

# Du portrait pictural aux *deepfakes* : le visage en tant que totalité<sup>1</sup>

*From Painted Portraiture to Deepfake: Face as a Totality*

< Maria Giulia Dondero <sup>1</sup> >

1. Fond National de la Recherche Scientifique ; Université de Liège  
mariagiulia.dondero@uliege.be

DOI : 10.25965/interfaces-numeriques.4855

## < RÉSUMÉ >

Cet article porte sur les stratégies d'accumulation et de manipulation des images du visage. Plus particulièrement, il met en contraste la *composite photography* de Francis Galton et l'accumulation/manipulation de photos de visage dans la production et la détection des *deepfakes*. L'article analyse la relation entre singularité et généralité dans ces deux cas d'étude, en focalisant préalablement l'attention sur la tradition et l'idéaltype du portrait pictural.

## < MOTS-CLÉS >

portrait, identité, photographie composite, deepfake

## < ABSTRACT >

This paper focuses on the representational strategies characterized by the accumulation and manipulation of facial images. More specifically, it contrasts Francis Galton's composite photography and the accumulation/manipulation of facial images in the production and detection of deepfakes. The article analyzes the relationship between singularity and generality in these two cases of study, focusing beforehand on the pictorial tradition of portraiture.

---

<sup>1</sup> Cet article est une version traduite et augmentée d'un article publié en anglais dans la revue *Lexia. Rivista di Semiotica* ayant pour titre "Composition and decomposition in artistic portraits, scientific photography and in deepfake videos" (*Lexia*, 37-38, 437-453, 2021, DOI 10.4399/978882553853321).

< **KEYWORDS** >Portraiture, Identity, Composite Photography, Deepfake

---

**1. Introduction**

Ce texte vise à décrire la spécificité des dispositifs de représentation du visage. Dans un premier temps, je vais prendre en considération le portrait en peinture, qui a façonné l'idéaltype du portrait — et qui demeure une référence obligée pour le portrait photographique qui le détourne — ; dans un second temps, je le mettrai en contraste avec les représentations scientifiques du visage, et notamment avec la *Composite Photography* dont Francis Galton a été le pionnier au XIX siècle. Enfin, je porterai mon attention sur les spécificités représentationnelles du visage dans les vidéos *deepfake* contemporaines<sup>2</sup>.

En effet, pour relier les deux univers du portrait pictural et des vidéos contemporaines —qui dépendent de la disponibilité de grandes collections d'images (big data) et des techniques de production et de détection du faux du *deep learning* —, l'expérience de la photographie composite est nécessaire pour comprendre la manière dont la représentation du visage passe de la mise en valeur de l'unicité à celle de la généralité et du caractère interchangeable des individualités. Je prendrai en considération les différentes stratégies de cadrage et d'utilisation des représentations du visage employées par les trois médias que sont la peinture artistique, la photographie scientifique et la vidéo, afin d'identifier les règles de représentation ainsi que les modèles d'individualité/généralité qui se sont succédés dans les siècles.

---

<sup>2</sup> Je tiens à préciser que la partie sur le portrait a été grandement inspirée par les travaux d'Anne Beyaert-Geslin sur le portrait, non seulement par son livre *Sémiotique du portrait. De Dibutade au Selfie* (2017) mais aussi par son HDR sur les figures de l'autre (*L'autre de l'autre*, 2008). Les parties de ce texte consacrées au deepfake ont en revanche été travaillées grâce aux invitations de Massimo Leone à prendre au sérieux certains phénomènes contemporains liés à l'esthétique et à l'éthique du visage, qu'il aborde en tant que Principal Investigator dans le cadre du projet européen ERC *Facets. Face Aesthetics in Contemporary E-Technological Societies* (2019-2025).

Les principales questions abordées seront : Comment, d'une part, représenter une identité *spécifique* ? Comment, d'autre part, représenter une identité ou un comportement *moyen* ou *typique* d'une classe d'individus ? Et enfin : Comment construire, par le biais d'images multiples, une identité typique de soi-même, afin de la rendre remplaçable par celle d'un autre ?

## 2. La compacité du portrait pictural

Dans cette première partie de mon article, j'aborderai les règles de composition des portraits, à savoir les règles qui font, qu'en histoire de l'art, un portrait soit reconnu comme tel. Au sein de la tradition historico-artistique, la compacité est considérée comme la caractéristique nécessaire de tout portrait. Elle permet notamment de distinguer les portraits des paysages contenant une figure humaine ou une scène d'actualité.

Les portraits apparaissent en effet comme peu articulés et excessivement immuables, surtout si on les compare aux autres grands genres de la tradition picturale. Il semble que le portrait ne mette en avant qu'un simple contraste, c'est-à-dire une articulation binaire figure/fond qui empêche l'appréhension de tout déploiement narratif de l'image. La caractéristique essentielle du portrait réside dans le fait que la figure doit être placée au centre, et produire une « totalité cohésive » opposée au vide formé par le fond (Dondero 2020).

La densité et la compacité de la figure sur le plan de l'expression doivent pouvoir signifier, sur le plan du contenu, l'addition et la *summa* des expériences de vie du sujet représenté, pris dans sa totalité. La perfection du portrait est atteinte lorsque le sujet réussit à faire coïncider son expérience passée et son destin avec le *je-ici-maintenant* de la présence au temps présent. C'est la raison pour laquelle le portrait met l'accent sur l'unité et l'unicité du visage et non sur les gestes ou les actions. Parmi toutes les parties du corps, ce sont les yeux qui sont les plus marqués par les expériences passées et qui condensent l'histoire d'une vie à exposer devant un spectateur. Le sujet est ainsi généralement présenté via la partie considérée comme la plus « noble », à savoir la zone entourant les yeux et la tête. Une focalisation sur le corps entier aurait le

désavantage de suggérer une proto-action du sujet du portrait, ce qui irait à l'encontre d'un face à face intensif entre l'observateur et l'individu représenté. Il n'est donc pas surprenant que l'utilisation du flou ait été exclue des portraits, car elle empêcherait la reconnaissance et la valorisation du sujet, qui forme par définition une totalité bien déterminée et circonscrite. Dans le portrait, le flou empêcherait la stabilisation de l'identité du sujet et suspendrait tout échange avec l'observateur ; ce qui résulterait du flou serait ainsi une identité flottante (Beyaert-Geslin et Lloveria 2014).

La question qui se pose maintenant est la suivante : Le visage constitue-t-il le véritable objet d'investigation du portrait, ou son objet est-il autre ? Une réponse classique serait celle de dire que les portraits représentent l'acte de regarder — et ce ne serait pas du tout faux. En même temps, je pense que dans les portraits traditionnels, ce qui est représenté, c'est surtout un système d'équilibre entre la figure et le fond, c'est-à-dire une structure délicate où le fond doit laisser de la place pour que la figure puisse émerger — mais pas trop de place pour éviter que la figure manque d'un soutien solide. En même temps, il ne faut pas qu'elle laisse trop peu de place non plus au fond, car alors la force de présence de la figure disparaîtrait, puisqu'elle serait engloutie par le fond.

### **3. La libération du visage et le portrait des types (autour de Galton et de Peirce)**

Ce que l'on peut appeler la « libération du visage du fond » se produit lorsque l'équilibre entre la figure et le fond est défait, et par conséquent la cohésion et la compacité du visage se défait également. Avant le développement des techniques de *deep learning* — qui ont permis la décomposition et la recomposition des visages —, plusieurs expériences esthétiques et scientifiques ont détourné la caractéristique essentielle du portrait, la compacité. Deux exemples de la libération du visage de la structure traditionnelle des portraits nous viennent de la peinture et des sciences.

Dans le domaine de la peinture, je ne mentionnerai que les visages flous de Francis Bacon. Les portraits de Bacon sont composés de zones chromatiques et matiériques en mouvement, en déplacement continu et,

plus généralement, de tensions entre des forces conflictuelles de contraction et d'expansion, de fusion et d'explosion (Deleuze 1981). L'identification d'une frontière entre le visage et le fond devient difficile : l'incertitude, l'instabilité et la dispersion de l'identité domine ce genre de portraits.

Dans le domaine des sciences, ce sont les expériences scientifiques de Francis Galton qui inaugurent la décomposition de l'équilibre entre figure et fond et qui génèrent des portraits de types (Figure 1). Il est important de rappeler que Francis Galton était une personnalité éminente de l'anthropologie et des statistiques britanniques. Il était également géographe, météorologue, écrivain, proto-ingénieur et psychométricien, et il est considéré comme le fondateur de la psychologie différentielle ou comparative. Galton est surtout connu pour ses portraits de types criminels mais, comme le remarque Allan Sekula dans un célèbre article publié en 1986, « The Body and the Archive » (Sekula 1986, p. 18), son travail n'était lié à la criminologie<sup>3</sup> que de manière périphérique. Les travaux de Galton s'inscrivaient plutôt dans le cadre d'une entreprise plus vaste visant à utiliser les statistiques pour l'« amélioration » de la race britannique. Ainsi, le processus composite a été utilisé non seulement pour identifier les types de criminels, mais aussi pour étudier les traits familiaux, la propension à la maladie, la reproduction en général et, plus tard, pour isoler les caractéristiques raciales typiques.

Il est mort en 1911, quelques années seulement avant Ch. S. Peirce, disparu en 1914.

Galton a inspiré une grande partie de la pensée théorique de Peirce concernant la notion de *photographie composite* (Peirce 1895, Basso et Dondero 2011). Les portraits composites de Galton étaient produits par l'enregistrement et l'exposition successifs d'images sur une seule plaque, avant de les soumettre à un processus de superposition. La nature scientifique de ses photographies composites provient du fait que chaque visage a été capturé en utilisant les mêmes paramètres, la même perspective, la même distance focale, et la même position par rapport à

---

3 Pour une interprétation bien documentée de l'utilisation des méthodes photographiques médico-légales de Bertillon et Galton et des empreintes digitales dans l'histoire de la criminologie, voir Leone (2021).

une grille de fond, organisée à travers des abscisses et des ordonnées. Ces paramètres ont été établis afin d'assurer la commensurabilité des visages superposés — comme montré dans la partie inférieure de cette image sous le titre « Combinations of portraits » (Figure 1).

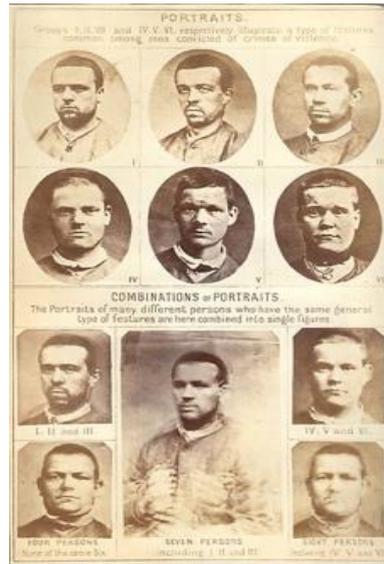


Figure 1. Francis Galton, *Portraits composites de types criminels*, 1877. The Galton Archive, University College London Special Collections.

Dans son article intitulé « Composite Portraits Made by Combining Those of Many Different Persons Into a Single Figure », publié en 1878, Galton déclare :

*Le procédé photographique nous permet d'obtenir avec une précision mécanique une image générale ; une image qui ne représente aucun homme en particulier, mais qui représente une figure imaginaire possédant les caractéristiques moyennes d'un groupe d'hommes donné. Ces visages idéaux possèdent un surprenant air de réalité. Personne, en regardant l'un d'entre eux pour la première fois, ne douterait qu'il s'agisse de la ressemblance avec une personne vivante. Pourtant, comme je l'ai dit, il n'en est rien ; c'est le portrait d'un type et non d'un individu. (Galton 1878, p. 13, nous traduisons et soulignons)*

Les portraits composites de Galton étaient considérés comme de véritables moyennes statistiques, représentant des types humains — un criminel, une prostituée, un Anglais, un Juif, etc. Galton a écrit à propos de ses images composites qu'elles sont :

*bien plus que des moyennes ; [...] Ce sont de véritables généralisations, car elles incluent l'ensemble des matériaux recueillis. Le flou de leurs contours, qui n'est jamais très important dans les photos composites véritablement générales, sauf dans des moindres détails, mesure la tendance des individus à s'écarter du type qui se positionne au centre de l'image (Galton 1879, p. 166, nous traduisons et soulignons).*

Le flou des contours dont parle Galton est très important dans ma démonstration. En effet, si le flou était traditionnellement interdit dans tout type de portrait, avec Galton et ses portraits de types, nous sommes confrontés au fait surprenant que l'individualité émerge du flou, car le flou des contours nous permet de voir ce qui se détache de la généralité du type et permet de dévoiler une individualité, voire des spécificités<sup>4</sup>.

Chiara Ambrosio, dans un article intitulé « Composite Photographs and the Quest for Generality » (2016), affirme que plusieurs aspects de la méthode photographique de Galton ont très tôt attiré l'attention de Peirce et ont marqué ses travaux. En effet, dans sa « Short Logic » (1895), Peirce était déjà en train de transformer la méthode de Galton en une métaphore conceptuelle, visant à concevoir un outil exploratoire pour comprendre la nature des idées. Ce qui avait fasciné Peirce, selon Ambrosio, était le fait que Galton non seulement affirmait que « les visages idéaux obtenus par la méthode du portrait composite semblent avoir beaucoup en commun avec [...] les soi-disant idées abstraites », mais qu'il proposait de

---

<sup>4</sup> Comme Peirce, Manovich (1995) a également pris connaissance des travaux de Galton. Ce qui, pour Manovich, est intéressant dans le travail de Galton, c'est qu'avec ses photographies, Galton n'a pas seulement proposé que la généralité puisse être représentée par des images : il l'a objectivée et matérialisée. Pour Manovich, ce phénomène peut être appelé « externalisation de l'esprit » : l'objectivation des processus mentaux internes, privés, par des formes visuelles externes que l'on peut manipuler permet de partager ce qui est caché dans l'esprit d'un individu. Manovich utilise les travaux de Galton comme exemple pour expliquer la naissance de la demande de normalisation de la société de masse moderne. Il remarque que les sujets sont censés être standardisés, et que les moyens par lesquels ils sont standardisés sont censés l'être également.

renommer les idées abstraites « idées cumulatives » (Galton, 1883, p. 183.)

Comme Ambrosio l'affirme dans cet article, les photographies composites ne visaient pas la comparaison empirique entre cerveau et appareil photo. Les métaphores de la *composition* et du *composite* visaient à signifier plutôt le « type de contrôle (effectué par la faculté de jugement) que l'esprit exerce sur le processus de généralisation » (Ambrosio 2016, p. 14, nous traduisons). Selon Ambrosio, Peirce a ainsi déplacé le processus de généralisation des visages individuels de Galton vers la faculté de jugement et la catégorisation. À ce propos, Ambrosio énonce les différences entre les productions des deux savants :

*Les photographies composites de Galton sont statiques : ce sont des présentations de types idéaux, dont la généralité est validée par la fiabilité du processus mécanique qui a servi à leur génération. Les composites de Peirce, en revanche, sont intrinsèquement dynamiques : ils ont une base expérientielle (certaines des nuances de la couleur jaune auxquelles nous comparons la couleur de notre chaise ont pu être vues, expérimentées), mais ils ont aussi une sorte de pouvoir prédictif (la photographie composite nous permettra de reconnaître d'autres nuances de jaune en tant que « jaune », et de les appliquer à d'autres perceptions). (Ambrosio 2016, p. 15, nous traduisons et soulignons)*

Cette vision de la nuance est cruciale car, selon Peirce, ce sont les nuances qui permettent de comprendre les limites des catégories et des classifications, ainsi que d'appliquer une catégorie à de nouveaux percepts — et on pourrait même dire, plus généralement, de rendre les catégories élastiques, plastiques. Pour le dire autrement,

*pour Galton, le centre de l'image est la partie essentielle de la photographie, car c'est au centre que se rassemblent les « traits typiques ». Pour Peirce, au contraire, le processus intéressant se produit dans la périphérie des images, les zones dans lesquelles le flou suggère d'autres façons, éventuellement nouvelles, d'appliquer le « modèle » composite à un nouveau contexte et de dégager de nouvelles relations grâce à cette application (Ambrosio 2016, p. 16, nous soulignons).*

Nous voyons ici à l'œuvre la célèbre logique du vague de Peirce<sup>5</sup>. Le vague nous ramène à la conception peircienne de la découverte et de l'acquisition de connaissances. Toutes les expressions telles que « pouvoir prédictif », « voir ultérieurement » ou « nouveau contexte » utilisées par Peirce et Ambrosio montrent, comme l'affirme Pierluigi Basso dans son travail sur Peirce et la photographie contenu dans *Sémiotique de la photographie*, que la composante iconique de la photographie composite « ouvre une mémorabilité livrée à l'expérience future » (Basso Fossali et Dondero 2011, p. 252). En effet, la photographie composite constitue une classe ouverte de photographies, dont les nouveaux éléments produisent un réajustement des caractéristiques définitoires d'une idée ou d'un percept, en maintenant toujours la sémiotique en mouvement. D'une certaine manière, on pourrait affirmer que la conception de la photographie de Galton était également orientée vers l'avenir, car le type d'homme formé par la superposition des visages individuels devait permettre d'identifier les futurs criminels, les malades, etc. La généralisation réalisée par le type photographique pouvait donc se transformer en schématisation, c'est-à-dire en un dispositif à la fois suffisamment général et reproductible, en plus d'être suffisamment singulier et unique, pour statuer sur de nouvelles occurrences de visages. Mais dans l'œuvre de Peirce, la photographie composite devient quelque chose d'encore plus fondamental : la métaphore d'une adéquation étroite avec le domaine de l'expérience, cette dernière étant caractérisée par la succession des « plans de la réalité », étroitement liés à nos réactions perceptives. Il est clair en effet que l'intérêt de Peirce pour la photographie composite va bien au-delà de sa technique et de l'actualité du débat scientifique qui l'entoure à la fin du XIXe siècle. Peirce va plus loin que Galton aussi dans la conception de la photographie : toute photographie « instantanée » résulte d'une opération d'assemblage ou plutôt de la réunification des intervalles d'exposition sur une seule plaque, rendant imperceptible *la stratification des différentes appréhensions du devenir* : même ce que l'on appelle une « photographie instantanée », prise avec un appareil photo, est *un composite/une*

---

<sup>5</sup> Voir à ce propos l'excellent livre sur le concept de vague dans l'œuvre de Peirce de Christiane Chauviré (1995).

*composition* « des intervalles d'exposition qui sont plus nombreux, et de loin, que les sables de la mer » (Peirce, 2.441, nous traduisons).

Basso compare cette fusion des intervalles d'exposition au fonctionnement de la perception, qui ne peut jamais se réduire à une photographie de l'état des choses, mais qui s'apparente à une photographie composite qui, guidée par la schématisation accomplie par l'imagination, se rapproche des ajustements situationnels et indiciels qui associent localement les percepts : « [Pour le pragmaticiste], tout ce qui est présent dans la substance de ses croyances peut être représenté dans les schématisations de son imagination, à savoir dans ce qui peut être comparé à des photographies composites de séries continues de modifications d'images ; ces composites étant accompagnés de résolutions conditionnelles quant à la conduite » (Peirce 5.517, nous traduisons).

Enfin, la réflexion de Peirce sur la variabilité du monde et l'instabilité de la schématisation accomplie par notre perception est une bonne introduction au caractère interchangeable des visages à l'époque contemporaine, où le visage est pertinent en particulier pour sa flexibilité.

#### **4. Big data et apprentissage profond. Le type individuel et le remplacement de l'identité**

Selon les mots de Lev Manovich (1995), le *deep learning* a perfectionné la standardisation de l'individu qui était déjà en germe à l'époque de Galton. Le *deep learning* a permis de développer des analyses fines de grandes bases de données, notamment des bases d'images faciales. Il a également conduit à de nombreuses formes d'utilisation malveillante de ces bases de données, comme l'usurpation d'identité dans la production de vidéos *deepfake*.

En effet, comme l'explique un article récent de Nguyen *et al.* (2019) intitulé « Deep Learning for Deepfakes Creation and Detection », le *deepfake* (de « deep learning » et « fake ») est une technique qui permet de superposer des images de visage d'une personne cible (Donald Trump) à une vidéo d'une personne source (Nicolas Cage) pour créer une

vidéo de la personne cible (Trump) faisant ou disant des choses appartenant à la personne source (Cage).

Bien qu'aujourd'hui les applications permettant de produire rapidement des superpositions de visages et de gestes à l'aide d'un petit nombre de photos se multiplient, les vidéos professionnelles dépendent des algorithmes de *deep learning* nécessitant une grande quantité de données d'images et de vidéos pour entraîner les modèles afin de créer des images et des vidéos réalistes. Des modèles d'apprentissage profond, tels que les autoencodeurs et les réseaux antagonistes génératifs (GAN, *generative adversarial networks*), ont récemment été utilisés par les algorithmes de *deepfake* pour synthétiser les expressions et les mouvements du visage d'une personne, pour ensuite les greffer sur d'autres identités.

Il est facile de voir qu'il s'agit à nouveau de superposer différentes images de visages comme dans le cadre de la photographie composite, mais ce qui change avec ces technologies est que les multiples photos concernent une seule et même personne. Contrairement aux portraits de types de Galton, qui font la moyenne entre une multiplicité de visages appartenant à des personnes différentes, le processus concerne ici l'accumulation d'images d'une *même* personne afin d'en produire, pour ainsi dire, le *type d'elle-même* —pour ensuite rendre ce type superposable au visage et aux gestes d'une autre personne. En d'autres termes, l'identité d'une personne peut ensuite être superposée au corps d'une autre personne, ce qui revient à transposer le type obtenu à partir des multiples occurrences du visage d'un *seul* individu sur une autre identité *singulière* qui accomplit une action particulière. L'association des multiples photos d'un individu amène à la construction d'un spectre d'images représentant l'individu qui vise non pas une image-type opérant la moyenne de toutes les différentes expressions de l'individu, mais bien une sorte de flexibilité et de manipulabilité : plus le nombre d'images disponibles est important, plus le spectre d'actions à prêter à l'autre est large.

La figure schématique ci-dessous (figure 2) montre un modèle de production de *deepfake* utilisant la paire encodeur-décodeur.

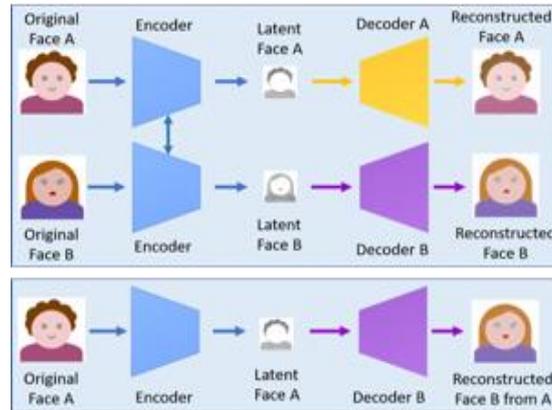


Figure 2. Nguyen et al. 2019, p. 3.

Dans la partie supérieure de la figure, les deux parcours de codification du visage (Original Face A et Original Face B) utilisent chacun le décodeur adéquat au visage pris en considération. Dans la partie inférieure de la figure, en revanche, le décodeur utilisé pour le visage A est celui du visage B. En effet, l'image du visage A qui est encodée en utilisant un encodeur commun (A et B) et décodée avec le décodeur du visage de l'autre personne, le décodeur B, crée un *deepfake*, à savoir la reconstruction du visage B à partir du visage A.

Les stratégies utilisées pour la *détection des deepfakes* sont multiples : certaines reposent sur le rythme de clignement des yeux, tandis que d'autres utilisent des caractéristiques spatio-temporelles, voire les interstices entre les cadrages et les incohérences temporelles.

Les stratégies qui s'appuient sur les caractéristiques temporelles des images vidéo utilisent une méthode de classification binaire, où des classificateurs sont utilisés pour distinguer les vidéos authentiques des vidéos falsifiées. Ce type de méthode nécessite évidemment une grande base de données de vidéos réelles et fausses pour entraîner la capacité de la machine à classer. Deux groupes de méthodes sont généralement employés, à savoir les méthodes qui visent à détecter les imperfections visuelles entre un cadre et l'autre (« visual artifacts *within* video frame-based methods ») et celles consacrées à la détection des caractéristiques

temporelles tout au long de la succession des cadres (« temporal features *across* frame-based methods »).

Le premier groupe explore les imperfections visuelles entre une image et l'autre, tandis que le second utilise les caractéristiques du lien temporel entre les images. Celle qui me semble la plus intéressante est sans doute la seconde méthode, car elle ne travaille pas sur des descripteurs de caractéristiques, mais plutôt sur des *descripteurs de séquences*, et se base sur la continuité du mouvement, sur la durée des expressions faciales.

Partant du constat que la cohérence temporelle n'est pas appliquée efficacement dans le processus de synthèse des *deepfakes*, Sabir *et al.* (2019) ont exploité l'utilisation des caractéristiques spatio-temporelles dans la durée des flux pour détecter les incohérences dans la succession des images. La manipulation vidéo est effectuée image par image, de sorte que les imperfections plastiques (low-level artefacts) ou les glitch visuels produits par les manipulations du visage se manifestent davantage sous la forme d'imperfections et de dissonances temporelles via des incohérences entre les cadres.

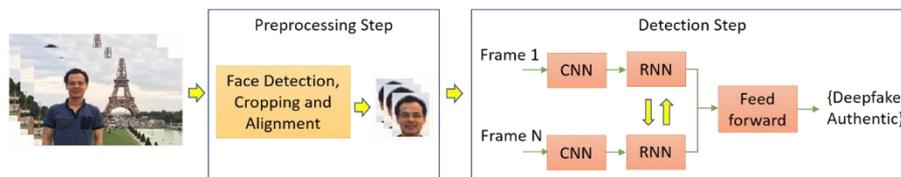


Figure 3. Nguyen *et al.* 2019, p. 5.

Dans cette image (figure 3), on est face à la description d'un processus en deux étapes pour la détection de la manipulation des visages. L'étape de prétraitement vise à détecter, recadrer et aligner les visages sur une séquence d'images, tandis que la deuxième étape distingue les images faciales manipulées des images authentiques en combinant des méthodes de réseau neuronal convolutif (CNN) et de réseau neuronal récurrent (RNN).

Le recadrage et l'alignement des visages sont des techniques utilisées par les détecteurs de vidéos *deepfake* qui étaient déjà connues et employées par Francis Galton : les visages photographiés des différents

individus devaient être totalement alignés et superposables sur une même plaque. Dans la détection de *deepfake*, les chercheurs utilisent l'alignement des différentes positions prises par le visage d'un individu lors de l'enregistrement de la vidéo. Alors que Galton cherchait à réduire la différence entre plusieurs individus à un seul type au sein d'une photographie composite, avec le *deepfake*, on cherche en revanche à construire *le type d'un visage* à partir de la multiplicité de ses mouvements et des expressions faciales se modifiant tout au long de la durée de la vidéo. Quant à la méthode de détection, c'est *dans les interstices* entre une image et une autre qu'il devient possible de détecter l'inauthenticité d'une vidéo.

Si certaines différences de rythme entre une séquence et une autre et entre un clignement des yeux et un autre permettent de reconnaître un fake, il faut aussi dire que ce qui nous surprend lorsque nous regardons des vidéos *deepfake* est le fait que certaines zones floues persistent et que la lumière qui tombe sur les joues n'est pas vraiment naturelle : certaines parties sont trop lumineuses et brillantes tandis que d'autres sont complètement opaques. La répartition de l'intensité lumineuse est sûrement l'une des caractéristiques plastiques sur lesquelles il faut s'appuyer pour la détection.

Le problème du flou est d'ailleurs une composante importante de la production des fakes – comme le donne à voir ce schéma (figure 4) qui présente des expériences plus ou moins réussies de permutation de visages (face swapping).



Figure 4. <https://github.com/shaoanlu/faceswap-GAN/blob/master/README.md>

On remarque que certaines de ces expériences sont réussies parce que le visage est bien exposé et bien droit, tandis que d'autres le sont moins parce que le visage est obstrué ou que les expressions faciales ne sont pas très symétriques. Mais ce qui est important est que ces séries d'images montrent qu'entre une image faciale et l'autre, une figure intermédiaire floue appelée « masque » émerge en combinant les caractéristiques des deux visages, source et cible. La partie interne du « visage intermédiaire » ou « médiateur » est particulièrement floue. En effet, les contours du masque sont bien marqués, tout comme les contours des yeux, du nez et de la bouche. C'est sur ces contours que s'opère la transformation d'un individu en un autre, tandis que ce qui ne fait pas partie de ces contours apparaît moins travaillé, moins parfait. En effet, la peau et tout ce qui, dans un visage, se rapproche d'une configuration de surface, comme les joues, sont traités de manière différente des contours de nos organes de la vision, de l'odorat, de la parole, ainsi que le contour général du visage. Ces derniers représentent des traits qui caractérisent la spécificité de chaque visage, tandis que les joues sont des parties du visage qui peuvent être considérées comme plus « communes » et donc, d'une certaine manière, plus facilement interchangeables.

Nous pourrions en conclure que, alors que dans la tradition du portrait, le flou était interdit afin de ne pas compromettre la stabilité de l'identité de l'individu représenté, dans le cas de Galton, le flou signifie l'émancipation des traits individuels de la généralisation en un type, cette généralisation étant manifestée sur la plaque photographique en tant que coïncidence parmi les multiples visages. Dans les expériences *deepfake*, en revanche, le flou est ce qui permet le passage d'une identité à une autre, et représente ce qui est commun entre les deux.

## 5. Conclusion

Dans la tradition du portrait, la compacité et la netteté de la figure étaient destinées à signifier l'intensité de la présence d'un individu face à un autre individu (l'observateur), le flou signifiant une individualité à l'identité instable. Le flou était interdit également pour une autre raison, cette fois-ci plutôt liée au plan de l'expression : le visage faisait partie d'un système d'équilibres constitué par la relation entre la figure et son fond, qui sont intimement liés<sup>6</sup>. Lorsque le système du portrait en tant que lien entre figure et fond est défait, le visage retrouve plus de liberté, comme dans les tableaux de Bacon. Mais en même temps la solennité de la mise en scène d'une individualité pleine et accomplie est totalement mise en échec.

Les choses changent avec la photographie composite de Galton où la netteté des zones centrales de la stratification des photos représente le type. La netteté, ici, est certes aussi un signifiant de stabilité, mais cette stabilité appartient au type, non pas à l'individu. C'est une netteté dans laquelle l'individu s'efface. C'est seulement dans les zones floues que l'individualité émerge, aux marges de la superposition des visages.

Les *deepfakes* produisent une sorte de type individuel. Par ailleurs et paradoxalement, c'est par le biais des big data, c'est-à-dire par la *schématisation d'une large collection des différentes expressions faciales d'un seul individu* que l'individualité dans sa complexité est produite.

---

<sup>6</sup> Dans (Dondero, 2020), j'analyse les différentes sortes de compositions de la figure et du fond en considérant des dispositifs d'émergence ou de retrait du visage par rapport au fond.

Dans le cas du portrait pictural, c'était le contraire : il fallait produire une seule image qui puisse condenser les expériences de vie d'un individu, non seulement celles passées et présentes mais aussi celles futures, à savoir celles que son destin avait préparées pour lui.

Les choses se passent encore différemment chez Galton. Si Galton visait à atteindre la généralisation d'un type d'homme (le criminel, le Juif, etc.) en superposant de multiples photographies d'individus différents, les techniques du *deep learning* utilisées dans les vidéos *deepfake*, au contraire, visent à rendre compte de toutes les expressions d'un seul individu en produisant une *diagrammatisation de ses expressions*, c'est-à-dire une identité élastique qui peut être transférée sur le corps d'une autre personne agissant dans une variété de situations.

La superposition de visages par Galton, que nous pourrions appeler « synthèse d'identités multiples », était construite sur l'immobilité des modèles ; dans les cas de la production et de la détection de vidéos *deepfake*, la machine cherche à disposer de la plus large gamme de photos des mouvements, des gestes et des expressions faciales d'un individu pour arriver à une schématisation suffisamment plastique pour s'accorder à la variabilité des mouvements et des expressions faciales de la personne sur laquelle l'identité sera « implantée ».

En ce sens, le « processus du masque » qui a lieu lors de la production de vidéos *deepfake* fonctionne comme un diagramme au sens de Peirce où les virtualités potentielles sont à l'œuvre et qui, les diagrammes étant par définition des dispositifs manipulables, est capable de couvrir toute la variété des gestes possibles que le visage d'une autre personne pourra assumer. La fonction de flou du processus du masque consiste à tester une commensurabilité entre le modèle d'un individu et l'autre. Les *deepfakes* rendent ainsi possible le passage de la multiplicité d'expressions faciales d'un individu à la synthèse diagrammatique qui rendra l'identité visuelle de l'individu suffisamment disponible aux gestes d'un autre corps et aux expressions faciales d'un autre visage.

## Bibliographie

- Ambrosio Chiara (2016). Composite Photographs and the Quest for Generality. *Critical Inquiry*, 42, pp.547-579. Disponible à cette adresse : [https://discovery.ucl.ac.uk/id/eprint/1477167/1/Composite%20Photographs\\_OA.pdf](https://discovery.ucl.ac.uk/id/eprint/1477167/1/Composite%20Photographs_OA.pdf).
- Basso Fossali Pierluigi & Dondero Maria Giulia (2011). *Sémiotique de la photographie*, Pulim, Limoges.
- Beyaert-Geslin Anne (2017). *Sémiotique du portrait. De Dibutade au Selfie*, De Boeck Supérieur, Louvain-la-Neuve.
- Beyaert-Geslin Anne, Lloveria Vivien (2014). Une approche épisémotique de la présence. Sur quelques portraits d'adolescents. *Versus* 117, pp. 95-112.
- Beyaert-Geslin Anne (2021). *L'invention de l'Autre. Le Juif, le Noir, le(s) paysan(s), l'Alien*, Garnier, Paris.
- Chauviré Christiane (1995). *Peirce et la signification. Introduction à la logique du vague*, PUF, Paris.
- Deleuze Gilles (2003). *Francis Bacon: The Logic of Sensation*, Continuum, London.
- Dondero Maria Giulia (2020). *Les Langages de l'image. De la peinture aux Big Visual Data*, Éditions Hermann, Paris.
- Galton Francis (1878). Composite Portraits made by combining those of many different persons into a single figure. *Nature*, 18, pp.97-100. Disponible à cette adresse: <http://galton.org/essays/1870-1879/galton-1878-Composite-Portraits-Extended.pdf>.
- Galton Francis (1879). Generic Images. *Proceedings of the Royal Institution*, 9, pp.161-170. Disponible à cette adresse: <http://galton.org/books/generic-images/galton-1879-generic-images.pdf>
- Galton Francis (1883). *Inquiries Into Human Faculty and Its Development*, Macmillan, London. Disponible à cette adresse : <http://galton.org/books/human-faculty/text/galton-1883-human-faculty-v4.pdf>
- Leone Massimo (2021). From Fingers to Faces: Visual Semiotics and Digital Forensics. *International Journal for the Semiotics of Law*, 34, pp. 579-599. DOI: <https://doi.org/10.1007/s11196-020-09766-x>.
- Manovich Lev (1995). *From the Externalization of the Psyche to the Implantation of Technology*. Disponible à cette adresse : [http://manovich.net/content/04-projects/006-from-the-externalization-of-the-psyche-to-the-implantation-of-technology/04\\_article\\_1995.pdf](http://manovich.net/content/04-projects/006-from-the-externalization-of-the-psyche-to-the-implantation-of-technology/04_article_1995.pdf)

- Nguyen Thanh T. *et al.* (2019). Deep Learning for Deepfakes Creation and Detection, *arXiv.org*. Disponible à cette adresse: <https://arxiv.org/abs/1909.11573v1>.
- Peirce Charles Sanders (1895). Short Logic. *Collected Papers* (fragm.) 2.282, 2.286-91, 2.295-96, 2.435-443, 2.444, 7.555-558.
- Peirce Charles Sanders (1931-35). *Collected Papers of Charles Sanders Peirce*, 8 vol., Hartshore, Weiss, Burks dirs., Cambridge, Harvard University Press.
- Sabir Ekraam, *et al.* (2019). Recurrent convolutional strategies for face manipulation detection in videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 80-87. Disponible à cette adresse : <https://arxiv.org/abs/1905.00582>.
- Sekula Alan (1986). The Body and the Archive. *October*, 39, pp.3-64.