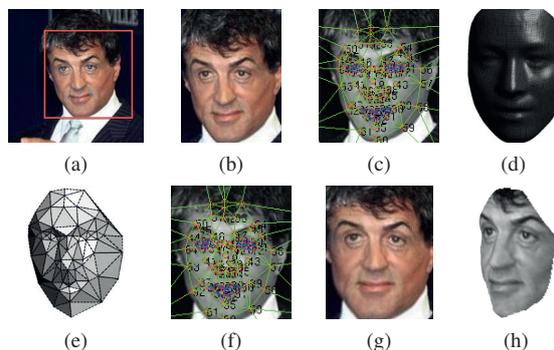


Figure 1. **Alignment pipeline.** (a) The detected face, with 6 initial fiducial points. (b) The induced 2D-aligned crop. (c) 67 fiducial points on the 2D-aligned crop with their corresponding Delaunay triangulation, we added triangles on the contour to avoid discontinuities. (d) The reference 3D shape transformed to the 2D-aligned crop image-plane. (e) Triangle visibility w.r.t. to the fitted 3D-2D camera; darker triangles are less visible. (f) The 67 fiducial points induced by the 3D model that are used to direct the piece-wise affine warping. (g) The final frontalized crop. (h) A new view generated by the 3D model (not used in this paper).

to demonstrate the effectiveness of the features, we keep the distance learning step trivial.

## 2. Face Alignment

Existing aligned versions of several face databases (*e.g.* *LFW*-a [29]) help to improve recognition algorithms by providing a normalized input [26]. However, aligning faces in the unconstrained scenario is still considered a difficult problem that has to account for many factors such as pose (due to the non-planarity of the face) and non-rigid expressions, which are hard to decouple from the identity-bearing facial morphology. Recent methods have shown successful ways that compensate for these difficulties by using sophisticated alignment techniques. These methods can use one or more from the following: (1) employing an analytical 3D model of the face [28, 32, 14], (2) searching for similar fiducial-points configurations from an external dataset to infer from [4], and (3) unsupervised methods that find a similarity transformation for the pixels [17, 15].

While alignment is widely employed, no complete physically correct solution is currently present in the context of unconstrained face verification. 3D models have fallen out of favor in recent years, especially in unconstrained environments. However, since faces are 3D objects, done correctly, we believe that it is the right way. In this paper, we describe a system that includes analytical 3D modeling of the face based on fiducial points, that is used to warp a detected facial crop to a 3D frontal mode (*frontalization*).

Similar to much of the recent alignment literature, our alignment is based on using fiducial point detectors to direct the alignment process. We use a relatively simple fiducial

point detector, but apply it in several iterations to refine its output. At each iteration, fiducial points are extracted by a Support Vector Regressor (SVR) trained to predict point configurations from an image descriptor. Our image descriptor is based on LBP Histograms [1], but other features can also be considered. By transforming the image using the induced similarity matrix  $T$  to a new image, we can run the fiducial detector again on a new feature space and refine the localization.

**2D Alignment** We start our alignment process by detecting 6 fiducial points inside the detection crop, centered at the center of the eyes, tip of the nose and mouth locations as illustrated in Fig. 1(a). They are used to approximately scale, rotate and translate the image into six anchor locations by fitting  $T_{2d}^j := (s_i, R_i, t_i)$  where:  $x_{anchor}^j := s_i[R_i|t_i] * x_{source}^j$  for points  $j = 1..6$  and iterate on the new warped image until there is no substantial change, eventually composing the final 2D similarity transformation:  $T_{2d} := T_{2d}^1 * \dots * T_{2d}^k$ . This aggregated transformation generates a 2D aligned crop, as shown in Fig. 1(b). This alignment method is similar to the one employed in LFW-a, which has been used frequently to boost recognition accuracy. However, similarity transformation fails to compensate for out-of-plane rotation, which is particularly important in unconstrained conditions.

**3D Alignment** In order to align faces undergoing out-of-plane rotations, we use a generic 3D shape model and register a 3D affine camera, which are used to warp the 2D-aligned crop to the image plane of the 3D shape. This generates the 3D-aligned version of the crop as illustrated in Fig. 1(g). This is achieved by localizing additional 67 fiducial points  $x_{2d}$  in the 2D-aligned crop (see Fig. 1(c)), using a second SVR. As a 3D generic shape model, we simply take the average of the 3D scans from the USF Human-ID database, which were post-processed to be represented as aligned vertices  $v_i = (x_i, y_i, z_i)_{i=1}^n$ . We manually place 67 anchor points on the 3D shape, and in this way achieve full correspondence between the 67 detected fiducial points and their 3D references. An affine 3D-to-2D camera  $P$  is then fitted using the generalized least squares solution to the linear system  $x_{2d} = X_{3d}\vec{P}$  with a known covariance matrix  $\Sigma$ , that is,  $\vec{P}$  that minimizes the following loss:  $loss(\vec{P}) = r^T \Sigma^{-1} r$  where  $r = (x_{2d} - X_{3d}\vec{P})$  is the residual vector and  $X_{3d}$  is a  $(67 * 2) \times 8$  matrix composed by stacking the  $(2 \times 8)$  matrices  $[x_{3d}^T(i), 1, \vec{0}; \vec{0}, x_{3d}^T(i), 1]$ , with  $\vec{0}$  denoting a row vector of four zeros, for each reference fiducial point  $x_{3d}(i)$ . The affine camera  $P$  of size  $2 \times 4$  is represented by the vector of 8 unknowns  $\vec{P}$ . The loss can be minimized using the Cholesky decomposition of  $\Sigma$ , that transforms the problem into ordinary least squares. Since, for example, detected points on the contour of the face tend to be more noisy, as their estimated location is largely influenced by the depth with respect to the camera angle, we

use a  $(67 * 2) \times (67 * 2)$  covariance matrix  $\Sigma$  given by the estimated covariances of the fiducial point errors.

**Frontalization** Since full perspective projections and non-rigid deformations are not modeled, the fitted camera  $P$  is only an approximation. In order to reduce the corruption of such important identity-bearing factors to the final warping, we add the corresponding residuals in  $r$  to the x-y components of each reference fiducial point  $x_{3d}$ , we denote this as  $\widetilde{x}_{3d}$ . Such a relaxation is plausible for the purpose of warping the 2D image with smaller distortions to the identity. Without it, faces would have been warped into the same shape in 3D, losing important discriminative factors. Finally, the frontalization is achieved by a piece-wise affine transformation  $T$  from  $x_{2d}$  (source) to  $\widetilde{x}_{3d}$  (target), directed by the Delaunay triangulation derived from the 67 fiducial points<sup>1</sup>. Also, invisible triangles w.r.t. to camera  $P$ , can be replaced using image blending with their symmetrical counterparts.

### 3. Representation

In recent years, the computer vision literature has attracted many research efforts in descriptor engineering. Such descriptors when applied to face-recognition, mostly use the same operator to all locations in the facial image. Recently, as more data has become available, learning-based methods have started to outperform engineered features, because they can discover and optimize features for the specific task at hand [19]. Here, we learn a generic representation of facial images through a large deep network.

**DNN Architecture and Training** We train our DNN on a multi-class face recognition task, namely to classify the identity of a face image. The overall architecture is shown in Fig. 2. A 3D-aligned 3-channels (RGB) face image of size 152 by 152 pixels is given to a convolutional layer (C1) with 32 filters of size 11x11x3 (we denote this by 32x11x11x3@152x152). The resulting 32 feature maps are then fed to a max-pooling layer (M2) which takes the max over 3x3 spatial neighborhoods with a stride of 2, separately for each channel. This is followed by another convolutional layer (C3) that has 16 filters of size 9x9x16. The purpose of these three layers is to extract low-level features, like simple edges and texture. Max-pooling layers make the output of convolution networks more robust to local translations. When applied to aligned facial images, they make the network more robust to small registration errors. However, several levels of pooling would cause the network to lose information about the precise position of detailed facial structure and micro-textures. Hence, we apply max-pooling only to the first convolutional layer. We interpret these first layers as a front-end adaptive pre-processing stage. While they are responsible for most of the computation, they hold

<sup>1</sup> $T_{2d}$  can be used here to avoid going through the 2D lossy warping.

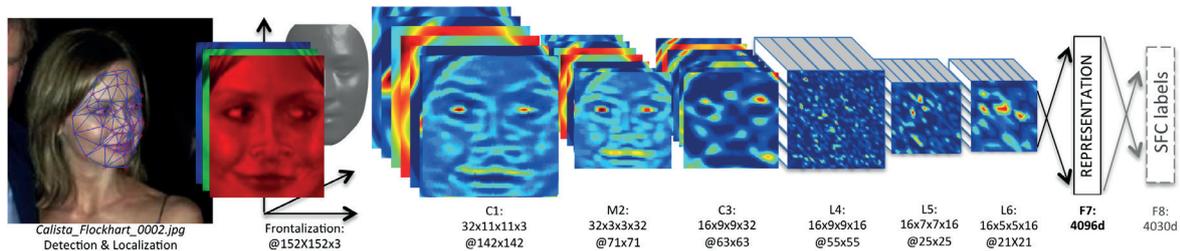


Figure 2. **Outline of the DeepFace architecture.** A front-end of a single convolution-pooling-convolution filtering on the rectified input, followed by three locally-connected layers and two fully-connected layers. Colors illustrate feature maps produced at each layer. The net includes more than 120 million parameters, where more than 95% come from the local and fully connected layers.

very few parameters. These layers merely expand the input into a set of simple local features.

The subsequent layers (L4, L5 and L6) are instead locally connected [13, 16], like a convolutional layer they apply a filter bank, but every location in the feature map learns a different set of filters. Since different regions of an aligned image have different local statistics, the spatial stationarity assumption of convolution cannot hold. For example, areas between the eyes and the eyebrows exhibit very different appearance and have much higher discrimination ability compared to areas between the nose and the mouth. In other words, we customize the architecture of the DNN by leveraging the fact that our input images are aligned. The use of local layers does not affect the computational burden of feature extraction, but does affect the number of parameters subject to training. Only because we have a large labeled dataset, we can afford three large locally connected layers. The use of locally connected layers (without weight sharing) can also be justified by the fact that each output unit of a locally connected layer is affected by a very large patch of the input. For instance, the output of L6 is influenced by a  $74 \times 74 \times 3$  patch at the input, and there is hardly any statistical sharing between such large patches in *aligned* faces.

Finally, the top two layers (F7 and F8) are fully connected: each output unit is connected to all inputs. These layers are able to capture correlations between features captured in distant parts of the face images, *e.g.*, position and shape of eyes and position and shape of mouth. The output of the first fully connected layer (F7) in the network will be used as our raw face representation feature vector throughout this paper. In terms of representation, this is in contrast to the existing LBP-based representations proposed in the literature, that normally pool very local descriptors (by computing histograms) and use this as input to a classifier.

The output of the last fully-connected layer is fed to a K-way softmax (where K is the number of classes) which produces a distribution over the class labels. If we denote by  $o_k$  the  $k$ -th output of the network on a given input, the probability assigned to the  $k$ -th class is the output of the softmax function:  $p_k = \exp(o_k) / \sum_h \exp(o_h)$ .

The goal of training is to maximize the probability of the correct class (face id). We achieve this by minimizing the cross-entropy loss for each training sample. If  $k$  is the index of the true label for a given input, the loss is:  $L = -\log p_k$ . The loss is minimized over the parameters by computing the gradient of  $L$  w.r.t. the parameters and by updating the parameters using stochastic gradient descent (SGD). The gradients are computed by standard back-propagation of the error [25, 21]. One interesting property of the features produced by this network is that they are very sparse. On average, 75% of the feature components in the topmost layers are exactly zero. This is mainly due to the use of the ReLU [10] activation function:  $\max(0, x)$ . This soft-thresholding non-linearity is applied after every convolution, locally connected and fully connected layer (except the last one), making the whole cascade produce highly non-linear and sparse features. Sparsity is also encouraged by the use of a regularization method called dropout [19] which sets random feature components to 0 during training. We have applied dropout only to the first fully-connected layer. Due to the large training set, we did not observe significant overfitting during training<sup>2</sup>.

Given an image  $I$ , the representation  $G(I)$  is then computed using the described feed-forward network. Any feed-forward neural network with  $L$  layers, can be seen as a composition of functions  $g_\phi^l$ . In our case, the representation is:  $G(I) = g_\phi^{F7}(g_\phi^{L6}(\dots g_\phi^{C1}(T(I, \theta_T))\dots))$  with the net's parameters  $\phi = \{C_1, \dots, F_7\}$  and  $\theta_T = \{x_{2d}, \vec{P}, \vec{r}\}$  as described in Section 2.

**Normalization** As a final stage we normalize the features to be between zero and one in order to reduce the sensitivity to illumination changes: Each component of the feature vector is divided by its largest value across the training set. This is then followed by  $L_2$ -normalization:  $f(I) := \tilde{G}(I) / \|\tilde{G}(I)\|_2$  where  $\tilde{G}(I)_i = G(I)_i / \max(G_i, \epsilon)$ <sup>3</sup>. Since we employ ReLU activations, our system is not invariant to re-scaling of the image intensities. Without bi-

<sup>2</sup>See the **supplementary** material for more details.

<sup>3</sup> $\epsilon = 0.05$  in order to avoid division by a small number.

ases in the DNN, perfect equivariance would have been achieved.

#### 4. Verification Metric

Verifying whether two input instances belong to the same class (identity) or not has been extensively researched in the domain of unconstrained face-recognition, with supervised methods showing a clear performance advantage over unsupervised ones. By training on the target-domain’s training set, one is able to fine-tune a feature vector (or classifier) to perform better within the particular distribution of the dataset. For instance, *LFW* has about 75% males, celebrities that were photographed by mostly professional photographers. As demonstrated in [5], training and testing within different domain distributions hurt performance considerably and requires further tuning to the representation (or classifier) in order to improve their generalization and performance. However, fitting a model to a relatively small dataset reduces its generalization to other datasets. In this work, we aim at learning an unsupervised metric that generalizes well to several datasets. Our unsupervised similarity is simply the inner product between the two normalized feature vectors. We have also experimented with a supervised metric, the  $\chi^2$  similarity and the Siamese network.

##### 4.1. Weighted $\chi^2$ distance

The normalized *DeepFace* feature vector in our method contains several similarities to histogram-based features, such as LBP [1]: (1) It contains non-negative values, (2) it is very sparse, and (3) its values are between [0, 1]. Hence, similarly to [1], we use the weighted- $\chi^2$  similarity:  $\chi^2(f_1, f_2) = \sum_i w_i (f_1[i] - f_2[i])^2 / (f_1[i] + f_2[i])$  where  $f_1$  and  $f_2$  are the *DeepFace* representations. The weight parameters are learned using a linear SVM, applied to vectors of the elements  $(f_1[i] - f_2[i])^2 / (f_1[i] + f_2[i])$ .

##### 4.2. Siamese network

We have also tested an end-to-end metric learning approach, known as Siamese network [8]: once learned, the face recognition network (without the top layer) is replicated twice (one for each input image) and the features are used to directly predict whether the two input images belong to the same person. This is accomplished by: a) taking the absolute difference between the features, followed by b) a top fully connected layer that maps into a single logistic unit (same/not same). The network has roughly the same number of parameters as the original one, since much of it is shared between the two replicas, but requires twice the computation. Notice that in order to prevent overfitting on the face verification task, we enable training for only the two topmost layers. The Siamese network’s induced distance is:  $d(f_1, f_2) = \sum_i \alpha_i |f_1[i] - f_2[i]|$ , where  $\alpha_i$  are

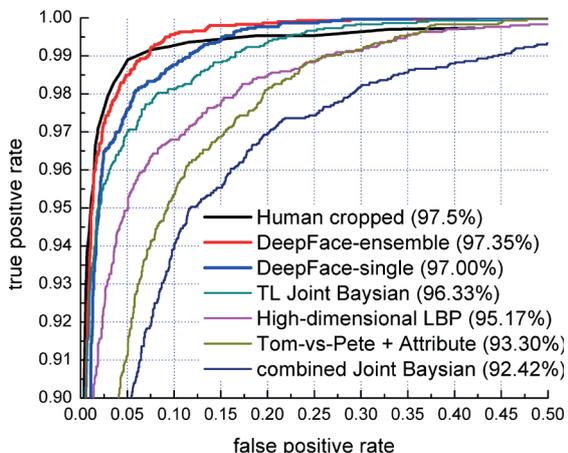


Figure 3. The ROC curves on the *LFW* dataset. Best viewed in color.

trainable parameters. The parameters of the Siamese network are trained by standard cross entropy loss and back-propagation of the error.

## 5. Experiments

We evaluate the proposed *DeepFace* system, by learning the face representation on a very large-scale labeled face dataset collected online. In this section, we first introduce the datasets used in the experiments, then present the detailed evaluation and comparison with the state-of-the-art, as well as some insights and findings about learning and transferring the deep face representations.

### 5.1. Datasets

The proposed face representation is learned from a large collection of photos from *Facebook*, referred to as the Social Face Classification (*SFC*) dataset. The representations are then applied to the Labeled Faces in the Wild database (*LFW*), which is the *de facto* benchmark dataset for face verification in unconstrained environments, and the YouTube Faces (*YTF*) dataset, which is modeled similarly to the *LFW* but focuses on video clips.

The *SFC* dataset includes 4.4 million labeled faces from 4,030 people each with 800 to 1200 faces, where the most recent 5% of face images of each identity are left out for testing. This is done according to the images’ time-stamp in order to simulate continuous identification through aging. The large number of images per person provides a unique opportunity for learning the invariance needed for the core problem of face recognition. We have validated using several automatic methods, that the identities used for training do not intersect with any of the identities in the below-mentioned datasets, by checking their name labels.

**The LFW dataset** [18] consists of 13,323 web photos of 5,749 celebrities which are divided into 6,000 face pairs in 10 splits. Performance is measured by mean recognition accuracy using A) the *restricted* protocol, in which only same and not same labels are available in training; B) the *unrestricted* protocol, where additional training pairs are accessible in training; and C) an *unsupervised* setting in which no training whatsoever is performed on LFW images.

**The YTF dataset** [30] collects 3,425 YouTube videos of 1,595 subjects (a subset of the celebrities in the LFW). These videos are divided into 5,000 video pairs and 10 splits and used to evaluate the video-level face verification.

The face identities in *SFC* were labeled by humans, which typically incorporate about 3% errors. Social face photos have even larger variations in image quality, lighting, and expressions than the web images of celebrities in the LFW and YTF which were normally taken by professional photographers rather than smartphones<sup>4</sup>.

## 5.2. Training on the SFC

We first train the deep neural network on the *SFC* as a multi-class classification problem using a GPU-based engine, implementing the standard back-propagation on feed-forward nets by stochastic gradient descent (SGD) with momentum (set to 0.9). Our mini-batch size is 128, and we have set an equal learning rate for all trainable layers to 0.01, which was manually decreased, each time by an order of magnitude once the validation error stopped decreasing, to a final rate of 0.0001. We initialized the weights in each layer from a zero-mean Gaussian distribution with  $\sigma = 0.01$ , and biases are set to 0.5. We trained the network for roughly 15 sweeps (epochs) over the whole data which took 3 days. As described in Sec. 3, the responses of the fully connected layer F7 are extracted to serve as the face representation.

We evaluated different design choices of DNN in terms of the classification error on 5% data of *SFC* as the test set. This validated the necessity of using a large-scale face dataset and a deep architecture. First, we vary the train/test dataset size by using a subset of the persons in the *SFC*. Subsets of sizes 1.5K, 3K and 4K persons (1.5M, 3.3M, and 4.4M faces, respectively) are used. Using the architecture in Fig. 2, we trained three networks, denoted by *DF-1.5K*, *DF-3.3K*, and *DF-4.4K*. Table 1 (left column) shows that the classification error grows only modestly from 7.0% on 1.5K persons to 7.2% when classifying 3K persons, which indicates that the capacity of the network can well accommodate the scale of 3M training images. The error rate rises to 8.7% for 4K persons with 4.4M images, showing the network scales comfortably to more persons. We've also varied the global number of samples in *SFC* to 10%, 20%, 50%,

<sup>4</sup>See the **supplementary** material for more details about *SFC*.

leaving the number of identities in place, denoted by *DF-10%*, *DF-20%*, *DF-50%* in the middle column of Table 1. We observed the test errors rise up to 20.7%, because of overfitting on the reduced training set. Since performance does not saturate at 4M images, this shows that the network would benefit from even larger datasets.

We also vary the depth of the networks by chopping off the C3 layer, the two local L4 and L5 layers, or all these 3 layers, referred respectively as *DF-sub1*, *DF-sub2*, and *DF-sub3*. For example, only four trainable layers remain in *DF-sub3* which is a considerably shallower structure compared to the 9 layers of the proposed network in Fig. 2. In training such networks with 4.4M faces, the classification errors stop decreasing after a few epochs and remains at a level higher than that of the deep network, as can be seen in Table 1 (right column). This verifies the necessity of network depth when training on a large face dataset.

## 5.3. Results on the LFW dataset

The vision community has made significant progress on face verification in unconstrained environments in recent years. The mean recognition accuracy on LFW [18] marches steadily towards the human performance of over 97.5% [20]. Given some very hard cases due to aging effects, large lighting and face pose variations in LFW, any improvement over the state-of-the-art is very remarkable and the system has to be composed by highly optimized modules. There is a strong diminishing return effect and any progress now requires a substantial effort to reduce the number of errors of state-of-the-art methods. *DeepFace* couples large feedforward-based models with fine 3D alignment. Regarding the importance of each component: 1) Without frontalization: when using only the 2D alignment, the obtained accuracy is “only” 94.3%. Without alignment at all, *i.e.*, using the center crop of face detection, the accuracy is 87.9% as parts of the facial region may fall out of the crop. 2) Without learning: when using frontalization only, and a naive LBP/SVM combination, the accuracy is 91.4% which is already notable given the simplicity of such a classifier.

All the LFW images are processed in the same pipeline that was used to train on the *SFC* dataset, denoted as *DeepFace-single*. To evaluate the discriminative capability of the face representation in isolation, we follow the **unsupervised** setting to directly compare the inner product of a pair of normalized features. Quite remarkably, this achieves a mean accuracy of 95.92% which is almost on par with the best performance to date, achieved by supervised transfer learning [5]. Next, we learn a kernel SVM (with  $C=1$ ) on top of the  $\chi^2$ -distance vector (Sec. 4.1) following the **restricted** protocol, *i.e.*, where only the 5,400 pair labels per split are available for the SVM training. This achieves an accuracy **97.00%**, reducing significantly the error of the state-of-the-art [7, 5], see Table 3.

Network	Error	Network	Error	Network	Error
<i>DF-1.5K</i>	7.00%	<i>DF-10%</i>	20.7%	<i>DF-sub1</i>	11.2%
<i>DF-3.3K</i>	7.22%	<i>DF-20%</i>	15.1%	<i>DF-sub2</i>	12.6%
<i>DF-4.4K</i>	8.74%	<i>DF-50%</i>	10.9%	<i>DF-sub3</i>	13.5%

Table 1. Comparison of the classification errors on the *SFC* w.r.t. training dataset size and network depth. See Sec. 5.2 for details.

Network	Error ( <i>SFC</i> )	Accuracy $\pm$ SE ( <i>LFW</i> )
<i>DeepFace-align2D</i>	9.5%	0.9430 $\pm$ 0.0043
<i>DeepFace-gradient</i>	8.9%	0.9582 $\pm$ 0.0037
<i>DeepFace-Siamese</i>	NA	0.9617 $\pm$ 0.0038

Table 2. The performance of various individual *DeepFace* networks and the Siamese network.

**Ensembles of DNNs** Next, we combine multiple networks trained by feeding different types of inputs to the DNN: 1) The network *DeepFace-single* described above based on 3D aligned RGB inputs; 2) The gray-level image plus image gradient magnitude and orientation; and 3) the 2D-aligned RGB images. We combine those distances using a non-linear SVM (with  $C=1$ ) with a simple sum of power CPD-kernels:  $K_{\text{Combined}} := K_{\text{single}} + K_{\text{gradient}} + K_{\text{align2d}}$ , where  $K(x, y) := -\|x - y\|_2$ , and following the restricted protocol, achieve an accuracy **97.15%**.

The *unrestricted* protocol provides the operator with knowledge about the identities in the training sets, hence enabling the generation of many more training pairs to be added to the training set. We further experiment with training a Siamese Network (Sec. 4.2) to learn a verification metric by fine-tuning the Siamese’s (shared) pre-trained feature extractor. Following this procedure, we have observed substantial overfitting to the training data. The training pairs generated using the *LFW* training data are redundant as they are generated out of roughly 9K photos, which are insufficient to reliably estimate more than 120M parameters. To address these issues, we have collected an additional dataset following the same procedure as with the *SFC*, containing an additional new 100K identities, each with only 30 samples to generate same and not-same pairs from. We then trained the Siamese Network on it followed by 2 training epochs on the *LFW* unrestricted training splits to correct for some of the data set dependent biases. The slightly-refined representation is handled similarly as before. Combining it into the above-mentioned ensemble, *i.e.*,  $K_{\text{Combined}} += K_{\text{Siamese}}$ , yields the accuracy **97.25%**, under the *unrestricted* protocol. By adding four additional *DeepFace-single* networks to the ensemble, each trained from scratch with different random seeds, *i.e.*,  $K_{\text{Combined}} += \sum K_{\text{DeepFace-Single}}$ , the obtained accuracy is **97.35%**. The performances of the individual networks, before combination, are presented in Table 2.

The comparisons with the recent state-of-the-art meth-

Method	Accuracy $\pm$ SE	Protocol
Joint Bayesian [6]	0.9242 $\pm$ 0.0108	restricted
Tom-vs-Pete [4]	0.9330 $\pm$ 0.0128	restricted
High-dim LBP [7]	0.9517 $\pm$ 0.0113	restricted
TL Joint Bayesian [5]	0.9633 $\pm$ 0.0108	restricted
<i>DeepFace-single</i>	<b>0.9592</b> $\pm$ 0.0029	unsupervised
<i>DeepFace-single</i>	<b>0.9700</b> $\pm$ 0.0028	restricted
<i>DeepFace-ensemble</i>	<b>0.9715</b> $\pm$ 0.0027	restricted
<i>DeepFace-ensemble</i>	<b>0.9735</b> $\pm$ 0.0025	unrestricted
Human, cropped	0.9753	

Table 3. Comparison with the state-of-the-art on the *LFW* dataset.

Method	Accuracy (%)	AUC	EER
MBGS+SVM- [31]	78.9 $\pm$ 1.9	86.9	21.2
APEM+FUSION [22]	79.1 $\pm$ 1.5	86.6	21.4
STFRD+PMML [9]	79.5 $\pm$ 2.5	88.6	19.9
VSOFF+OSS [23]	79.7 $\pm$ 1.8	89.4	20.0
<i>DeepFace-single</i>	<b>91.4</b> $\pm$ 1.1	96.3	8.6

Table 4. Comparison with the state-of-the-art on the *YTF* dataset.

ods in terms of the mean accuracy and ROC curves are presented in Table 3 and Fig. 3, including human performance on the cropped faces. The proposed *DeepFace* method advances the state-of-the-art, closely approaching human performance in face verification.

#### 5.4. Results on the *YTF* dataset

We further validate *DeepFace* on the recent video-level face verification dataset. The image quality of YouTube video frames is generally worse than that of web photos, mainly due to motion blur or viewing distance. We employ the *DeepFace-single* representation directly by creating, for every pair of training videos, 50 pairs of frames, one from each video, and label these as same or not-same in accordance with the video training pair. Then a weighted  $\chi^2$  model is learned as in Sec. 4.1. Given a test-pair, we sample 100 random pairs of frames, one from each video, and use the mean value of the learned weighed similarity.

The comparison with recent methods is shown in Table 4 and Fig. 4. We report an accuracy of **91.4%** which reduces the error of the previous best methods by more than 50%. Note that there are about 100 wrong labels for video pairs, recently updated to the *YTF* webpage. After these are corrected, *DeepFace-single* actually reaches 92.5%. This experiment verifies again that the *DeepFace* method easily generalizes to a new target domain.

#### 5.5. Computational efficiency

We have efficiently implemented a CPU-based feedforward operator, which exploits both the CPU’s Single Instruction Multiple Data (SIMD) instructions and its cache by leveraging the locality of floating-point computations

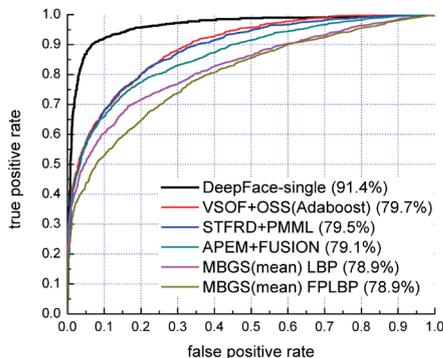


Figure 4. The ROC curves on the YTF dataset. Best viewed in color.

across the kernels and the image. Using a single core Intel 2.2GHz CPU, the operator takes 0.18 seconds to extract features from the raw input pixels. Efficient warping techniques were implemented for alignment; alignment alone takes about 0.05 seconds. Overall, the *DeepFace* runs at 0.33 seconds per image, accounting for image decoding, face detection and alignment, the feedforward network, and the final classification output.

## 6. Conclusion

An ideal face classifier would recognize faces in accuracy that is only matched by humans. The underlying face descriptor would need to be invariant to pose, illumination, expression, and image quality. It should also be general, in the sense that it could be applied to various populations with little modifications, if any at all. In addition, short descriptors are preferable, and if possible, sparse features. Certainly, rapid computation time is also a concern. We believe that this work, which departs from the recent trend of using more features and employing a more powerful metric learning technique, has addressed this challenge, closing the vast majority of this performance gap. Our work demonstrates that coupling a 3D model-based alignment with large capacity feedforward models can effectively learn from many examples to overcome the drawbacks and limitations of previous methods. The ability to present a marked improvement in face recognition, attests to the potential of such coupling to become significant in other vision domains as well.

## References

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *PAMI*, 2006. 3, 5
- [2] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz. Fast high dimensional vector multiplication face recognition. In *ICCV*, 2013. 1, 2
- [3] Y. Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2009. 1
- [4] T. Berg and P. N. Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In *BMVC*, 2012. 2, 7
- [5] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun. A practical transfer learning algorithm for face verification. In *ICCV*, 2013. 1, 2, 5, 6, 7
- [6] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In *ECCV*, 2012. 2, 7
- [7] D. Chen, X. Cao, F. Wen, and J. Sun. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *CVPR*, 2013. 1, 2, 6, 7
- [8] S. Chopra, R. Hadsell, and Y. LeCun. Learning a similarity metric discriminatively, with application to face verification. In *CVPR*, 2005. 2, 5
- [9] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *CVPR*, 2013. 7
- [10] G. E. Dahl, T. N. Sainath, and G. E. Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *ICASSP*, 2013. 4
- [11] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, Q. Le, M. Mao, M. Ranzato, A. Senior, P. Tucker, K. Yang, and A. Ng. Large scale distributed deep networks. In *NIPS*, 2012. 2
- [12] P. J. P. et al. An introduction to the good, the bad, & the ugly face recognition challenge problem. In *FG*, 2011. 2
- [13] K. Gregor and Y. LeCun. Emergence of complex-like cells in a temporal product network with local receptive fields. arXiv:1006.0448, 2010. 4
- [14] T. Hassner. Viewing real-world faces in 3D. In *International Conference on Computer Vision (ICCV)*, Dec. 2013. 2
- [15] G. B. Huang, V. Jain, and E. G. Learned-Miller. Unsupervised joint alignment of complex images. In *ICCV*, 2007. 2
- [16] G. B. Huang, H. Lee, and E. Learned-Miller. Learning hierarchical representations for face verification with convolutional deep belief networks. In *CVPR*, 2012. 2, 4
- [17] G. B. Huang, M. A. Mattar, H. Lee, and E. G. Learned-Miller. Learning to align from scratch. In *NIPS*, pages 773–781, 2012. 2
- [18] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *ECCV Workshop on Faces in Real-life Images*, 2008. 1, 6
- [19] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In *ANIPS*, 2012. 1, 2, 3, 4
- [20] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009. 6
- [21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient based learning applied to document recognition. *Proc. IEEE*, 1998. 1, 2, 4
- [22] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang. Probabilistic elastic matching for pose variant face verification. In *CVPR*, 2013. 7
- [23] H. Mendez-Vazquez, Y. Martinez-Diaz, and Z. Chai. Volume structured ordinal features with background similarity measure for video face recognition. In *Int'l Conf. on Biometrics*, 2013. 7
- [24] M. Osadchy, Y. LeCun, and M. Miller. Synergistic face detection and pose estimation with energy-based models. *JMLR*, 2007. 2
- [25] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 1986. 2, 4
- [26] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *BMVC*, 2013. 2
- [27] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *CVPR*, 2013. 2
- [28] Y. Taigman and L. Wolf. Leveraging billions of faces to overcome performance barriers in unconstrained face recognition. arXiv:1108.1122, 2011. 2
- [29] Y. Taigman, L. Wolf, and T. Hassner. Multiple one-shots for utilizing class label information. In *BMVC*, 2009. 2
- [30] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR*, 2011. 1, 6
- [31] L. Wolf and N. Levy. The SVM-minus similarity score for video face recognition. In *CVPR*, 2013. 7
- [32] D. Yi, Z. Lei, and S. Z. Li. Towards pose robust face recognition. In *CVPR*, 2013. 2

**Annexe 3, Totems - Partie pratique de mémoire, Victor Poullain**



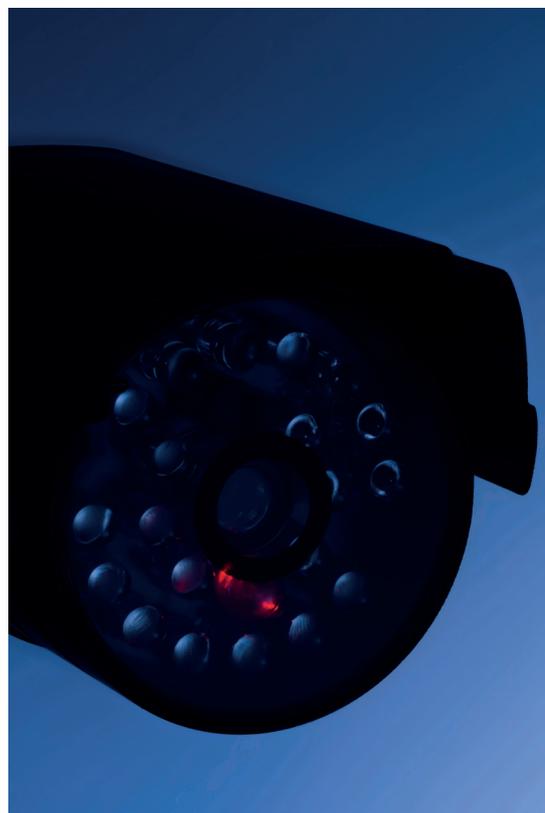
*Totems, 1/7, Victor Poullain, photographie, 2016*



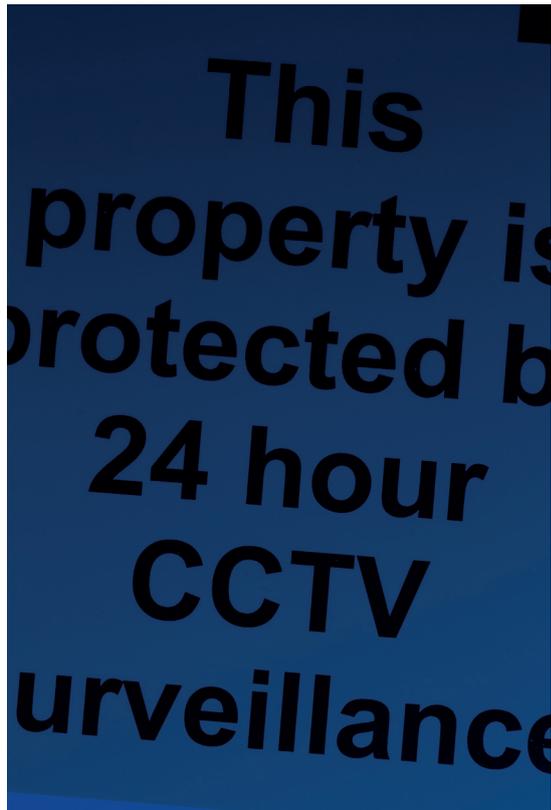
*Totems, 2/7, Victor Poullain, photographie, 2016*



*Totems, 3/7, Victor Poullain, photographie, 2016*



*Totems, 4/7, Victor Poullain, photographie, 2016*



*Totems, 5/7, Victor Poullain, photographie, 2016*



*Totems, 6/7, Victor Poullain, photographie, 2016*



*Totems, 7/7, Victor Poullain, photographie, 2016*