Miroir Magique : représentation virtuelle de la nouvelle apparence physique du patient oncologique après une chirurgie altérante du visage

par

Antoine DESSOLIES

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLEÀ L'OBTENTION DE
LA MAITRISE EN GENIE CONCENTRATION PERSONNALISÉE
M. Sc. A.

MONTRÉAL, LE 28 JANVIER 2026

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. Carlos Vàzquez, directeur de mémoire
Département de génie logiciel et TI à l'École de technologie supérieure

M. David Labbé, président du jury
Département de génie logiciel et TI à l'École de technologie supérieure

M. Simon Drouin, membre du jury
Département de génie logiciel et TI à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 18 DÉCEMBRE 2025

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEUR

REMERCIEMENTS

J’aimerai remercier mon directeur de recherche, Carlos, ainsi que mon co-directeur de recherche, Thierry, pour tous le temps qu’ils m’ont accordé et la patience qu’ils ont eu envers mon égard, ainsi que pour m’avoir donné la chance de pouvoir réaliser ma maitrise avec eux.

Merci beaucoup à tous les copains (nouveaux et anciens) du Canada, sans vous la maitrise n’aurait juste pas été possible, et plus particulièrement Mathieu et Eloi, qui m’ont sauvé, je n’aurai pas pu rêver de meilleurs colocs.

Je remercie également mes amis en France, pour ceux qui sont venus me voir : c’était incroyable et j’espère que vous avez autant aimé votre voyage que j’ai aimé votre visite. Pour ceux qui ne sont pas venus : vous me manquez et j’ai hâte de vous revoir.

Mille mercis à Arthur, Claire, Justine et Mattéo qui ont su garder ma santé mentale à flot malgré les obstacles. Je reviens bientôt pour vous voir !

Un dernier merci à ma famille, qui me guide et me réconforte même depuis 6000km, sans les hivers auraient été beaucoup plus rudes.

Miroir Magique : représentation virtuelle de la nouvelle apparence physique du patient oncologique après une chirurgie altérante du visage

Antoine DESSOLIES

RÉSUMÉ

Les patients atteints de cancer du visage subissant une ablation du nez font face à des défis psychologiques importants liés à la perte de leur identité visuelle. La méthode actuelle de création d'épithèses nasales est longue, coûteuse et intrusive, nécessitant plusieurs essais physiques. Ce mémoire présente le « Miroir Magique », une solution non invasive permettant aux patients de visualiser leur nouvelle apparence avec différentes épithèses virtuelles en temps réel.

La méthode développée combine les technologies 3DDFA V3 pour générer une texture réaliste du nez et 3DDFA V2 pour la reconstruction faciale rapide. Le suivi robuste du visage est assuré par BlazeFace via MediaPipe, capable de détecter les visages sans nez. Les algorithmes de recalage 3D, utilisant l'Analyse en Composantes Principales (ACP) couplée à l'algorithme ICP, permettent un positionnement précis des modèles d'épithèse. Le filtre Minilag garantit la fluidité des mouvements sans retard perceptible.

Les résultats expérimentaux démontrent une précision moyenne de placement de 9,70 pixels, une dérive inférieure à un pixel, et une latence moyenne de 23 ms respectant le seuil de confort pour les applications immersives. La robustesse du système a été validée sur des vidéos de visages avec et sans nez, respectant le critère de 95% du modèle d'épithèse dans la zone d'intérêt dans la majorité des cas. La vitesse de traitement actuelle (4,33 images par seconde) reste limitée, mais l'utilisation d'un processeur graphique permettrait d'atteindre les performances nécessaires au temps réel.

Le « Miroir Magique » offre une méthode innovante et respectueuse pour accompagner les patients dans le choix de leur épithèse, s'intégrant naturellement dans le projet AUDACE et ouvrant la voie à une nouvelle approche de création d'épithèses faciales adaptée aux besoins médicaux modernes.

Mots-clés : reconstruction faciale 3D, épithèse nasale, réalité augmentée, modèle déformable 3D, suivi de visage sans marqueurs, recalage 3D, ICP, analyse en composantes principales, filtrage de position

Magic Mirror : virtual representation of the new physical appearance of an oncological patient after an appearance altering surgery

Antoine DESSOLIES

ABSTRACT

Facial cancer patients undergoing nasal ablation face significant psychological challenges related to the loss of their visual identity. The current method for creating nasal prostheses is time-consuming, expensive, and invasive, requiring multiple physical fittings. This thesis presents "Magic Mirror," a non-invasive solution enabling patients to visualize their new appearance with different virtual prostheses in real-time.

The developed method combines 3DDFA V3 technologies for realistic nose texture generation and 3DDFA V2 for fast facial reconstruction. Robust face tracking is ensured by BlazeFace via MediaPipe, capable of detecting faces without noses. 3D registration algorithms, using Principal Component Analysis (PCA) coupled with the ICP algorithm, enable precise positioning of prosthesis models. The Minilag filter ensures smooth movements without perceptible delay.

Experimental results demonstrate an average placement accuracy of 9.70 pixels, drift below one pixel, and an average latency of 23 ms meeting the comfort threshold for immersive applications. System robustness was validated on videos of faces with and without noses, meeting the criterion of 95% of the prosthesis model within the region of interest in most cases. Current processing speed (4.33 frames per second) remains limited, but using a graphics processor would achieve the performance necessary for real-time operation.

"Magic Mirror" offers an innovative and respectful method for supporting patients in choosing their prosthesis, naturally integrating into the AUDACE project and paving the way for a new approach to facial prosthesis creation adapted to modern medical needs.

Keywords: 3D facial reconstruction, nasal prosthesis, augmented reality, 3D morphable model, markerless face tracking, 3D registration, ICP, principal component analysis, position filtering

TABLE DES MATIÈRES

Page

LISTE DES TABLEAUX

Page

LISTE DES FIGURES

Page

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

2D	Bidimensionnel
3D	Tridimensionnel
3DMM	Masque déformable tridimensionnel
ACP	Analyse en composantes principales
AVG	Moyenne
BFM	Basel Face Model
CHUM	Centre hospitalier de l'université de Montréal
ICP	Iterative closest point
PFH	Point feature histogram
RANSAC	Random sample consensus
RGB	Red Green Blue
RNC	Réseau de neurones convolutif
STD	Écart type

LISTE DES SYMBOLES ET UNITÉS DE MESURE

UNITÉS GÉOMÉTRIQUES

Angle plan

° degré

Longueur

mm millimètre

px pixel

UNITÉ DE TEMPS

ms milliseconde

INTRODUCTION

En 2022, 20 millions de nouveaux cas de cancers ont été détectés(Filho et al., 2025). Parmi ces patients, 3-4% sont atteints de cancers du visage ou du cou. Pour certains patients atteints de ce type de cancer, l'unique solution est l'ablation d'une partie du visage, qui leur laisse des séquelles irréversibles. Au Centre Hospitalier de l'Université de Montréal (CHUM), environ 400 patients par an sont traités pour des cancers de la tête, et une vingtaine doit subir une opération du visage altérante.

Lorsqu'un patient perd une partie de son visage, son identité telle qu'il la connaît est détruite (Destruhaut, Vigarios, Andrieu, & Pomar, 2011), et les retombées sur tous les aspects de sa vie (relationnel, financier, etc.) peuvent amener une grande détresse psychologique. Toutefois, un soutien psychologique, émotionnel et identitaire, bien que nécessaire n'est pas encore systématiquement mis à disposition des patients. (Jackson et al., 2023)

C'est pour répondre à ce besoin qu'a vu le jour le projet AUDACE, un projet qui souhaite accompagner les patients ayant récemment perdu une partie de leur visage, afin de rendre plus facile la reconstruction de leur nouvelle identité, et leur transition vers leur nouvelle vie. Nous nous intéressons ici plus particulièrement aux patients ayant perdu leur nez et souhaitant poser une épithèse (prothèse de visage)(Vigarios, Destruhaut, Pomar, Dichamp, & Toulouse, 2015). Cette prothèse ne nécessite aucune chirurgie et peut être placée directement sur le visage du patient, afin de remplacer la partie qui a été retiré à cause du cancer. Elle cherche à donner au patient l'apparence qu'il avait, avant son altération du visage.

Le projet est composé d'une équipe pluridisciplinaire, ainsi que de patients volontaires, ayant déjà effectué leur transition, et prêt à soutenir les nouveaux patients. A la suite de l'opération, la plupart des patients décident de poser une épithèse à la place de la partie qui a été retirée. Malheureusement, la création et la personnalisation des épithèses n'utilise pas les technologies les plus récentes, et ne permet pas au patient de modifier l'épithèse à son goût avant de la poser. En effet, la méthode de création d'épithèse actuellement mise en place consiste à demander à un artiste spécialisé dans la création de prothèse de nez de créer un nez à partir de l'apparence du patient avant la perte de son nez. Si le patient n'est pas satisfait de l'épithèse créée, il doit le signaler à son médecin qui doit ensuite le signaler à l'artiste, qui créé une nouvelle épithèse

suivant les remarques du patient, ce qui ralentit le processus.(Bannink et al., 2024) De plus cette méthode de création d'épithèse est intrusive car il faut placer l'épithèse dans la cavité nasale du patient afin de savoir si elle lui correspond ou pas.

Cette méthode est lente, invasive et coûteuse. Afin de l'améliorer, il faudrait donc une nouvelle méthode, l'automatiser le plus possible, et la rendre le plus facile d'accès possible au patient. Le projet « Miroir Magique » cherche donc à créer une méthode immersive, capable de reproduire la sensation de se regarder dans un miroir, en plaçant une épithèse virtuelle au bon endroit, en temps réel, sans avoir recours à des marqueurs faciaux, utilisés pour faciliter le suivi de visage (Gao, Lim, Lin, & Green, s.d.). Ainsi si le client n'est pas satisfait du modèle virtuel d'épithèse, il peut en choisir un autre et l'essayer tout de suite. Afin de rendre l'utilisation du « Miroir Magique » la plus simple possible, nous utiliserons uniquement des vidéos 2D en tant que jeu de données d'entrée. Pour cela, il est pertinent d'étudier des techniques telles que la reconstruction automatique du visage à partir d'une image, le suivi de mouvement ou la déformation d'un masque virtuel afin de l'adapter au patient.

Une fois le modèle d'épithèse sélectionné, il peut directement être imprimé en 3D, et sera ensuite disponible environ 24h après. Cette nouvelle méthode serait un gain de temps pour le patient ainsi que pour le médecin, un gain d'argent pour l'hôpital, et serait moins intrusif pour le patient.

Ce mémoire est organisé comme suit. Le chapitre 1 présente une revue de littérature sur les différentes solutions existantes se rapprochant du concept de « Miroir Magique ». Le chapitre 2 présente les objectifs et sous objectifs du projet, ainsi que sa contribution à l'état de l'art. Le chapitre 3 détaille la méthodologie utilisée pour la mise en place du « Miroir Magique ». Les résultats sont présentés dans le chapitre 4, et discutés dans le chapitre 5. Le mémoire se termine par une conclusion.

CHAPITRE 1

REVUE DE LITTÉRATURE

La littérature actuelle comporte déjà des méthodes qui se rapprochent du concept de « Miroir Magique » (Grishchenko et al., s.d. ; J. Guo et al., 2020 ; Thomas & Taniguchi, 2016), dans le sens ou les utilisateurs peuvent voir une reconstruction de leur visage dans un flux vidéo en direct sans avoir besoin de poser de marqueurs faciaux. Ces méthodes peuvent être divisées en deux catégories : utilisation des points caractéristiques ou reconstruction de visage directe. Les différences entre ces deux méthodes sont illustrées dans la Figure 1.1.

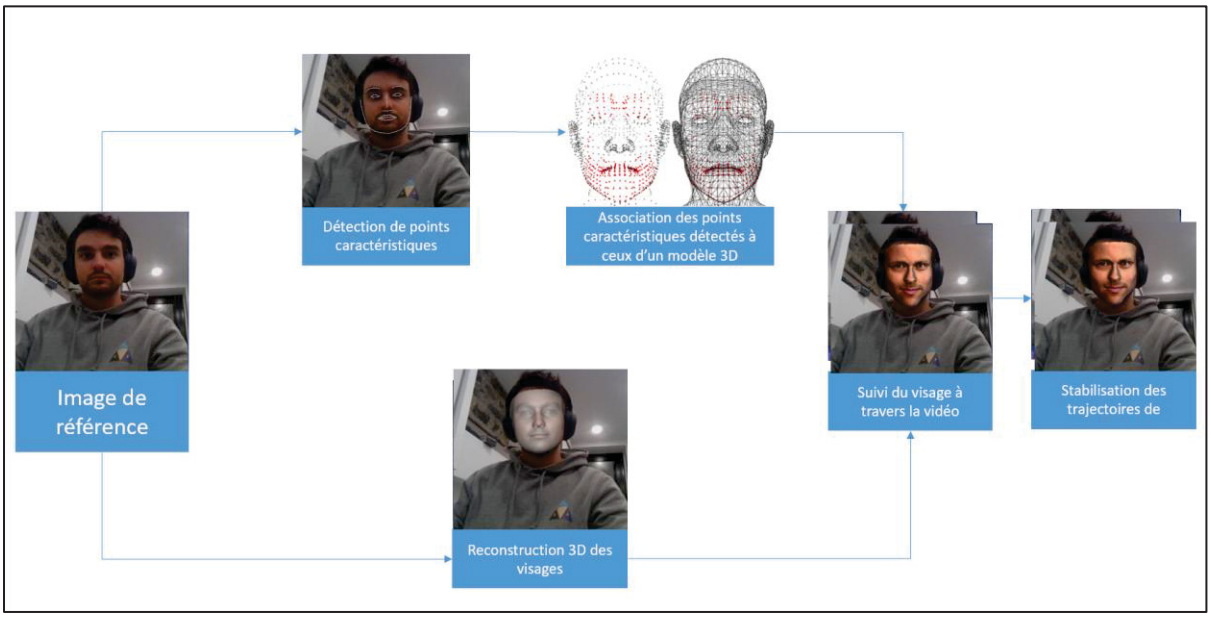

Figure 1.1 Illustration des différences entre les catégories principales de "Miroir Magique"

Ces méthodes, qui sont de plus en plus nombreuses, s'appuient pratiquement toutes sur l'intelligence artificielle et plus particulièrement les réseaux de neurones (Amirgaliyev, Mussabek, Rakhimzhanova, & Zhumadillayeva, 2025 ; Sharma & Kumar, 2022). Cet expansion peut s'expliquer par les différentes industries qui y portent attention, comme le cinéma et les jeux vidéo, qui utilisent la reconstruction de visage pour animer des personnages ou des dessins animés à partir de véritables visages d'acteurs (Shakir & Al-Azza, 2022); la

sécurité, qui utilisent les réseaux de neurones pour reconstruire, identifier, et traquer des visages(Kajalo & Lindblom, 2016), ou bien le médical (Nguyen, Nguyen, Dakpé, Ho Ba Tho, & Dao, 2022).

La revue de littérature traite d'abord des méthodes de « Miroir Magique » par utilisation de points caractéristiques puis par reconstruction 3D des visages directe. Elle traite ensuite des méthodes de suivi de visage et de filtrage des positions. Enfin, elle se termine par une revue des méthodes d'évaluation, et une synthèse générale.

1.1 Utilisation de points caractéristiques

1.1.1 Détection de points de repères

Dans cette partie de la revue de littérature, on définit « point de repère » des points d'intérêts sur le visage. Depuis leur création, les méthodes de détection de points de repère n'ont cessé d'évoluer : les méthodes holistiques (Baker & Matthews, 2004 ; Cootes, Edwards, & Taylor, 2001) ont été remplacées par l'utilisation de réseaux de neurones profonds, plus précis, et plus rapides (Y. Wu & Ji, 2019). Les développeurs peuvent utiliser leur propre méthode de détection de point de repère, ou bien utiliser des méthodes qui sont publiques et qui ont déjà été démontrées comme efficaces. Dans les deux cas on retrouve deux catégories majeures dans les méthodes de détection de points de repères : les régressions directes de coordonnées de points de repères, et l'utilisation de cartes thermiques, qui permettent également de trouver les coordonnées (Khabarlak & Koriashkina, 2022 ; Yin, Wang, Chen, Chen, & Liang, 2020).

La **méthode de régression directe** de coordonnées utilise un réseau de neurones convolutif pour extraire les caractéristiques importantes d'une image, pour finalement prédire les coordonnées de chaque point de repère sur l'image(Z.-H. Feng, Kittler, Awais, Huber, & Wu, 2018). Cette méthode est rapide et peu coûteuse en calcul comparée à celle des cartes thermiques, comme indiqué dans le Tableau 1.1, mais elle a une précision un peu inférieure selon (Khabarlak & Koriashkina, 2022).

Tableau 1.1 Comparaison de la précision et du temps de calcul des méthodes « Régression directe » et « Cartes thermiques »

Tiré de Khabarlak & Koriashkina (2022)

	Régression Directe	Cartes thermiques
Temps de calcul (ms)	1,2	4,0
Précision (Erreur Moyenne Normalisée %)	4,21	3,72

Elle est présente par exemple chez (Prados-Torreblanca, Buenaposada, & Baumela, 2022) et (X. Guo et al., 2019), qui la couple avec des graphs afin de conserver une structure de visage cohérente. De leur côté (Micaelli, Vahdat, Yin, Kautz, & Molchanov, 2023) utilisent un modèle d'équilibre profond, pour produire des prédictions cohérentes au fil du temps, tout en maintenant une complexité de calcul constante. (Y. Wu & Ji, 2015) ainsi que (Zhu, Shi, Zheng, & Sadiq, 2019) emploient également la régression de coordonnées directe en ajoutant un module capable de déterminer si les points de repères sont visibles ou pas, afin d'améliorer la robustesse et la précision des prédictions. (Xu, Li, Yuan, & Geng, 2021) introduisent une méthode fondée sur des points "ancres", dont les voisins sont connus. Ces points servent de modèle pour localiser les points environnants et estimer avec précision les coordonnées de l'ensemble des repères faciaux. Le domaine le plus étudié concernant la régression directe de coordonnées est sans aucun doute les fonctions de pertes des RNC, afin de trouver une fonction qui donne des résultats plus précis et plus robuste. La fonction la plus utilisée reste L2 (Z.-H. Feng et al., 2018), mais il y'a eu de grandes avancées pour promouvoir la fonction WingLoss (Z.-H. Feng et al., 2018) par exemple, qui améliore L1 en utilisant une fonction de perte basée sur un logarithme près de l'origine, ce qui permet à la fonction d'être moins sensibles aux erreurs petites. RetinaFace (Deng, Guo, Ververas, Kotsia, & Zafeiriou, 2020) est une méthode qui prédit en même temps la boite englobante de visage, la position des points de repères (par régression directe) et la position des sommets 3D du visage, projetés sur l'image.

La méthode de prédiction de points de repères en utilisant des **cartes thermiques** est une méthode qui utilise également des RNC, qui génèrent une carte thermique pour chaque point de repère. Les coordonnées probables de chaque point de repère sont ensuite extraites grâce à une fonction argmax (Khabarlak & Koriashkina, 2022). Les deux méthodes (Régression directe et Cartes thermiques) sont illustrées dans la Figure 1.2

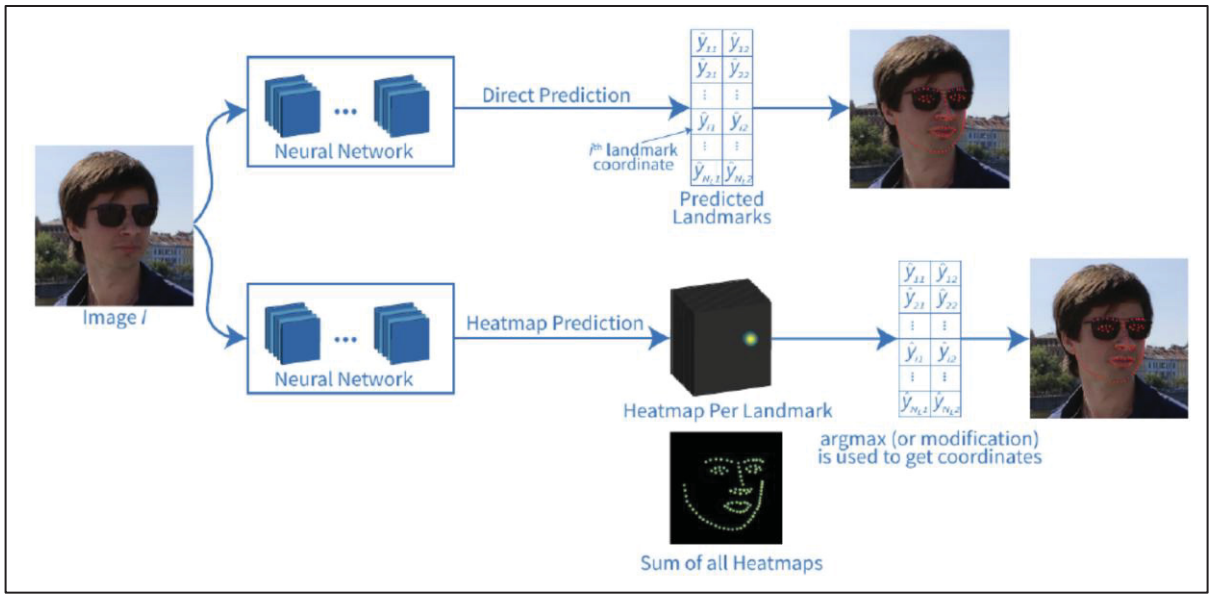

Figure 1.2 Schéma de l'utilisation de cartes thermiques et de la régression directe
Tirée de Khabarlak & Koriashkina (2022, p. 5)

Cette méthode est de plus en plus répandue dans la littérature (Wan et al., 2023) malgré sa complexité de calcul : la génération d'une carte thermique pour chaque point de repère peut devenir couteux en fonction du nombre de repère (Yin et al., 2020). En effet des nouvelles méthodes de détection de points de repères basées sur les cartes thermique sortent chaque année (Memon et al., 2025). La génération de cartes thermiques est au cœur de nombreuses méthodes, comme LUVLi (Kumar et al., 2020) , qui prédit la visibilité des points de repères en même temps que leur positions, ou PropagationNet (Huang, Deng, Shen, Zhang, & Ye, 2020), qui utilise les cartes thermiques pour dégager les contours du visages, et utilise ces contours comme région d'intérêt pour augmenter la précision des prédictions. (Jin, Liao, & Shao, 2021a) renforcent la robustesse des prédictions en y intégrant les prédictions des points de repères

voisins, et en combinant les cartes thermiques avec la régression directe de coordonnées, ce qui augmente également la précision.

Bien que plus rare, il existe également d'autres méthodes de détection de points de repères qui n'utilisent pas la régression directe de coordonnées ou bien les cartes thermiques. Il existe par exemple la détection de points de repères en utilisant des graphes, afin de conserver la relation de distance entre les points de repère (W. Li et al., 2020 ; Lin et al., 2021), qui permet de conserver une cohérence de forme et de renforcer la robustesse face aux occlusions. Une fois que les points de repères ont été détectés, ils sont mis en relation avec des modèles 3D de visage (ou tout autre avatar) afin de placer les modèles 3D à l'endroit adéquat, et de le modifier pour qu'il ait une expression correspondant à celle du visage de référence (Grishchenko et al., s.d.). Pour cela il est commun d'utiliser des Blendshapes.

1.1.2 Utilisation de Blendshapes

Les Blendshapes sont le produit d'un effort issu des années 1980 (Lewis et al., s.d.), qui cherchait à réduire la complexité d'animation d'un visage, pour cette raison, le client principal était l'industrie du cinéma, et plus particulièrement de l'animation. Ils sont présents par exemple dans la création du personnage « Gollum » dans le seigneur des anneaux (Lewis et al., s.d.).

Le fonctionnement des blendshapes est très simple, et très intuitif, ce qui a contribué à son succès dans le domaine de l'animation. Les blendshapes sont une base de modèles de visage 3D, qui représentent chacun une expression (neutre, sourire, demi-sourire, pleurs…). Ces modèles peuvent être scannés, ou bien créés par un ou plusieurs artistes, qui créent les visages sur des logiciels de modélisation. Les animateurs peuvent ensuite créer une nouvelle expression en effectuant une combinaison linéaire des autres, ce qui laisse une grande liberté à l'animateur, tout en l'empêchant de modifier la forme générale du visage qu'il anime, étant donné qu'il ne modifie que les expressions. Cela permet de s'assurer de toujours garder le visage d'un personnage intact à travers un film, tout en indiquant ce qu'il ressent. Au cours des années, les blendshapes se sont développés, et les modèles de visage 3D ont laissé place à des groupes de données qui correspondent aux déplacements des sommets composant les

blendshapes entre les modèles « expressifs » et le modèle neutre. Les groupes de données ont ensuite été triées par régions (par exemple, haut et bas du visage) afin de rendre l'utilisation encore plus simple, toutefois, le principe de fonctionnement est resté le même(Lewis et al., s.d.).

La plupart du temps, le modèle de visage est préparé en avance, comme dans (Ming, Li, Ling, Zhang, & Xu, 2025) et (Garrido, Valgaert, Wu, & Theobalt, 2013). Il peut provenir par exemple d'un scan précis d'un visage effectué au préalable, ou bien d'un modèle 3D créé entièrement par un artiste sur un logiciel de modélisation 3D. La personnalisation du visage peut également être automatisée, comme dans (Thomas & Taniguchi, 2016) où les auteurs récupèrent un modèle 3D de visage grossier à partir d'une caméra RGBD, puis affinent les traits du visage en utilisant les données de profondeur de la caméra, et en incluant des déplacements selon les normales de chaque point du modèle 3D de visage pour correspondre au mieux aux données de profondeur. Ils utilisent également ces données de profondeurs pour réussir à positionner correctement le blendshape dans la vidéo. (Ming et al., 2025) et (Garrido et al., 2013) utilisent une méthode de détection des points de repères caractéristiques (yeux, bouche, sourcils…) sur le visage, qui sont présentées dans le chapitre 1.3, puis les font correspondre aux points correspondants sur le blendshape, relevés au préalable, ce qui permet de placer correctement le blendshape vis-à-vis de l'image. Cela leur permet également de déformer le modèle 3D pour représenter l'expression du visage.

1.2 Reconstruction de visage directe

L'autre méthode pour créer un « Miroir Magique » consiste à créer un modèle 3D du visage de référence pour ensuite le placer dans la vidéo.

1.2.1 Génération d'un modèle de visage 3D grâce à un réseau de neurones génératif

Les réseaux de neurones peuvent être utilisés pour créer directement des modèles 3D correspondant aux visages présents dans l'image. C'est ce qu'ont fait (Qin et al., 2024) dans un projet qui vise à reconstruire des modèles 3D de visages entiers à partir de scan de visage dont ils ont supprimé la région du nez. Les auteurs ont récupéré 400 scans 3D de visages sains,

ont enlevé la région du nez, et ont entraîné un réseau de neurones génératif à recréer le nez du patient à partir de son visage incomplet. Le réseau est basé sur une structure encodeur-décodeur qui leur permet d'obtenir des résultats précis : en moyenne 1.45 ± 0.24 mm d'écart moyen 3D entre le nez reconstruit et le nez réel. L'objectif de ce dernier projet est proche de celui de (Y. Li et al., 2024), qui complète des scans 3D de visage. Les auteurs utilisent un réseau de neurone pour créer le maillage manquant dans le masque, puis optimisent la position des différents sommets du maillage pour ajuster les visages complets aux données incomplètes.

(Meng et al., 2022) quant à eux reprennent une structure encodeur-décodeur, en y apportant des modifications. Le programme commence en analysant une image de visage grâce à un réseau de neurones convolutif (RNC), qui extrait les caractéristiques de l'image puis, dans la partie décodeur, utilise un réseau convolutif de graph, et les caractéristiques extraites par le RNC, pour générer les positions des points d'un modèle 3D de visage.

En temps normal, les détecteurs de points caractéristiques ne génèrent pas assez de points pour pouvoir construire un modèle 3D de visage, à partir du visage présent dans l'image. C'est le constat dont sont parti (Wood et al., 2022) pour créer leur propre réseau capable de créer un modèle 3D en se basant uniquement sur la détection de points caractéristiques. Pour cela ils ont créé une base de données synthétique, c'est-à-dire entièrement générée par une intelligence artificielle génératrice, ce qui garantit le nombre de points de repères nécessaire à la création du modèle 3D. De plus les réseaux de neurones pour la reconstruction de visage 3D entrainés sur des données synthétiques performent mieux que ceux entrainés sur des données réelles selon (Wood et al., 2022). Ces approches, bien que prometteuses, requièrent des ressources considérables tant pour la création de bases de données volumineuses (400 scans 3D pour (Qin et al., 2024)) que pour l'entraînement de réseaux complexes encodeur-décodeur, limitant ainsi leur déploiement à grande échelle.

1.2.2 Utilisation de modèles surfaciques 3D déformables

Le 3D Morphable Mask (modèle surfacique 3D déformables), ou 3DMM, est un modèle génératif pour la forme et l'apparence du visage(Blanz & Vetter, 1999). Il est composé d'une base de données de visages avec une correspondance point par point entre chaque visage. Cette correspondance permet d'effectuer des combinaisons linéaires des visages de la base de données (combinaisons linéaires des coordonnées des points correspondants) afin d'obtenir un visage cohérent. Le 3DMM permet aussi de découpler la couleur et la forme du visage, ce qui permet de rendre ces paramètres indépendants des facteurs extérieurs comme l'illumination ou les paramètres de caméra. Ils créent également des modèles 3D de visage complet malgré les occlusions et les déformations de visage. C'est sur ce constat que se sont basés (Nguyen et al., 2022), qui utilisent les 3DMM pour reconstruire un visage de patients atteints de paralysie faciale, afin de pouvoir créer un programme de réhabilitation personnalisé. Le programme est capable de créer un modèle 3D correspondant à la forme du visage malgré l'asymétrie provoquée par la paralysie faciale.

Les créateurs des 3DMM se sont inspirés d'EigenFaces par (Sirovich & Kirby, 1987). EigenFaces considère les images de visage 2D comme un espace vectoriel, dont il est possible d'extraire les composants principaux (vecteurs propres). Les visages peuvent être ensuite représentés comme des combinaisons linéaires de ces composants principaux. Les créateurs du 3DMM appliquent ce principe à des modèles de visage 3D au lieu d'image de visage 2D, comme illustré dans la Figure 1.3.

Dans un premier temps (Blanz & Vetter, 1999) le réalisent en découplant la forme et la couleur du visage, afin de les isoler des facteurs externes tels que l'illumination ou les paramètres de caméra. Ainsi, les visages peuvent être représentés sous la forme :

$$S = \bar{S} + A_{id}\alpha + A_{color}\beta \qquad (1.1)$$

Avec

S : le nouveau visage

$\bar{S}$: l'élément neutre (visage neutre)

A_{id}/A_{color} : les bases de la forme / couleur

α/β : les coefficients de forme / couleur

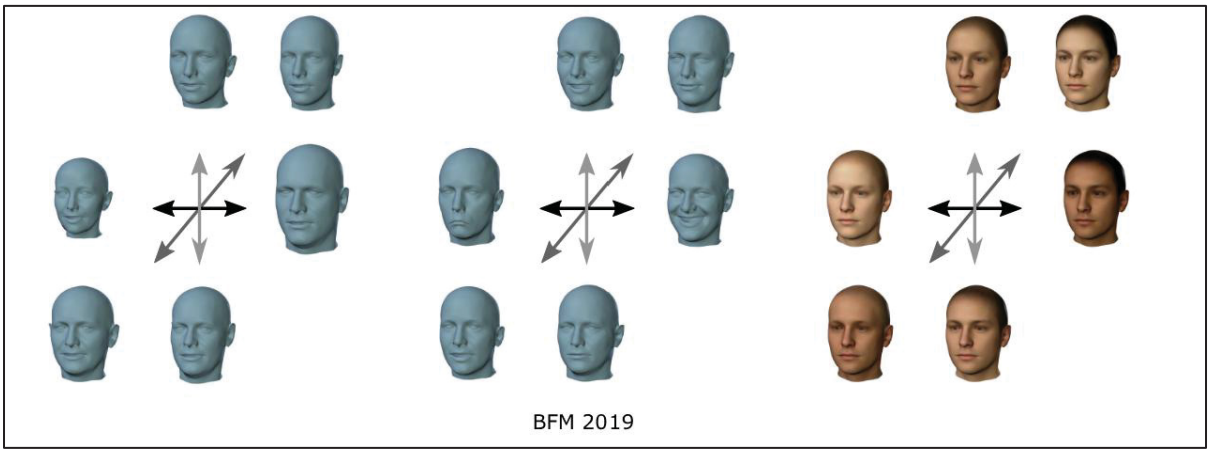

Figure 1.3 Représentation des variations principales de l'espace des visages
Tirée de B. Egger et al. (2020, p. 14)

En 2009 est publiée la première base de données publique de 3DMM : Basel Face Model (BFM) (Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009), qui a depuis eu 2 mises à jour en 2017 (Gerig et al., 2018) (BFM 2017) et 2019 (« Models | Shape modelling @ University of Basel », s.d.) (BFM 2019), et il s'agit encore aujourd'hui de la base de données la plus utilisée dans le cadre des 3DMM malgré l'apparition de quelques autres comme (T. Li, Bolkart, Black, Li, & Romero, 2017),(L. Wang et al., 2022),(Booth, Roussos, Ponniah, Dunaway, & Zafeiriou, 2018) et (Huber et al., 2016). Au fil du temps, les bases de données se sont améliorées et il est maintenant possible d'ajuster l'expression du visage grâce aux 3DMM.

L'utilisation des 3DMM est particulièrement intéressante lorsqu'elle est couplée avec des réseaux de neurones profonds, entrainés à prédire les coefficients corrects et l'orientation des visages détectés dans une image, afin d'en créer des modèles 3D. Cette méthode peut également être utilisée sur des images consécutives pour avoir un modèle de visage 3D depuis une vidéo. C'est le cas chez (Y. Guo, zhang, Cai, Jiang, & Zheng, 2019) qui utilise un 3DMM avec 3 RNC qui traitent tous des niveaux de caractéristiques différents, ce qui permet d'accélérer le processus jusqu'à atteindre une performance en temps réel. (Tiwari, Kurmi, Venkatesh, & Chen, 2022) quant à eux, couplent l'utilisation du 3DMM avec un réseau qui

prend en compte le « contexte » de l'image (l'environnement) afin de réduire les effets des occlusions.

Les utilisations sont nombreuses et pluridisciplinaire, les 3DMM sont utilisés avec des réseaux de neurones profonds dans la reconnaissance faciale, le montage vidéo ou la modification de visage (Diao, Jiang, Fan, Li, & Wu, 2024). Les plus connus de ces programmes sont la gamme des 3DDFA (X. Zhu, Liu, Lei, & Li, 2019),(J. Guo et al., 2020) et (Z. Wang, Zhu, Zhang, Wang, & Lei, 2024). Plus particulièrement les deux plus récents : 3DDFAV2 (J. Guo et al., 2020) et 3DDFAV3 (Z. Wang et al., 2024). 3DDFAV2 a pour objectif de créer un modèle 3D correspondant à un visage, en temps réel, à partir d'une image, d'une vidéo, ou bien d'un flux vidéo en temps réel contenant le visage de référence.

Pour cela, 3DDFAV2 utilise un réseau de neurones profond léger (MobileNet (Howard et al., 2017)) qui prédit les coefficients d'expression, de forme, et la pose du modèle 3D (orientation et position), à partir d'une image redimensionnée par un détecteur de visage (ici il s'agit de Retinaface (Deng et al., 2020)). Cette image redimensionnée permet de fournir seulement le visage à MobileNet, qui est entrainé sur des images avec seulement des visages. Afin d'être le plus rapide possible, l'auteur a effectué une analyse en composante principale sur la base de données BFM. Il s'agit d'un outil statistique (ACP) qui permet de trouver les composants les plus importants d'un jeu de données, il est présenté plus en détails dans le chapitre 1.3.4. L'ACP permet d'extraire les 40 paramètres de forme et 10 paramètres d'expression les plus importants (les paramètres de couleurs ne sont pas pris en compte). Cela permet d'augmenter significativement la vitesse de calcul de la forme générale du modèle 3D à obtenir, mais signifie également la perte de détails du visage. Étant donné que seuls les paramètres les plus importants ont été conservés, ceux qui permettaient d'affiner le modèle 3D pour correspondre au visage de référence, ne sont plus calculés.

Les auteurs modifient la base de données BFM en ajoutant des images de synthèses proches des images données, afin de créer des courtes vidéos, pour inciter le réseau de neurones à prédire des modèles 3D de visage cohérents temporellement, c'est-à-dire qui ont une forme identique, et seuls les paramètres de pose et d'expressions changent. Lors de sa publication en 2020, il s'agissait du programme en temps réel le plus performant en terme de vitesse, de précision de forme et de placement, comme indiqué dans le tableau Tableau 1.2 où 3DDFAV2

est comparé à PRNet (Yao Feng, Wu, Shao, Wang, & Zhou, 2018), l'algorithme qui était jusque-là le plus performant en précision de reconstruction de visage par utilisation de 3DMM dans des vidéos. Les résultats sont évalués sur la base de données Menpo-3D (Zafeiriou et al., 2017), qui fournit un ensemble d'images et de séquences de visages annotés avec des repères faciaux tridimensionnels cohérents, dans des conditions variées de pose et d'éclairage issue de 300-VW (Shen et al., 2015).

Tableau 1.2 Comparaison des résultats de 3DDFAV2
et PRNet selon le standard de Menpo-3D

	3DDFA V2	PRNet
Vitesse (ms)	2.1	9.8
Erreur Moyenne Normalisée (%)	1.71	1.90

Il est à noter que l'auteur de 3DDFA V2 propose une version du modèle 3D avec de la texture, mais il ne s'agit pas de la texture prise en charge par BFM, mais d'une extraction de la texture du visage directement depuis l'image, puis plaquée sur le modèle 3D. Cela implique que si le visage est caché par une occlusion, la couleur de l'objet cachant le visage sera considérée comme faisant partie de la couleur du visage. Il est intéressant de remarquer qu'une solution de suivi existe dans le 3DDFAV2, qui permet de réutiliser le modèle 3D de visage créé à la première image et de le déplacer au bon endroit avec la bonne expression dans les images suivantes. Pour cela 3DDFAV2 créé une zone d'intérêt à partir des sommets du modèle de visage reconstruit dans l'image précédente, puis utilise cette zone pour trouver les paramètres de 3DMM correspondants, sans changer les paramètres de forme, afin de garder une cohérence de forme.

3DDFAV3 (Z. Wang et al., 2024) quant à lui a un tout autre objectif : obtenir le modèle le plus détaillé possible et correspondant de la façon la plus précise possible au visage visible dans l'image. Pour cela il se base sur ResNet50 (He, Zhang, Ren, & Sun, 2016) pour prédire les paramètres du 3DMM (BFM) ainsi que la position du modèle à partir de l'image. Le modèle

3D de visage est ensuite segmenté selon une sémantique cohérente avec la segmentation 2D standard, comme illustré dans la Figure 1.4, où la segmentation 2D standard est tirée de (Zheng, Deng, Zhu, Li, & Zafeiriou, 2022).

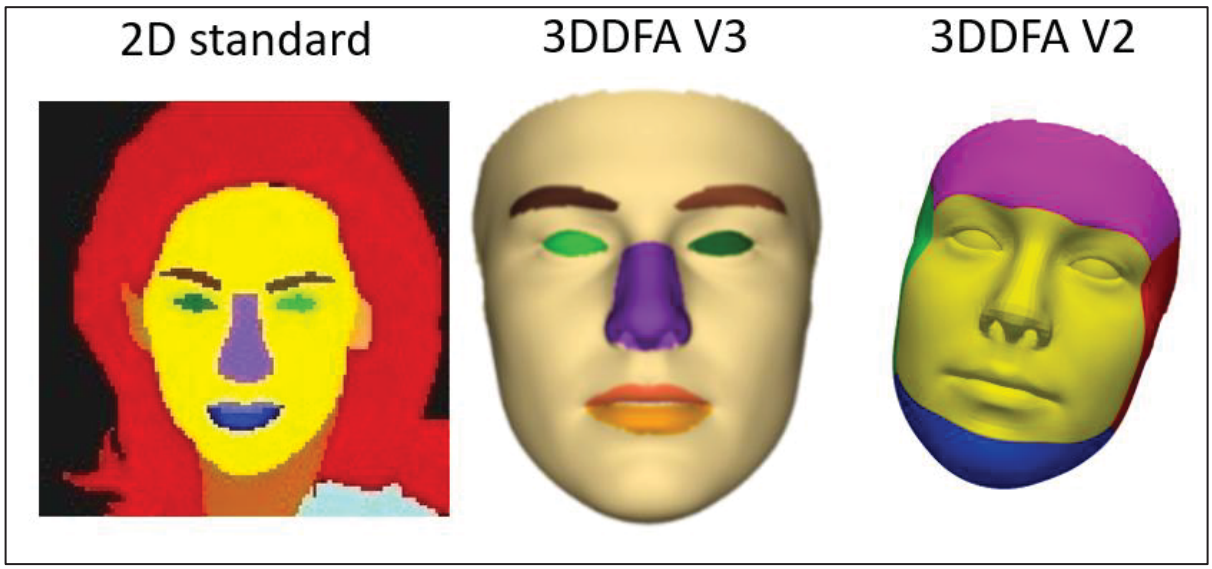

Figure 1.4 Visualisation de la segmentation 2D standard, 3DDFAV3 et 3DDFAV2

Les sommets du modèle 3D sont ensuite projetés sur le plan de l'image et les segmentation 3D et 2D sont comparées et ajustées par un réseau de neurone régit par une fonction de perte qui a pour objectif principal de minimiser la différence de distance statistique entre les groupes de points 3D (projetés) et 2D correspondants. Cette méthode est plus lente que 3DDFAV2 (5.8 secondes contre 2.1ms) et n'est pas adaptée aux vidéos, mais elle donne des modèles de visage 3D beaucoup plus précis comme illustré dans la Figure 1.5 et indiqué dans Tableau 1.3 où les deux méthodes ont été testées selon le standard REALY (Chai et al., 2022). Dans ce tableau « avg » signifie moyenne et « std » signifie écart type.

Tableau 1.3 Comparaison des résultats de 3DDFAV2 et 3DDFAV3 selon le standard REALY

Méthode	Vue Frontale (mm)				
	Nez avg±std	Bouche avg±std	Front avg±std	Joue avg±std	Moyenne
3DDFA V2	1.903±0.517	1.597±0.478	2.447±0.647	1.072±0.333	1.537
3DDFA V3	1.586±0.306	1.238±0.373	1.810±0.394	1.111±0.327	1.436

Figure 1.5 Image de référence et modèles de visages issus de 3DDFA V2 et 3DDFA V3

La couleur du modèle 3D issu de 3DDFA V3 est insensible aux occlusions comme illustré dans la Figure 1.6 qui montre un modèle 3D de visage créé à partir de l'image d'une personne sans nez, ce qui est très intéressant pour obtenir un rendu semi-réaliste.

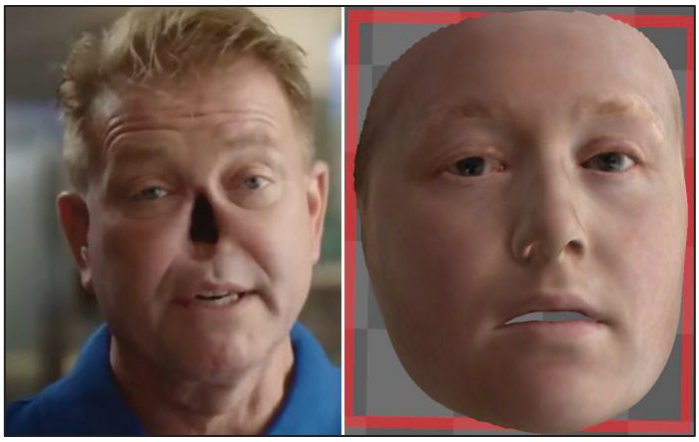

Figure 1.6 Reconstruction de visage avec texture malgré l'absence de nez (droite)

1.3 Suivi de visage dans une vidéo

Pour gagner de la vitesse de calcul, si deux images sont assez similaires, il est possible de déplacer le modèle de visage sans en recréer de nouveau. Pour cela le suivi de visage est crucial. Le suivi d'objet (et par conséquent le suivi de visage) dans une vidéo repose sur deux grandes catégories d'algorithmes : le suivi par méthode continue, et le suivi par détection (Luo et al., 2021).

1.3.1 Suivi de visage par méthodes continues

Le suivi de visage par méthode continue consiste à suivre des caractéristiques définies lors d'une initialisation, à travers une vidéo (Ross, Lim, Lin, & Yang, 2008). Une de ces méthodes est le suivi par flux optique, qui consiste à suivre une zone d'intérêt en mesurant la différence d'intensité entre les pixels de deux images consécutives (Beauchemin & Barron, 1995). Elle est utilisée pour suivre des zones d'intérêt du visage comme (Ranftl, Alonso-Fernandez, & Karlsson, 2015). Une autre méthode très connue est la méthode développée par (Lucas & Kanade, 1981), qui est un suivi par modèle. L'objectif de cette méthode est d'aligner au mieux une image « source » avec une image « cible », en utilisant un gradient d'intensité spatiale et un algorithme de minimisation pour trouver une bonne correspondance. Cette méthode est utilisée pour suivre des parties de visages qui ont été extraites auparavant comme le fait (Kim,

Jang, & Choi, 2007), qui extrait des zones d'intérêt du visage en utilisant l'algorithme Kanade-Lucas-Tomasi (Tomasi, 1991), qui lui-même permet d'extraire des zones faciles à suivre. Les méthodes de flux optique et de Lucas-Kanade peuvent également être combinées pour obtenir des suivis de visage temporellement cohérent, comme dans (Dong et al., 2018) qui utilise le flux optique pour suivre des points de repères du visage, et Lucas-Kanade pour affiner la prédiction et assurer une cohérence temporelle. Les problèmes principaux des méthodes de suivi continu sont qu'elles n'arrivent pas à suivre les visages lors d'occlusions ou lors de changement d'illumination ou de pose du visage extrême au cours de la vidéo (Chrysos, Antonakos, Snape, Asthana, & Zafeiriou, 2018). Pour résoudre ce problème il faut se tourner vers une autre méthode de suivi de visage : le suivi par détection.

1.3.2 Suivi de visage par détection

Le suivi de visage par détection contient deux catégories principales : le suivi de visage par détection des points de repères, qui est présenté dans le chapitre 1.1.1, et la détection de visage sans points de repères.

1.3.3 Détection de visage sans les points de repères

Vers le début des années 2000, beaucoup de méthodes de reconstruction de visage et de détection de points de repères s'appuyaient sur le détecteur de visage (Viola & Jones, 2001). Celui-ci extrait des caractéristiques simples dans l'image, les classe via des classifieurs entraînés, puis organise ces classifieurs en cascade pour localiser les zones de l'image susceptibles de contenir un visage. Cette méthode pose les bases de la détection de visage moderne et sera utilisé par une majorité des méthodes de détection de visage par la suite (Wijaya et al., 2025).

Les autres méthodes de détection du visage sont variées, certains auteurs, comme (Saito, Li, & Li, 2016), utilisent un classifieur pour déterminer les pixels faisant parti du visage dans l'image d'entrée et en extraire un masque binaire indiquant les parties de l'image où le masque est visible, ce qui permet même de segmenter le visage dans les images. De son côté (J. Guo et al., 2020) utilise une fonction de perte spécialisée sur des réseaux de neurones convolutifs profonds

qui permettent de détecter et extraire une zone générale où se trouve le visage (zone d'intérêt). De son côté Google a développé Mediapipe (Lugaresi et al., s.d.), un cadre spécialisé pour l'utilisation d'apprentissage machine directement depuis les données fournies par une caméra. La détection de visage dans ce cadre est effectuée par le modèle BlazeFace (Bazarevsky, Kartynnik, Vakunov, Raveendran, & Grundmann, 2019), un réseau de neurones qui a été conçu pour fonctionner sur téléphone, qui est donc léger et rapide (0.6ms sur un iPhone XS) et une très bonne précision (98.61% de précision moyenne, avec un seuil standard de 0,5 pour l'intersection sur union des boîtes englobantes). La détection de visage via MediaPipe est assez rapide pour fonctionner en temps réel, et assez robuste pour pouvoir détecter le visage dans n'importe quelle situation, ce qui en fait un excellent candidat pour notre application. Mediapipe peut envoyer les boîtes englobantes au programme de reconstruction 3D, qui peut ensuite générer un modèle 3D de visage compris dans la boîte englobante.

1.3.4 Recalage de deux modèles 3D de visages entre eux

Une fois que le visage est détecté dans la nouvelle image, il est possible d'y superposer le modèle de visage de l'image précédente. Pour cela il faut utiliser le recalage de modèles 3D.

La méthode de recalage la plus répandue est sans aucun doute l'ICP (Iterative Closest Point) de (Besl & McKay, 1992). L'ICP est une méthode itérative qui permet d'aligner deux modèles 3D avec des formes similaires, à partir d'une première estimation grossière de leur alignement (Tam et al., 2013). Si l'estimation initiale de l'alignement des deux modèles à recaler n'est pas assez bonne, l'ICP risque de ne pas converger (Mitra, Gelfand, Pottmann, & Guibas, 2004). Cette estimation peut être trouvée grâce à des données extérieures (par exemple des marqueurs (Allen, Curless, & Popović, 2002)), les modèles peuvent être déplacés à la main par un utilisateur (H. Li, Adams, Guibas, & Pauly, 2009) , ou bien utiliser un algorithme de placement approximatif, présenté plus bas (Liu, Zhou, Yang, Deng, & Liu, 2014). La seconde étape de l'ICP consiste à trouver les points les plus proches dans le modèle cible, pour chaque point du modèle source. Pour cela l'algorithme utilise en général un arbre k-dimension (Bentley, 1975),

qui accélère grandement la recherche de plus proches voisins. Les voisins sont associés par paires (un point sur le modèle source et son voisin le plus proche sur le modèle cible) qui seront utilisées pour trouver la transformation qui superpose les deux modèles.

L'étape suivante consiste à trouver une transformation optimale afin de réduire le plus possible la distance entre chacune des paires. Cette transformation est la solution du problème d'optimisation des moindres carrés :

$$\left(\hat{R}, \hat{t}\right) = arg \min_{R,t} \sum \left|\left|(R \cdot p_i + t) - q_i\right|\right|^2 \quad (1.2)$$

Avec p et q les points associés entre eux

Afin d'optimiser ce processus il est utile d'utiliser la méthode de décomposition en valeur singulières (Sorkine-Hornung & Rabinovich, s.d.).

Une fois que la transformation est calculée, le processus est répété de façon itérative jusqu'à atteindre un minimum local ou une condition d'arrêt (exemple : un nombre d'itération maximum) (Bouaziz, Tagliasacchi, Li, & Pauly, 2016).

Depuis sa création, de nombreuses variantes sont apparues pour compenser les points faibles de l'ICP, pour pouvoir écarter les points aberrants (auxquelles la méthode des moindres carrés est sensible) (Youyang Feng, Wang, & Zhang, 2019), pour réduire le coût de calcul (Zhang, Yao, & Deng, 2022), ou, comme évoqué précédemment, pour automatiser l'initialisation.

Pour résoudre ce dernier problème il est devenu de plus en plus commun d'utiliser la méthode de recalage global (Global registration) (Rusu, Blodow, & Beetz, 2009). Étant donné que le recalage global vise à aligner deux modèles 3D qui sont dans des positions et des poses inconnues, mais avec des formes similaires, il faut dans un premier temps réussir à trouver des caractéristiques qui ne varient pas en fonction de la pose. Ces caractéristiques peuvent être calculée grâce au "Persistent Feature Histogram" (PFH) (Rusu, Blodow, Marton, & Beetz, 2008), un algorithme qui détermine les caractéristiques d'un modèle 3D grâce aux normales des points, qui peuvent être estimées, ou données par l'utilisateur si elles sont disponibles (dans le cas d'un scan par infrarouge par exemple). Le PFH calcule la géométrie aux alentours de chacun des points grâce aux normales des plus proches voisins (trouvé à l'aide d'un arbre k dimensions) et établit un histogramme pour chacun des points, c'est à dire une empreinte de la

distribution de la géométrie autour de chaque point, qui comprend la distance entre les points, les normales et autres éléments géométriques. Une fois une empreinte générée pour chaque point, tous les points ainsi que leur empreinte sont compris dans un espace de k dimensions. Un arbre K dimensionnel est utilisé pour établir une “carte” de l’espace, qui nous permettra ensuite de faire correspondre les points du nuage source et du nuage cible. C’est ici qu’intervient RANSAC (Random Sample Consensus) (Fischler & Bolles, 1981), un algorithme itératif qui sélectionne au hasard des points du nuage source, cherchent leur plus proche voisin dans l’espace crée par le PFH, à l’aide de l’arbre K dimensionnel, puis calcule la transformation optimale pour que ces points soient au plus proche de leur correspondance. La transformation est ensuite évaluée en comparant les histogrammes du reste des points qui n’ont pas été sélectionnés, aux points le plus proche d’eux dans l’espace 3D. La méthode est répétée jusqu’à ce que le nombre d’itération maximum soit atteint, ou qu’une transformation satisfaisante soit trouvée. Afin de réduire le nombre de transformation à calculer, il est coutume d’ajouter une opération de « rognage » qui élimine les résultats incohérents (exemple : si les modèles ne sont pas situés à côté). Si un critère de rognage n’est pas rempli, alors la transformation est rejetée sans avoir besoin de la valider sur le reste de l’ensemble des données des nuages de points.

Dans de rares cas, il est possible d’avoir déjà une liste de points qui correspondent entre le nuage cible et le nuage source (par exemple des marqueurs) (Allen et al., 2002), il est alors beaucoup plus facile d’aligner les modèles 3D car il suffit de calculer la transformation qui fera correspondre les points puis l’appliquer au reste des points. Cette transformation est la solution du problème d’optimisation des moindres carrés :

$$(R, t) = arg \min_{R,t} \sum \left|\left|(R \cdot p_i + t) - q_i\right|\right|^2 \tag{1.3}$$

Avec p et q les points associés entre eux

Ce processus ressemble à l’ICP mais la paire de points exacte est déjà connue, il n’y a donc pas besoin de plusieurs itérations.

Lorsque les modèles 3D présentent des directions claires de répartition des points, il est possible de recourir à une Analyse en Composantes Principales (ACP) (Pearson, 1901) pour identifier les axes dominants de cette répartition. L’ACP est une méthode statistique qui consiste à dégager les axes principaux de répartition des points d’un ensemble de données.

Mathématiquement l'ACP calcule une matrice de covariance entre les données pour en dégager les relations les plus fortes entre les variables. Il faut ensuite effectuer des combinaisons linéaires de variables afin d'obtenir la variance la plus élevée possible, ces combinaisons linéaires sont non corrélées et sont appelées composantes principales (la direction de répartition des points dans l'exemple précédent). L'ACP est utilisée par (C.-W. Wang & Peng, 2021) pour initialiser le recalage de plusieurs nuages de points représentant des visages. Pour cela une ACP est appliquée à chaque point et ses voisins afin d'en extraire des directions normales ainsi que la courbure principale, calculée par :

$$\sigma = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} \tag{1.4}$$

Avec λ_1, λ_2, λ_3, les vecteurs propres issus de l'ACP.

Ces caractéristiques sont ensuite comparées à des références jusqu'à trouver des points caractéristiques comme le bout du nez. Ces points servent ensuite de références pour créer un premier alignement grossier avant d'utiliser l'ICP pour l'affiner.

1.4 Filtrage des positions d'objets virtuels

Dans le domaine de réalité virtuelle ou réalité augmentée, afin que l'immersion soit réussie au mieux, il est essentiel d'obtenir le placement le plus précis possible d'un objet virtuel, sans vibration, qui est un phénomène où l'objet adopte plusieurs positions proches dans des images successives, dû à l'incertitude sur sa position (Batmaz & Stuerzlinger, 2023). La littérature rapporte plusieurs approches de filtrage, dont le but est de prédire au mieux quelle sera la position de l'objet à l'image suivante, afin de placer l'objet dans l'image actuelle en conséquence. L'un de ces filtres est le filtre de « moyenne mobile » (Zhuang, Chen, Wang, & Lian, 2007). Ce filtre prédit la position actuelle en effectuant la moyenne des k dernières positions. Bien qu'efficace pour convertir des signaux simples, ce dernier n'est pas adapté pour des signaux avec des variations brusques et fréquentes, ce qui est le cas pour une personne qui découvre sa nouvelle apparence avec une épithèse par exemple. Pour remédier à cela, il existe le filtre exponentiel simple (Ahuja, Ofek, Gonzalez-Franco, Holz, & Wilson, 2021), qui associe des poids aux données en fonction de leur nouveauté, plus précisément, si la donnée

est récente, elle aura plus de poids. Le problème principal de ce filtre est qu'il n'utilise pas les valeurs précédentes au-delà de l'avant dernière, ce qui l'empêche de prendre en compte les mouvements lents.

Les filtres de Kalman (Kalman, 1960) permettent d'estimer des valeurs de sortie en se basant sur les données d'entrée disponibles et sur la précision associée à ces données. En effet le filtre utilise les matrices de covariance issues du périphérique d'entrée, ainsi qu'une estimation de la donnée actuelle, comparée à la mesure, pour déterminer la fiabilité de la nouvelle mesure et son poids dans la création de la nouvelle valeur filtrée. La valeur estimée est souvent créée à partir d'un modèle linéaire, bien qu'il existe des filtres de Kalman plus évolués qui sont capables d'utiliser des modèles non linéaires (Welch & Bishop, 1995). Ces filtres sont encore aujourd'hui utilisé pour filtrer les positions de méthodes de détections de points caractéristiques comme dans (Dong et al., 2018).

En revanche ces filtres utilisent la matrice de covariance des périphériques d'entrée, qui n'est pas toujours accessible et doit donc être estimée avant l'utilisation du filtre, et changée si l'outil qui relève les données change.

Les avancées les plus récentes ont amené au filtre Minilag (Song et al., 2023) basé sur le 1€ filter (Casiez, Roussel, & Vogel, 2012) qui lui-même est un filtre du premier ordre passe bas dont la fréquence de coupure change. L'objectif du filtre Minilag est de concevoir un filtre destiné aux applications de réalité augmentée, capable de lisser les positions des modèles 3D afin de rendre leurs déplacements plus fluides, tout en évitant tout retard entre la position filtrée et la position non filtrée comme illustré dans la Figure 1.7. En effet l'utilisation d'un filtre pour éliminer le bruit d'une courbe peut introduire un « retard » entre la courbe filtrée et la courbe brute. Il faut donc réaliser un compromis entre la suppression de bruit et le retard induit par le filtre (Song et al., 2023).

L'objectif de Minilag correspond parfaitement à nos critères pour réduire les vibrations des modèles 3D et prolonger l'immersion du patient. Comme dit précédemment, un retard peut apparaître lors de l'utilisation d'un filtre passe bas. Cela arrive si la fréquence de coupure est trop petite, et supprime les mouvements du modèle 3D qui ont une fréquence plus haute.

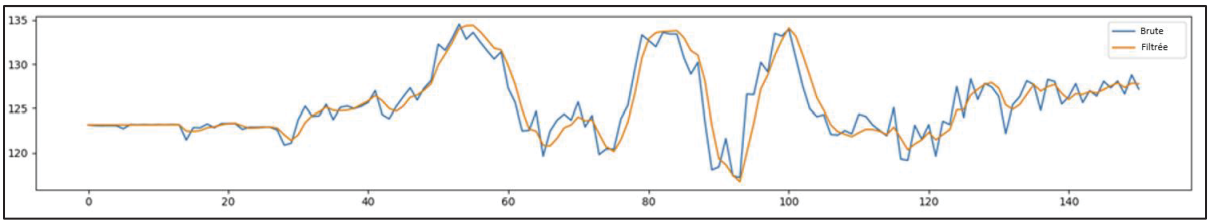

Figure 1.7 Courbe brute et courbe filtrée de la coordonnée en x d'un modèle 3D en fonction du temps

Minilag repose donc sur une équation de filtre passe bas avec une fréquence de coupure qui s'adapte aux mouvements du modèle :

$$\hat{P}_t = \alpha P_t + (1-\alpha)\hat{P}_{t-1} \tag{1.5}$$

$$\alpha = \frac{1}{1+\frac{\tau}{T_e}} \tag{1.6}$$

$$\tau = \frac{1}{2\pi f_c} \tag{1.7}$$

$$f_c = f_{cmin} + \beta \left|\dot{\hat{P}}_i\right| \tag{1.8}$$

Ici P_t représente la position du modèle à l'instant t, f_{cmin} représente la fréquence de coupure minimum et β est un coefficient d'influence des données précédentes sur la valeur actuelle du filtre. Le filtre augmente sa fréquence de coupure lorsque les mouvements sont grands et brusques, mais la fréquence est basse lorsqu'il n'y a pas de mouvements, ce qui permet de supprimer le bruit.

Afin d'améliorer la précision du filtre, les auteurs introduisent trois nouveautés principales. Dans un premier temps, le filtre Minilag utilise une « mise à jour en retour sur ses pas », c'est-à-dire qu'il met à jour les données qui ont déjà été filtrées auparavant. Cette mise à jour permet d'avoir des données plus précises pour la prédiction de la nouvelle donnée. Afin d'effectuer cette mise à jour, il faut ajuster une courbe polynomiale d'ordre 3 aux n dernières données non filtrées, puis l'utiliser afin de mettre à jour la dernière donnée filtrée.

$$\hat{P}_{t-1} = \varphi \hat{P}_{t-1} + (1-\varphi) P_{t-2}^{Ajusté} \tag{1.9}$$

Avec φ un coefficient d'ajustement.

Il est intéressant de noter qu'il n'est pas possible d'utiliser directement $P_{t-1}^{Ajusté}$ car les données sur les extrêmes des courbes ne sont pas assez fiables.

Dans un second temps, ce nouveau $\hat{P}_{t-1}$ est utilisé pour prédire $\hat{P}_t$ grâce à l'équation issue du 1euro filter. Les auteurs sont partis du principe qu'une partie du retard était due à la différence entre la valeur originale et la valeur filtrée.

$$\Delta P_t = P_t - \hat{P}_t \tag{1.10}$$

Cette valeur est ensuite utilisée pour compenser le retard, or P_t comporte du bruit, il faut donc filtrer ΔP_t également.

$$\Delta\hat{P}_t = \gamma\Delta P_t + (1-\gamma)\Delta\hat{P}_{t-1} \tag{1.11}$$

Avec γ un paramètre d'ajustement.

Il est important de noter que $\hat{P}_{t-1}$ est mis à jour dans l'étape précédente, il faut donc également mettre à jour $\Delta\hat{P}_{t-1}$:

$$\Delta\hat{P}_{t-1} = \eta\Delta\hat{P}_{t-1} + (1-\eta)(P_{t-1} - \hat{P}_{t-1}) \tag{1.12}$$

Avec η un paramètre d'ajustement

Une fois que ΔP_t a été filtré, le retard sur P_t peut enfin être compensé:

$$\hat{P}_t = \hat{P}_t + \Delta\hat{P}_t \tag{1.13}$$

Ce qui donne la prédiction finale de Pt.

La troisième et dernière nouveauté est la façon dont le filtre traite les rotations d'un modèle 3D. En effet les matrices de rotations influent aussi sur la forme et la taille de l'objet, en filtrant les éléments, il existe donc un risque de déformer l'objet, ce qui n'est pas désiré. Pour éviter cela, les auteurs ont décidé d'utiliser l'algèbre de Lie (Hall, 2015) , une façon de représenter les rotations sans risquer de déformer l'objet. Pour utiliser l'algèbre de Lie il faut tout d'abord établir un lien entre les matrices de rotations et l'algèbre de Lie. Les matrices de rotations peuvent être changées en éléments de l'algèbre de Lie grâce à la fonction exponentielle. Il existe donc avec R la matrice de rotation (matrice de pose) à l'instant t :

$$R = \exp(\dot{\omega}) \tag{1.14}$$

$$\dot{\omega} = \ln(R) \tag{1.15}$$

La matrice $\dot{\omega}$ est une exponentielle de matrice, il s'agit donc d'une matrice antisymétrique, sous la forme :

$$\dot{\omega} = \begin{pmatrix} 0 & W_1 & W_2 \\ -W_1 & 0 & W_3 \\ -W_2 & -W_3 & 0 \end{pmatrix} \quad (1.16)$$

Un vecteur de dimension 3x1 en est extrait afin de filtrer moins d'éléments :

$$\omega = (W_1 \quad W_2 \quad W_3) \quad (1.17)$$

Le filtrage doit être effectué dans un espace euclidien, l'algèbre de Lie n'est euclidien qu'autour de son élément neutre ([0 0 0]). Il est donc nécessaire d'obtenir des vecteur ω_t proches de l'élément neutre. Afin de respecter cette condition, il faut filtrer la différence de pose entre l'image précédente et l'image actuelle au lieu de filtrer la pose de l'objet :

$$\delta R_t = R_t * R_{t-1}^T \quad (1.18)$$

$$\delta\dot{\omega}_t = \ln(\delta R_t) \quad (1.19)$$

Les éléments de $\delta\omega_t$ sont ensuite filtrés, puis les transformations dans le sens inverse sont appliquées afin d'obtenir la matrice de pose adaptée.

Cette nouvelle méthode permet d'obtenir un placement d'objet virtuel précis et stable, sans induire de retard dans les déplacements, ce qui répond aux problèmes soulevés par les filtres précédents. A notre connaissance cette méthode n'est pas encore utilisée dans une méthode de suivi ou de reconstruction de visage, mais est spécialisée dans le suivi d'objet 3D.

1.5 Métriques d'évaluation

Les applications de génération de modèles 3D sont généralement testées avec des métriques telles que l'erreur absolue moyenne, l'erreur moyenne carrée, l'erreur moyenne normalisée… (Sharma & Kumar, 2022). Toutes ces métriques reposent sur une base de données annotée, dans laquelle les reconstructions peuvent être comparées à des données précises. Dans notre cas il n'existe pas de base de données vidéos de patient ayant perdu leur nez, dont on connait la position et l'orientation du nez. La validation de la méthode est donc beaucoup plus difficile.

La **latence** « motion-to-photon » est une métrique utilisée dans les applications de réalité virtuelle qui consiste à mesurer le temps entre le mouvement effectué et sa représentation dans l'application. (Carmack, 2013) indique que la latence commence à être remarquable à partir de 20 ms pour les application en réalité virtuelles et recommande de ne pas dépasser 50 ms, tandis que (Stauffert, Niebling, & Latoschik, 2020) indique que les humains peuvent détecter une latence de 17ms. Selon (Stauffert et al., 2020), une trop grande latence pourrait causer un inconfort chez les utilisateurs d'application de réalité augmenté et virtuelle. Ce seuil est difficile à quantifier car il dépend de chaque utilisateur.

La **robustesse** peut être quantifiée grâce à la Dérive (Engel, Schöps, & Cremers, 2014), qui est l'accumulation d'erreur au cours d'une opération de suivi, qui peut faire dévier un modèle 3D de sa cible. Elle peut être calculé sur une vidéo de visage fixe, où le modèle est donc sensé être immobile, grâce à :

$$Drift = \left\| p_t - p_{t_0} \right\| \tag{1.20}$$

Avec *pt* la position du modèle 3D à l'image finale, et p_{t0} la position à la première image

La **vibration** peut être calculé grâce à une formule donnée par (Song et al., 2023), à condition d'avoir une base de donnée de vidéos de visage dont la position et l'orientation sont connues. La vibration peut alors être calculée grâce à la formule :

$$J_R = \frac{1}{3n}\sum_{i=1}^{n}\left(\left(\omega_1^{i\,\prime\prime}\right)^2 + \left(\omega_2^{i\,\prime\prime}\right)^2 + \left(\omega_3^{i\,\prime\prime}\right)^2\right) \tag{1.21}$$

$$J_t = \frac{1}{3n}\sum_{i=1}^{n}\left(\left(t_1^{i\,\prime\prime}\right)^2 + \left(t_2^{i\,\prime\prime}\right)^2 + \left(t_3^{i\,\prime\prime}\right)^2\right) \tag{1.22}$$

Ou J_R (respectivement J_t) est la vibration en rotation (respectivement translation), ω_i (respectivement t_i) représente chacun des éléments du vecteur de rotation (respectivement translation) et ω'' (respectivement t'') est la dérivée seconde de ω (respectivement t).

1.6 Synthèse Générale

Cette revue de littérature a permis d'identifier les principales approches disponibles pour la création d'un « Miroir Magique » permettant aux patients ayant perdu leur nez de visualiser leur apparence avec une épithèse virtuelle. Les méthodes existantes peuvent être regroupées en deux grandes catégories : l'utilisation de points caractéristiques avec des blendshapes, et la reconstruction 3D directe du visage.

1.6.1 Limites des approches par points caractéristiques

Les méthodes basées sur la détection de points de repères faciaux (Khabarlak & Koriashkina, 2022 ; Y. Wu & Ji, 2019) présentent des limitations importantes dans le contexte de patients présentant des difformités faciales. Bien que les réseaux de neurones profonds aient considérablement amélioré la précision de détection sur des visages sains, avec des taux d'erreur normalisés aussi bas que 3,72% pour les méthodes par cartes thermiques (Khabarlak & Koriashkina, 2022), ces performances se dégradent significativement en présence d'occlusions importantes ou de déformations faciales atypiques. Les méthodes de blendshapes (Grishchenko et al., s.d. ; Lewis et al., s.d.), bien qu'offrant une grande flexibilité artistique et une représentation sémantique intuitive des expressions faciales, nécessitent soit une calibration préalable avec des marqueurs faciaux (Thomas & Taniguchi, 2016), soit des correspondances précises entre les points détectés et le modèle 3D (Garrido et al., 2013 ; Ming et al., 2025). Cette dépendance à une détection fiable des points de repères constitue un obstacle majeur pour leur application aux patients sans nez.

1.6.2 Avancées et défis des modèles surfaciques 3D déformables (3DMM)

Les modèles surfaciques 3D déformables (3DMM) se sont imposés comme la méthode dominante pour la reconstruction faciale 3D depuis les travaux fondateurs de (Blanz & Vetter, 1999). En découplant la forme et la texture du visage, et en utilisant l'analyse en composantes principales (ACP) pour représenter les variations faciales par combinaisons linéaires, les 3DMM offrent une robustesse théorique aux occlusions partielles. Cette propriété, mise en

évidence dans le domaine médical par (Nguyen et al., 2022) pour la reconstruction de visages asymétriques dus à une paralysie faciale, suggère une applicabilité potentielle aux patients sans nez.

Parmi les implémentations récentes, deux méthodes se distinguent par des objectifs complémentaires. Le 3DDFA V2 (J. Guo et al., 2020) privilégie la vitesse de traitement (2,1 ms par image) et la cohérence temporelle sur des séquences vidéo, grâce à une réduction du nombre de paramètres de forme (40) et d'expression (10) via ACP appliquée au Basel Face Model (BFM). Cette optimisation permet un fonctionnement en temps réel avec une erreur moyenne normalisée de 1,71% sur la base Menpo-3D, tout en préservant la cohérence de forme entre images successives. Toutefois, cette réduction de dimensionnalité se traduit par une perte de détails fins du visage. À l'opposé, le 3DDFA V3 (Z. Wang et al., 2024) vise la précision maximale en intégrant une segmentation sémantique 3D alignée sur les standards 2D, permettant une reconstruction détaillée avec une erreur moyenne de 1,436 mm selon le standard REALY (Chai et al., 2022). Cette méthode génère également une texture faciale robuste aux occlusions, reconstituant la couleur du nez même en son absence physique. Cependant, son temps de traitement (5,8 secondes par image) et son optimisation pour des images isolées limitent son utilisation directe dans un contexte vidéo en temps réel.

La littérature révèle ainsi une dichotomie entre vitesse/cohérence temporelle et précision/qualité de texture, aucune méthode existante ne combinant l'ensemble de ces propriétés essentielles pour une application clinique immersive.

1.6.3 Enjeux du suivi de visage pour les patients sans nez

Le suivi de visage dans des vidéos représente un défi distinct de la reconstruction 3D initiale. Les méthodes continues, telles que le flux optique (Beauchemin & Barron, 1995) ou l'algorithme de Lucas-Kanade (Lucas & Kanade, 1981), présentent une sensibilité importante aux changements d'illumination, aux occlusions et aux poses extrêmes (Chrysos et al., 2018). Ces limitations sont exacerbées dans le contexte de patients présentant des difformités faciales, où les zones d'intérêt traditionnelles peuvent être absentes ou modifiées.

Les approches par détection répétée offrent une robustesse supérieure en détectant le visage dans chaque image indépendamment. Les détecteurs basés sur les caractéristiques de Haar (Viola & Jones, 2001) ont posé les fondements de cette approche, mais les réseaux de neurones modernes comme BlazeFace (Bazarevsky et al., 2019) atteignent désormais des performances remarquables avec une vitesse de traitement de 0,6 ms et une précision de 98,61% d'IoU. Cette robustesse face aux variations de pose et d'apparence constitue un avantage crucial pour les applications médicales nécessitant une grande fiabilité.

1.6.4 Problématique du recalage 3D pour le suivi

Le suivi de visage par reconstruction 3D répétée nécessite un recalage efficace entre les modèles 3D successifs. L'algorithme ICP (Iterative Closest Point) de (Besl & McKay, 1992) demeure la référence pour l'alignement de modèles 3D de formes similaires, mais requiert une estimation initiale d'alignement suffisamment précise pour éviter la convergence vers des minima locaux (Mitra et al., 2004). Le recalage global (Rusu et al., 2008), utilisant les histogrammes de caractéristiques persistantes (PFH) couplés à RANSAC, automatise cette initialisation mais au prix d'un temps de calcul incompatible avec le temps réel.

L'ACP offre une alternative pour l'extraction rapide de la pose de modèles 3D présentant des directions claires de répartition de points. (C.-W. Wang & Peng, 2021) ont démontré son efficacité pour l'alignement initial de nuages de points faciaux en calculant les axes principaux de distribution et la courbure locale. Cette approche, bien que produisant un alignement approximatif, peut être raffinée par ICP pour obtenir une précision finale satisfaisante dans des temps compatibles avec le traitement vidéo.

Dans le cas de modèles 3D ayant une topologie constante, comme ceux générés par un même 3DMM, le recalage peut être simplifié en utilisant directement des correspondances de points connues. Cette propriété permet un calcul direct de la transformation optimale sans itérations multiples, réduisant significativement le temps de traitement.

1.6.5 Nécessité du filtrage pour la stabilité visuelle

La précision de reconstruction 3D, même avec les méthodes les plus avancées, présente toujours des variations légères entre images successives. Ces micro-variations, invisibles sur des images fixes, se manifestent sous forme de vibrations lors de l'affichage vidéo, perturbant l'immersion de l'utilisateur (Wilmott, Erkelens, Murdison, & Rio, 2022). Les filtres de lissage traditionnels, tels que la moyenne mobile (Zhuang et al., 2007) ou le filtre exponentiel simple (Ahuja et al., 2021), réduisent ces vibrations mais introduisent soit un retard perceptible, soit une réponse inadéquate aux mouvements rapides.

Le filtre de Kalman (Kalman, 1960) et ses variantes non linéaires (Welch & Bishop, 1995) offrent une solution théoriquement optimale en modélisant explicitement l'incertitude des mesures. Cependant, leur application nécessite la connaissance ou l'estimation des matrices de covariance des capteurs, information rarement accessible dans les systèmes de vision par ordinateur.

Le filtre Minilag (Song et al., 2023), développé spécifiquement pour les applications de réalité augmentée, propose une approche novatrice combinant un filtre passe-bas adaptatif, une mise à jour rétroactive des données filtrées, une compensation active du retard, et un traitement spécialisé des rotations via l'algèbre de Lie. Cette méthode élimine les vibrations tout en maintenant un retard imperceptible entre signal brut et signal filtré, répondant ainsi aux exigences contradictoires de stabilité et de réactivité pour une expérience immersive.

1.6.6 Lacunes identifiées dans la littérature

Malgré l'abondance de recherches en reconstruction et suivi de visages, deux lacunes majeures émergent pour l'application visée. Premièrement, très peu d'études considèrent les difformités faciales importantes : seuls (Nguyen et al., 2022) pour la paralysie faciale et (Qin et al., 2024) pour la reconstruction de nez à partir de scans 3D abordent cette problématique. Deuxièmement, aucune méthode publiée ne combine simultanément : (1) la génération de texture réaliste pour le nez en présence d'une absence physique, (2) un suivi de visage robuste en temps réel sur des visages atypiques, (3) la stabilité visuelle nécessaire à une expérience

immersive, et (4) un fonctionnement sans marqueurs faciaux pour une utilisation clinique non invasive.

La reconstruction de nez par réseaux génératifs de (Qin et al., 2024), bien qu'atteignant une précision remarquable de 1,45 ± 0,24 mm, repose sur des scans 3D nécessitant un équipement spécialisé coûteux et un temps d'acquisition incompatible avec une interaction en temps réel. De plus, aucune indication n'est donnée sur la capacité de ce système à fonctionner en mode interactif. Ces limitations soulignent le besoin d'une approche alternative basée sur l'imagerie 2D accessible (caméra standard) tout en maintenant une qualité de reconstruction suffisante pour l'évaluation clinique d'épithèses.

CHAPITRE 2

PROBLÉMATIQUE ET OBJECTIFS DU PROJET

2.1 Problématique et objectifs

La méthode actuellement utilisée pour que les patients choisissent l'épithèse qui leur convient, consiste à demander à un artiste de créer une épithèse selon les besoins du patient, faire essayer cette épithèse au dit patient, puis l'ajuster, ou en créer une nouvelle, selon les remarques et le niveau de satisfaction du patient (Ruse, Calhoun, & Davis, 2024). Cette méthode est longue, couteuse et intrusive, les épithèses devant être positionnées directement dans la cavité nasale. Cependant l'essai d'épithèse est une étape essentielle dans le processus d'apprivoisement de la nouvelle identité d'un patient à la suite de la perte de son nez.

L'objectif principal de ce projet est donc de proposer une méthode pour que les patients ayant perdu leur nez aient l'occasion d'avoir un aperçu de leur nouvelle apparence, sans utiliser de méthode intrusive. L'objectif est de simuler la découverte de leur nouvelle apparence devant un miroir, afin qu'ils puissent observer en temps réel les angles qui les intéressent ou les préoccupent, et évaluer leur future épithèse comme si elle était déjà posée sur leur visage.

Pour cela, le projet peut être divisé en sous-objectifs :

- Avoir un suivi de visage précis, c'est-à-dire qui superpose le modèle de visage virtuel reconstruit à partir d'une image du patient, à celui du patient dans le rendu vidéo. On se fixe pour objectif une précision de 10 pixels, lorsque l'utilisateur se trouve à environ 30 centimètres de la caméra.
- Avoir un suivi de visage robuste : le programme doit fonctionner sur une grande variété de visages, les patients ayant un visage inhabituel et une grande diversité de visage parmi les patients. On cherche à ce que le modèle de nez soit correctement placé sur au moins 95% des images de vidéos de divers patients. Le nez est déclaré comme « correctement placé » lorsqu'au moins 95% des points qui le composent sont dans la zone définie par les yeux et la bouche.
- Avoir un suivi de visage fluide : éviter les vibrations des modèles 3D supérieures à 1 millimètre, ou les dérives supérieures à 5 pixels. Si une seule des deux conditions

précédentes n'est pas respectée, le patient pourrait être sorti de son immersion, et cela réduirait l'impact bénéfique de la méthode.

- Besoin de faire tourner le programme en temps réel, idéalement à 30 images par secondes, afin de simuler la découverte devant un « miroir » et immerger le patient dans l'expérience avec une latence inférieure à 30 ms qui correspond à un compromis entre les deux valeurs extrêmes de la revue de littérature (Stauffert et al., 2020).

2.2 Contribution

L'application d'un programme de suivi de visage n'a, à notre connaissance, pas été exploré dans le cadre de patient ayant subi une perte de leur nez. Le projet qui s'en rapproche le plus est (Qin et al., 2024), qui utilisent les scans 3D de patients ayant perdus leurs nez pour recréer automatiquement des modèles de prothèse.

La proposition d'une nouvelle méthode non intrusive, c'est-à-dire sans avoir recours à des marqueurs faciaux, pour apercevoir la nouvelle apparence d'un patient avec une épithèse virtuelle en temps réel est quant à elle innovante et rien ne s'en approche dans la littérature actuelle. Le programme reconstruit un visage malgré une absence physique de nez, extrait la texture du visage pour l'appliquer sur le modèle de nez et affiche le nez sur le visage du patient, avec une latence inférieure à 30ms, dans un flux vidéo en direct de 30 images par secondes.

CHAPITRE 3

MÉTHODOLOGIE

3.1 Introduction

La revue de littérature indique qu'il existe deux méthodes principales pour créer un « Miroir Magique » : l'utilisation de points caractéristiques et la reconstruction directe de visage. Une étude préliminaire (3.2) démontre qu'il est plus simple d'utiliser la méthode de reconstruction directe de visage. Deux méthodes de reconstruction directe présentes dans la littérature se démarquent du reste. Il s'agit de 3DDFA V2, qui est capable de générer des modèles 3D de visage grossier mais très rapidement, et 3DDFA V3 qui est capable de créer des modèles 3D de visage très précis et avec de la texture, mais qui est trop lent pour être utilisé en temps réel. On relève également le filtre Minilag, qui est adapté au filtrage de position d'objets tridimensionnels. L'objectif de ce chapitre est de combiner les points forts de ces méthodes afin de créer une créer une application qui répond à nos objectifs de précision, robustesse, fluidité et vitesse de calcul.

3.2 Étude préliminaire

Tous les résultats de la littérature ont été obtenus en réalisant des expériences sur des visages complets. Il a donc été nécessaire d'effectuer une étude préliminaire simple, afin de vérifier si l'absence de nez impacte négativement les performances des différentes méthodes utilisées dans la littérature.

Dans un premier temps les méthodes évaluées sont les méthodes de détection de points caractéristiques de (Jin, Liao, & Shao, 2021b ; W. Li et al., 2020 ; Lin et al., 2021 ; X. Wang, Bo, & Fuxin, 2019 ; W. Wu et al., 2018 ; Xie, Wan, Shen, & Lai, 2021) . Les méthodes sont utilisées sur une vidéo de patient ayant perdu son nez (*Man Who Lost His Nose To Cancer Gets Incredible Prosthetic Nose | Body Parts*, 2022) et les résultats sont qualitatifs.

Aucune des méthodes utilisé n'a réussi à donner des points de repères de façon précise (particulièrement pour le nez) tout au long de la vidéo comme illustré dans la Figure 3.1, qui représente le suivi de (Micaelli et al., 2023).

Figure 3.1 Exemple de suivi de visage imparfait,
le visage est détecté en bas à droite

Dans un second temps, ce sont les méthodes de reconstruction directe de visage qui sont évaluées (J. Guo et al., 2020 ; Z. Wang et al., 2024 ; C.-Y. Wu, Xu, & Neumann, 2021). Ces dernières arrivent à créer des modèles 3D de visage malgré l'absence de nez. Ces résultats préliminaires nous permettent de conclure que les méthodes de reconstruction directe sont plus adaptées pour l'utilisation prévue de notre « Miroir Magique ».

L'étude préliminaire a également permis de mettre en évidence des problèmes lors de l'utilisation de méthodes de reconstruction directe. La fonction de suivi de visage de 3DDFA V2 n'est pas capable de suivre le visage en cas de mouvement brusque (ce qui peut arriver dans notre cas), à partir du moment où le visage n'est plus détecté, le programme met plusieurs images à superposer le modèle 3D avec le visage à nouveau. Ce problème est dû au code de 3DDFAV2, où le visage est cherché dans une zone d'intérêt. Si le visage n'est plus dans la zone d'intérêt, à cause d'un mouvement brusque ou d'une pose extrême par exemple, le programme prédit des paramètres pour un visage qui n'existe pas, et la condition pour rechercher une nouvelle zone d'intérêt prend plusieurs itérations à être atteinte. Une fois la

condition atteinte, le programme recherche une nouvelle zone d'intérêt dans l'image grâce à RetinaFace (Deng et al., 2020). Ce problème peut être résolu en obligeant le programme à chercher une nouvelle zone d'intérêt à chaque itération. Malgré la mise en place de la solution précédente, le programme n'arrive pas à détecter correctement le visage de patients sur certaines images avec des poses extrêmes, le détecteur de visage RetinaFace est donc remplacé par BlazeFace (Bazarevsky et al., 2019), qui est plus robuste.

3.3 Introduction et présentation générale de l'application « miroir magique »

La revue de littérature nous indique actuellement aucune méthode disponible libre de droit qui soit capable de remplir tous les critères sélectionnés. Nous allons donc utiliser plusieurs méthodes existantes et leur apporter les modifications nécessaires pour obtenir le résultat désiré.

Nous nous reposons tout d'abord sur la méthode 3DDFAV2 pour la détection et la reconstruction de visage, ce qui nous permet de créer un modèle 3D de visage à chaque image, assurant le suivi et la reconstruction de visage.

Nous utilisons également la méthode 3DDFAV3 lors de l'initialisation afin d'extraire une texture du visage, et plus particulièrement du nez, qui est ensuite transférée sur le modèle d'épithèse, afin d'obtenir un rendu semi réaliste. Le modèle d'épithèse, c'est-à-dire le modèle 3D de nez du patient, est ensuite affiché à l'endroit adéquat pour obtenir un rendu semi réaliste du visage du patient avec une épithèse virtuelle.

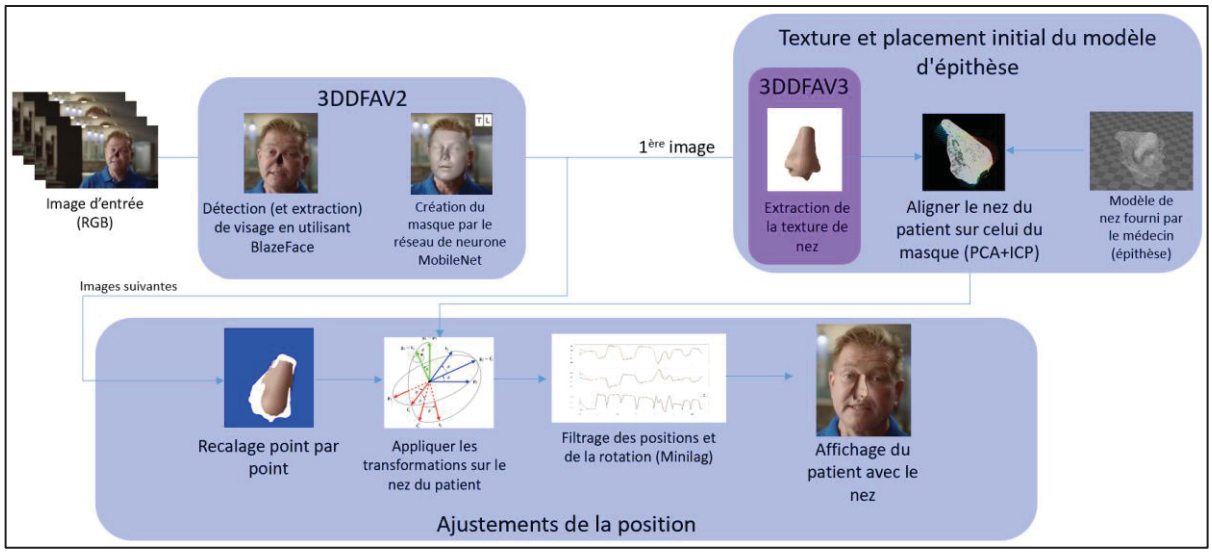

Figure 3.2 Schéma du fonctionnement de la méthode utilisée

Description générale de la méthode (illustrée dans la Figure 3.2) :

- Flux vidéo découpé en images d'entrées (avec le visage du patient)
- Première image de la vidéo où un visage est présent : **création de la texture et placement initial du modèle d'épithèse** (3DDFAV3)
    - Reconstruction 3D du modèle de visage
        - Détection du visage en utilisant RetinaFace, à partir de l'image d'entrée (3DDFAV2)
        - Création du modèle de visage par MobileNet (3DDFAV2)
    - Plaquage de la texture du nez issu de 3DDFAV3 sur le modèle de nez donné par le médecin (l'épithèse)
    - Extraire la zone du nez du modèle 3D de visage généré par 3DDFAV2
    - Trouver et appliquer la transformation qui aligne le modèle d'épithèse donné par le médecin sur celui extrait du modèle de visage généré par 3DDFAV2
    - Affichage uniquement du modèle d'épithèse (le modèle de visage est sauvegardé pour l'itération suivante)
- **Après initialisation** (images suivantes) :
    - Reconstruction 3D du modèle de visage

        - Détection du visage en utilisant BlazeFace, à partir de l'image d'entrée (3DDFAV2)
        - Création du modèle de visage par MobileNet (3DDFAV2)
    - Trouver la transformation qui aligne le modèle de visage généré par 3DDFAV2 issu de l'image précédente sur celui de l'image actuelle
    - Appliquer la transformation sur le modèle d'épithèse donné par le médecin
    - Filtrer la position du modèle d'épithèse
    - Afficher uniquement le modèle d'épithèse (le modèle de visage est sauvegardé pour l'itération suivante)

3.4 Création des modèles de nez

3.4.1 Reconstruction 3D du modèle de visage

Afin d'utiliser le « Miroir Magique » il faut d'abord créer les modèles nécessaires. Pour cela, la première image du patient est utilisée avec 3DDFAV2 afin d'en extraire un modèle 3D de visage (*visage_patient*), illustré dans la Figure 3.3. Dans un premier temps, 3DDFAV2 utilise BlazeFace afin de détecter le visage et de créer une image à partir de la région d'intérêt (le visage), puis il ajuste les paramètres d'un 3D Morphable Model (3DMM) pour qu'il s'ajuste au mieux au visage présent dans l'image. Ce modèle 3D sera utilisé pour la suite de l'initialisation.

Figure 3.3 Image en entrée du 3DDFAV2, où la zone de visage détectée par BlazeFace est indiquée en bleue, et image avec le modèle 3D créé par 3DDFAV2

3.4.2 Préparation des modèles de nez

Afin de pouvoir utiliser le plus de modèles d'épithèses différents et automatiser le plus possible la méthode, il est nécessaire de traiter les modèles de nez donnés par le médecin sous forme de modèle 3D. Effectivement, le recalage (superposition de deux modèles de nez) (3.4.3) est plus efficace si les deux modèles de nez (issu du médecin (nez*_épithèse)* et de l'image du patient (*nez_mesh*)) ont déjà une orientation similaire. S'ils ne partagent pas la même orientation initiale, un des modèles risque de se retourner par rapport à l'autre (Figure 3.4).

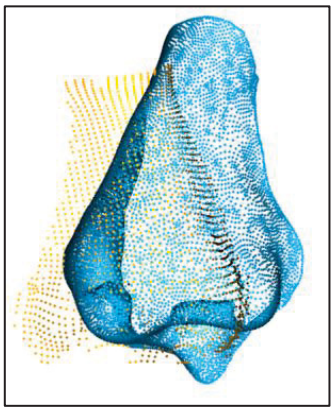

Figure 3.4 Nuages de points issus des modèles de nez, l'inscription en a retourné *nez_épithèse* (bleu) par rapport à *nez_mesh* (jaune)

Étant donné que le patient aura son visage à la verticale, le *nez_mesh* sera presque aligné sur l'axe y (Figure 3.5), l'alignement est donc effectué *nez_épithèse* sur l'axe y également.

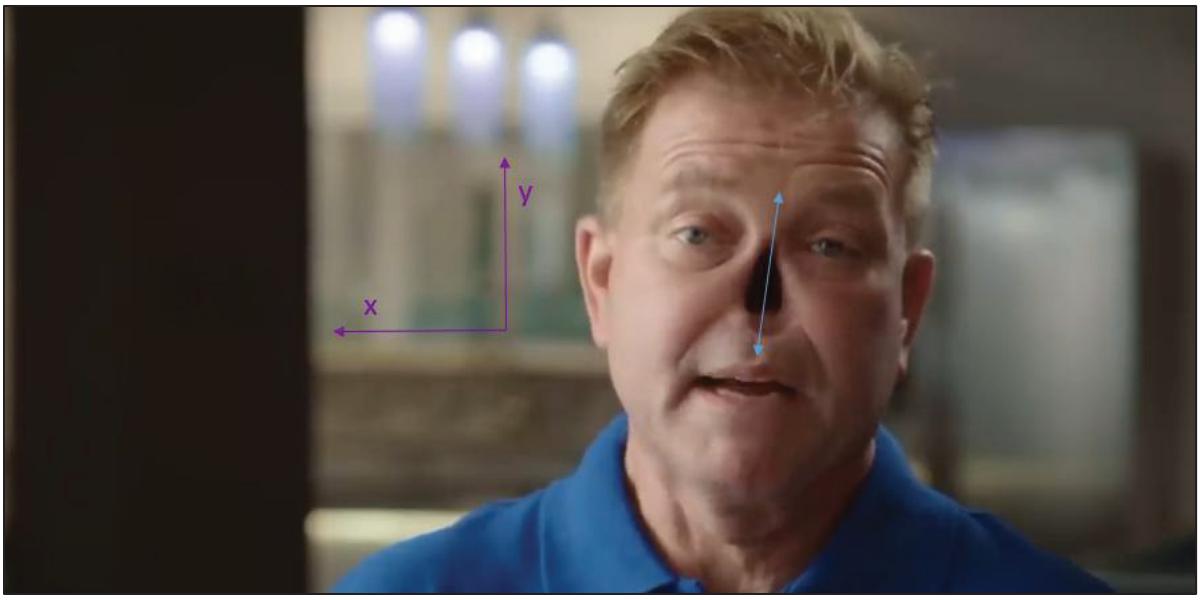

Figure 3.5 Patient à l'initialisation

Pour cela une Analyse des Composantes Principales (ACP) est utilisée afin de déterminer les axes principaux selon lesquels sont répartis les points qui composent les modèles de nez, c'est-à-dire la pose du modèle. Ici les nez possèdent 3 axes de répartition des points clairs comme illustré dans la Figure 3.6.

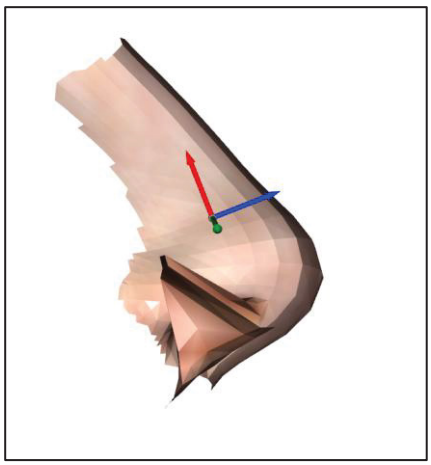

Figure 3.6 Visualisation des axes principaux d'un nez

Une translation est ensuite appliquée au *nez_épithèse* pour le placer à l'origine, pour ensuite y appliquer une rotation (qui correspond à l'inverse de la pose du modèle), afin de l'aligner à l'axe y comme illustré dans la Figure 3.7.

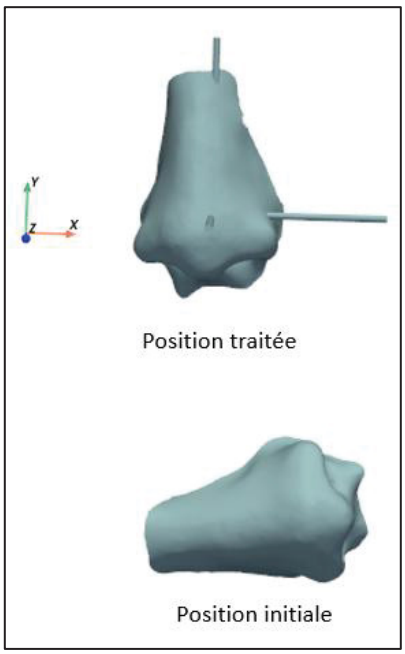

Figure 3.7 *nez_épithèse* avant et après repositionnement

Une fois le *nez_épithèse* aligné avec l'axe y, il est injecté dans la suite du programme, afin de plaquer la texture dessus.

3.4.3 Appliquer la texture sur le modèle de nez

Afin de pouvoir appliquer la texture sur *nez_épithèse*, il faut dans un premier temps générer la texture du nez. Pour cela la première image extraite par RetinaFace à partir de la première image est utilisée avec 3DDFAV3, qui génère un modèle 3D du visage du patient (*visage_texture*), avec de la texture (Figure 3.8).

Figure 3.8 Exemple de création *visage_texture* grâce à 3DDFAV3, à partir d'un visage sans nez

Une fois ce modèle 3D généré, il faut en extraire le nez (*nez_texture*). Contrairement à 3DDFAV2, la segmentation de 3DDFAV3 inclut le nez, qui peut donc être extrait facilement.

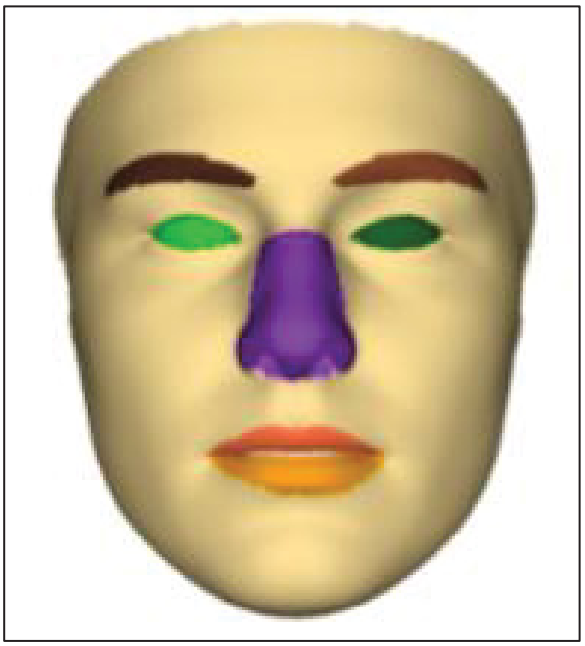

Figure 3.9 Représentation de la segmentation du modèle 3D de visage dans 3DDFAV3
Tirée de (Z. Wang et al., 2024)

Ce qui permet d'obtenir un modèle de nez avec de la texture. Afin de transférer cette texture à *nez_epithèse*, il est nécessaire de superposer les deux modèles l'un par rapport à l'autre. Pour cela, deux nuages de points sont extraits à partir de *nez_épithèse* et *nez_texture*. Une mise à l'échelle est ici nécessaire car les deux modèles ont des tailles très différentes, comme illustré dans Figure 3.10 et Figure 3.11, ce qui empêche le recalage des deux modèles, car leurs géométries sont trop différentes à cause de la différence de taille.

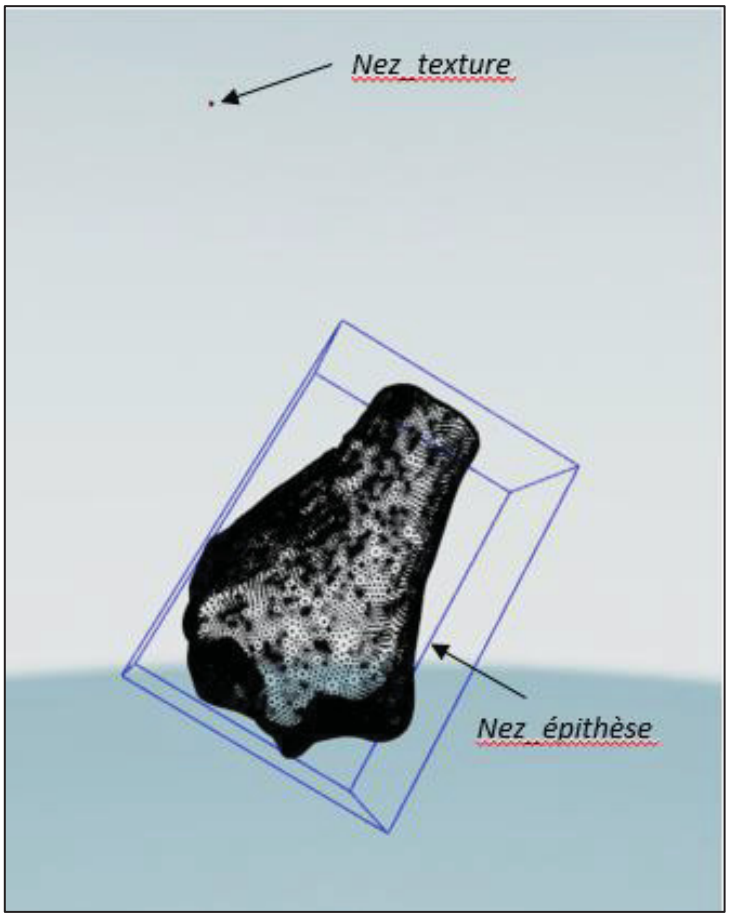

Figure 3.10 Visualisation de *nez_texture* et *nez_épithèse* ainsi que leur boite englobante associées

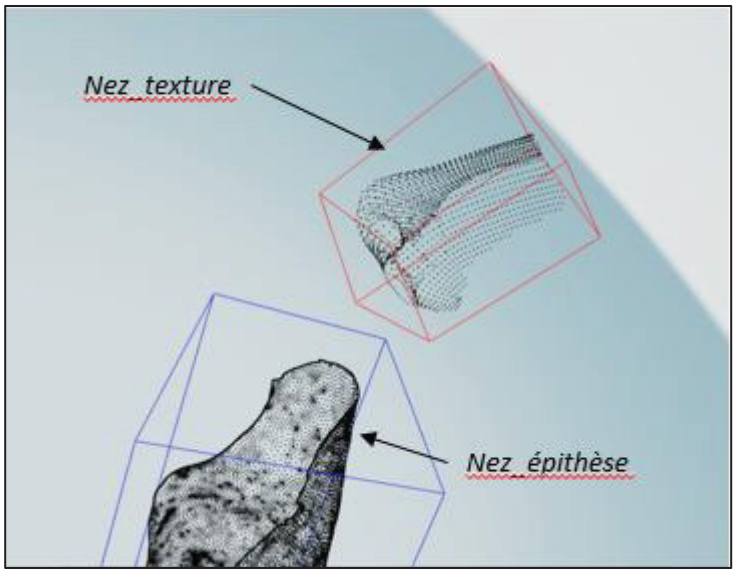

Figure 3.11 Visualisation de *nez_texture* et *nez_épithèse*
et leurs boites englobantes associées avec
un zoom sur *nez_texture*

Pour obtenir la même taille, la différence de taille entre les boites englobantes de chaque nuage de points est calculée, dans toutes les dimensions de l'espace dans le système de référence global.

$$BoiteEnglobante_{épithèse} \rightarrow (Lx_{épithèse}, Ly_{épithèse}, Lz_{épithèse}) \quad (3.1)$$

$$BoiteEnglobante_{texture} \rightarrow (Lx_{texture}, Ly_{texture}, Lz_{texture}) \quad (3.2)$$

$$Diff_x = \quad abs(Lx_{épithèse} - Lx_{texture}) \quad (3.3)$$

$$Diff_y = \quad abs(Ly_{épithèse} - Ly_{texture}) \quad (3.4)$$

$$Diff_z = \quad abs(Lz_{épithèse} - Lz_{texture}) \quad (3.5)$$

De la différence la plus grande entre les deux boîtes est extrait un facteur de grandissement, appliqué à tous les points de *nez_texture* via une matrice de transformation.

$$DimMax = \begin{cases} x \; si \; max(Diff_x, Diff_y, Diff_z) = Diff_x \\ y \; si \; max(Diff_x, Diff_y, Diff_z) = Diff_y \\ z \; si \; max(Diff_x, Diff_y, Diff_z) = Diff_z \end{cases} \quad (3.6)$$

$$Coef = \frac{L_{DimMax_{épithèse}}}{L_{DimMax_{texture}}} \tag{3.7}$$

$$Mat_{scale} = \begin{pmatrix} Coef & 0 & 0 & 0 \\ 0 & Coef & 0 & 0 \\ 0 & 0 & Coef & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.8}$$

Cette approche garantit un changement d'échelle uniforme tout en préservant le ratio d'aspect du modèle. La position et l'orientation de ces nuages de points est ensuite calculée grâce à une ACP, comme pour *nez_épithèse* dans la partie précédente.

Les matrices de poses (rotation) des modèles ainsi que les coordonnées de leur centre sont alors obtenues :

$$MatPose_{épithèse} = \begin{pmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{pmatrix} \tag{3.9}$$

$$MatPose_{texture} = \begin{pmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{pmatrix} \tag{3.10}$$

$$MatPose_{épithèse} = \begin{pmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{pmatrix} \tag{3.11}$$

$$MatRota = MatPose_{texture}{}^{T} * MatPose_{épithèse} \tag{3.12}$$

$$Centre_{texture} = [t_x \quad t_y \quad t_z] \tag{3.13}$$

$$Centre_{épithèse} = [e_x \quad e_y \quad e_z] \tag{3.14}$$

$$Centre_{texture} = [t_x \quad t_y \quad t_z] \tag{3.15}$$

Avec ces nouvelles informations, il est possible de déplacer *nez_épithèse* à l'origine (note : le modèle est normalement positionné à l'origine grâce aux opérations décrites dans la section précédente. Cependant, la mise à l'échelle peut légèrement déplacer son centre ; il est donc nécessaire de le recentrer à l'origine), appliquer les rotations appropriées pour aligner les deux modèles sur le même axe, puis une nouvelle translation pour les superposer comme illustré dans la Figure 3.12.

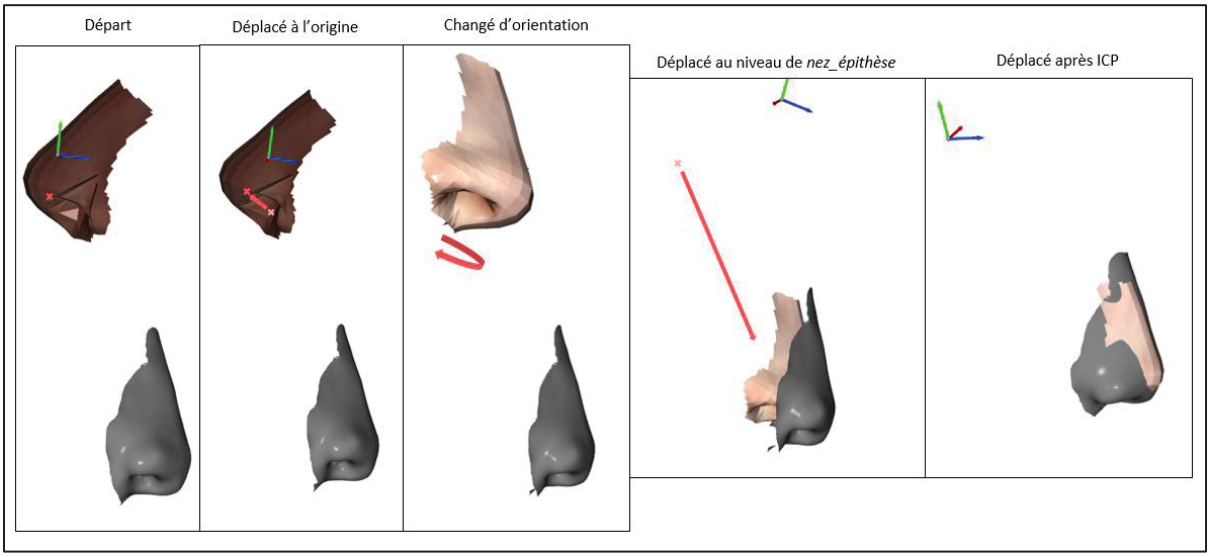

Figure 3.12 Visualisation du principe de recalage appliqué sur *nez_texture* vers *nez_épithèse*

Pour faire cela, des opérations de déplacement sont appliquées à tous les points de *nez_texture* :

$$M = [m_x \quad m_y \quad m_z] \tag{3.16}$$

$$M = M - Centre_{texture} \tag{3.17}$$

$$M = M * MatRota \tag{3.18}$$

$$M = M + Centre_{épithèse} \tag{3.19}$$

Avec M un point de *nez_texture*

Une fois les transformations issues de l'ACP effectuées, une ICP peut être utilisée pour minimiser la distance entre les deux modèles, et corriger les erreurs encore présentes. Pour cela l'ICP essaye des transformations de positions sur *nez_texture* jusqu'à trouver un minimum de distance avec ses plus proches voisins, qui satisfait un seuil de distance :

$$d_{min} = \min(\sum \|p - Mq\|^2) \; si \; d_{min} < d_{seuil} \tag{3.20}$$

Avec p et q deux points voisins de *nez_texture* et *nez_épithèse*, et d_{seuil} une valeur seuil définie au préalable.

Après avoir superposé *nez_épithèse* et *nez_texture*, la texture de *nez_texture* peut être transférée vers *nez_épithèse* : pour chaque point du nuage de point issu de *nez_épithèse,* on cherche son voisin le plus proche sur le nuage de point issu de *nez_texture,* en se basant sur la distance euclidienne entre deux points :

$$M = \begin{bmatrix} x_M & y_M & z_M \end{bmatrix} \tag{3.21}$$

$$P = \begin{bmatrix} x_P & y_P & z_P \end{bmatrix} \tag{3.22}$$

$$dist = \sqrt{(x_M - x_P)^2 + (y_M - y_P)^2 + (z_M - z_P)^2} \tag{3.23}$$

Ce processus est accéléré en utilisant un arbre k-dimension (Bentley, 1975).

La couleur du point du modèle créé par 3DDFAV3 est transférée vers le sommet correspondant de l'épithèse. Afin d'avoir un rendu plus réaliste, le nez est coloré avec les couleurs des faces au lieu des couleurs des sommets, la couleur des faces est la moyenne des couleurs des sommets qui la composent :

$$P_{i\in[1,3]} = \begin{bmatrix} R_i & G_i & B_i \end{bmatrix} \tag{3.24}$$

$$F = \frac{1}{3} \sum_{i\in[1,3]} \begin{bmatrix} R_i & G_i & B_i \end{bmatrix} \tag{3.25}$$

Avec P les 3 points qui composent la face triangulaire F

La coloration par face permet de réduire le nombre d'artefacts laissés par le transfert de couleur de *nez_texture* vers *nez_épithèse*, comme illustré dans Figure 3.13.

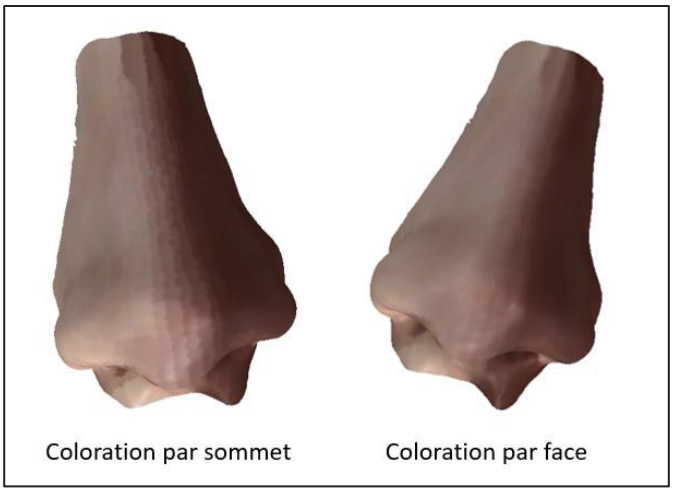

Figure 3.13 Comparaison de la coloration
par sommet et par face

Le rendu est ensuite effectué grâce au rendu de Cook-Torrance (Cook & Torrance, 1982).

3.4.4 Extraction du nez depuis le modèle de visage de 3DDFAV2

Une fois que *nez_épithèse* a de la texture, il faut réussir à l'aligner sur le nez du modèle issu de 3DDFAV2. En effet le modèle de visage issu de 3DDFAV2 est correctement placé vis-à-vis de l'image entière, tandis que 3DDFAV3 ne connait que la position du visage dans la partie extraite par RetinaFace. La position de *nez_épithèse* sur la première image est cruciale, car c'est à partir de cette position que les suivantes seront calculées.

Pour cela le nez du modèle 3D de visage généré lors de la première image (*nez_visage*) est extrait du modèle 3D de visage, afin de pouvoir le recaler avec *nez_épithèse*. Malheureusement la segmentation du modèle de visage utilisé par 3DDFAV2 ne permet pas d'extraire facilement le nez.

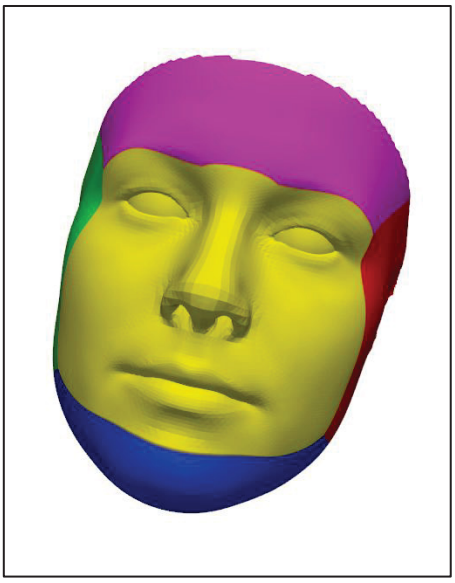

Figure 3.14 Modèle 3D généré par 3DDFAV2 coloré par segment

Pour cela la région du modèle de nez est extraite manuellement. L'extraction consiste à relever les différents indices de chacun des sommets du nez, puis créer un nouveau modèle 3D composé uniquement des sommets dont les indices ont été relevés.

3.4.5 Recalage de nez_visage et nez_épithèse

Grâce aux indices des points du nez dans le modèle 3D issu de 3DDFAV2 trouvés dans la partie précédente, *nez_visage* peut facilement être isolé du reste du modèle 3D. *Nez_épithèse* peut ensuite être recalé vers *nez_visage*, comme utilisé dans la section précédente, grâce à l'ACP et une ICP ce qui permet de placer précisément *nez_épithèse*. Une fois *nez_épithèse* en place, deux copies en sont créées, qui seront utilisées pour l'affichage, et pour le filtrage des positions.

3.4.6 Fin de l'initialisation

Une fois l'initialisation terminée, les différents éléments nécessaires à la suite du déroulement du programme sont prêts:

- Modèle d'épithèse avec texture, placé au bon endroit
- Modèle de visage généré par 3DDFAV2, issu de l'image d'initialisation, placé au bon endroit

3.5 Déroulement des tâches lors de l'utilisation du Miroir Magique

3.5.1 Préparation du suivi de visage

Une fois l'initialisation générale du programme effectuée, les images peuvent être traitées une par une. La fonction de suivi de 3DDFAV2 n'étant pas adaptée à notre cas, nous obligeons 3DDFAV2 à créer un modèle 3D de visage à chaque nouvelle image donnée (*masque_actuel*). Étant donné que les vidéos sur lesquelles le « Miroir Magique » sera utilisé seront continues, les modèles 3D successifs seront tous très similaires, sous réserve qu'il fonctionne en temps réel. En effet la fréquence de traitement sera assez élevée pour qu'il ne soit pas possible de faire des mouvements de visages assez important pour créer deux modèles 3D successifs qui ne se ressemblent pas.

Un nuage de points est extrait de *masque_actuel*, ainsi que du modèle 3D de visage de l'image précédente (*masque_précédent*).

Le suivi va être effectué à l'aide d'un recalage par correspondance, il faut donc créer deux listes d'indices qui indiqueront les points à faire correspondre entre eux. Les deux modèles 3D sont tous les deux créés par 3DDFAV2, et possèdent donc la même topologie, il est donc théoriquement possible d'utiliser l'entièreté des points en tant que points de correspondance. En pratique, utiliser la totalité des points ralentit énormément le programme (60ms contre 500ms), la liste de correspondance est donc réduite à un point sur dix. Le choix de sélectionner un point sur dix a été décidé de façon empirique : il s'agit d'une valeur qui permet d'accélérer grandement la vitesse de calcul, sans que l'impact sur la précision du recalage ne soit trop important. Étant donné que les points sont répartis par régions, les points dont les indices sont proches sont également proches dans l'espace, en choisissant un point sur dix une répartition uniforme par région est conservée.

3.5.2 Recalage des modèles 3D par correspondance pour chaque image

Masque_actuel est ensuite aligné avec le *masque_précédent* en utilisant un recalage par correspondance. L'inscription par correspondance consiste à sélectionner un sous ensemble de points dans chacun des nuages de points qui doit être positionné, ces sous-ensembles de points doivent correspondre de façon bijective (1 à 1) comme illustré dans la Figure 3.15.

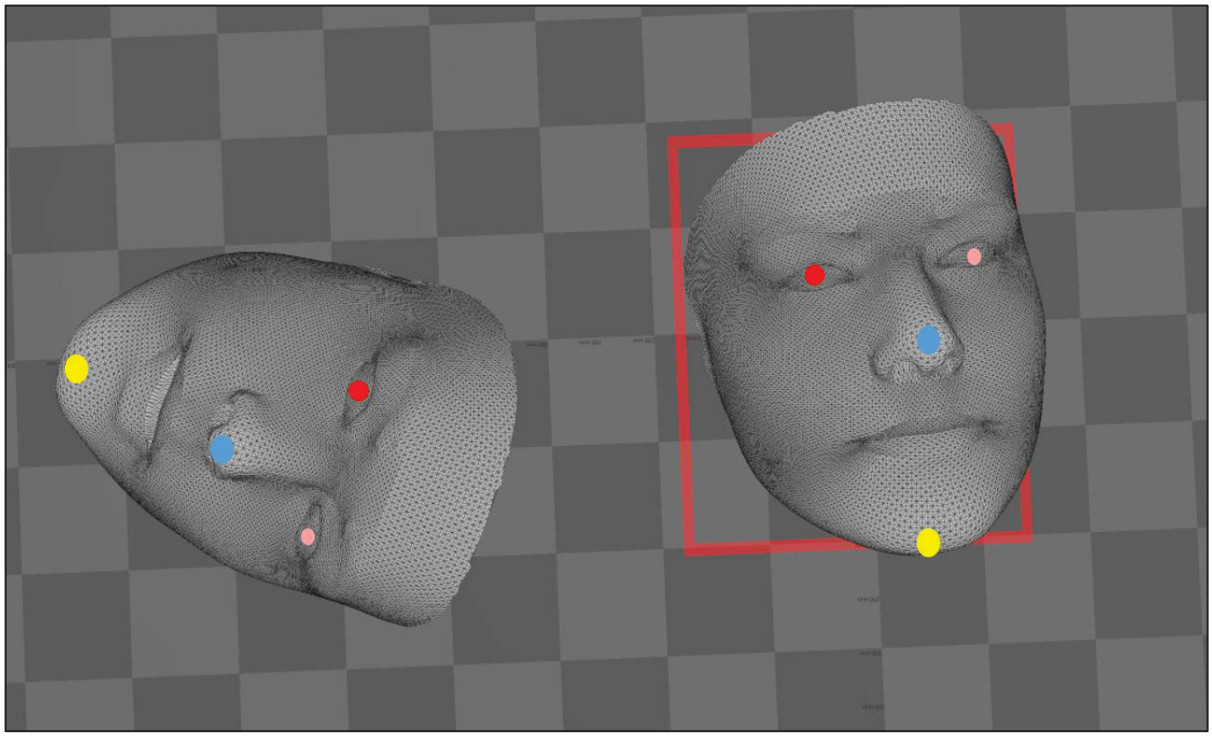

Figure 3.15 Exemple de sous ensemble de points qui correspondent de façon bijective

Ensuite l'algorithme essaie de trouver la transformation qui minimise la distance entre les points et leur correspondant. On cherche la matrice de transformation M qui minimise la distance d calculée par :

$$d = \sum_{(p,q)\in K} \|p - Mq\|^2 \tag{3.26}$$

Avec K le set de points de correspondance et M une matrice de transformation

Cette méthode ne fonctionne qu'avec des nuages de points avec des formes très proches, et dont les indices de chaque point sont connus, pour pouvoir établir une correspondance, ce qui

est le cas ici. En plus de la superposition des modèles 3D, l'inscription par correspondance permet également d'ajuster l'échelle du modèle 3D modifié pour pouvoir l'aligner sur celle du masque nouveau. Cette fonction est particulièrement utile lorsque l'utilisateur s'approche ou s'éloigne de la caméra, ce qui rend son visage plus gros ou plus petit sur la vidéo.

Le recalage par correspondance génère une matrice de transformation, qui correspond à la différence de rotation, taille et position des deux modèles 3D, soit le mouvement du visage entre l'image actuelle et précédente. Cette matrice de transformation est appliquée à *nez_épithèse* (qui se trouve au même endroit que le *masque_précédent* grâce à l'itération précédente). Le *masque_actuel* servira à faire le recalage par correspondance lors de l'itération suivante tandis que le *nez_épithèse* est utilisé pour afficher le nez sur la vidéo.

3.5.3 Filtrage des positions des modèles 3D pour chaque image

Le modèle 3D étant généré par 3DDFAV2 à chaque image, la position et l'orientation changent un petit peu à chaque image bien que le patient ne bouge pas, ou peu. Cet effet a pour conséquence un effet de vibration sur la vidéo finale, qui peut être perturbant, distraire le patient, ou le sortir de son immersion (Wilmott et al., 2022). Le filtre Minilag (Song et al., 2023) est utilisé pour éliminer cette vibration, sans induire de retard, ce qui était un problème importants dans les autres filtres de positions pour les applications de réalité virtuelle. Afin de réduire le temps de calcul du filtrage, seules la position et la rotation du centre sont filtrées. Grâce à la méthode d'ACP, ces caractéristiques sont extraites pour qu'elles puissent être filtrées. Minilag génère la rotation et la position du modèle filtré, il faut donc créer la matrice de rotation et de translation à appliquer au modèle d'épithèse. Pour cela la matrice de pose après filtrage est multipliée par la transposée de la matrice de pose avant filtrage.

3.5.4 Rendu du miroir magique

Afin d'afficher le nez avec texture, il a été nécessaire de créer une méthode d'affichage personnalisée, les outils précédents (3DDFAV2, 3DDFAV3) n'ayant pas les caractéristiques désirées, c'est-à-dire la capacité d'afficher n'importe quel modèle 3D, à l'endroit désiré, avec la texture associée au modèle. Pour créer la méthode d'affichage, on a utilisé le bibliothèque

python pyrender, qui est une bibliothèque spécialisée dans l'affichage de modèles et de scènes 3D.

L'initialisation de la méthode d'affichage s'effectue en plusieurs étapes, tout d'abord il faut créer la scène, qui contiendra tous les objets nécessaires à l'affichage. Le premier de ces objets est « l'écran », qui affiche l'image du patient en fond. Il s'agit en réalité d'un modèle 3D de parallélépipède rectangle, sur lequel a été appliqué une texture, la texture étant l'image du patient. Une caméra est ensuite ajoutée à la scène, placée de façon à ce que les bords de la vue de la caméra soient alignés avec les bords de « l'écran ». La position de la caméra est réglée de façon empirique, cependant, la taille de l'écran est fixe, donc la position n'a pas besoin d'être modifiée si la résolution de la caméra change. Une caméra orthographique (sans perspective) est utilisée afin de ne pas déformer le modèle de nez lorsqu'il sera affiché. Le modèle de nez quant à lui est ajouté entre l'écran et la caméra, selon les positions données par 3DDFAV2. Il est intéressant de noter que 3DDFAV2 inverse les coordonnées en y, nous avons donc dû inverser les coordonnées de l'écran, ce qui donne un rendu inverse par rapport à l'image de départ. Une visualisation de toute la scène vue de côté est présentée dans la Figure 3.16

Figure 3.16 Vue de côté de la scène à afficher,
le modèle d'épithèse est entouré en noir

Une fois la scène mise en place, le point de vue est créé depuis la caméra, puis affiché grâce à OpenCV.
A chaque nouvelle image, la scène de l'itération précédente est récupérée, puis « l'écran » et la position du nez sont actualisés avant d'afficher à nouveau la scène.

3.5.5 Méthodologie de validation

La validation de la méthode n'a pas pu être faite sur des patients, et il n'existe pas à notre connaissance de base de données de vidéo de patients ayant perdu leur nez dont l'orientation et la position sont connues. S'il est possible d'effectuer une expérience avec des patients dans le futur, il serait intéressant d'utiliser le « Miroir Magique » dans des positions extrêmes, et avec des mouvements brusques, puis de faire remplir un questionnaire au patient permettant de relever leur niveau de satisfaction sur différents critères comme l'immersion ou le réalisme qui sont des critères difficiles à évaluer objectivement. La validation a été réalisée avec des vidéos de personnes ayant perdu leur nez, qui étaient présentes sur Youtube et libres de droit (*After losing her nose to cancer, Becky shares her story to help others*, 2012 ; *J'ai perdu mon nez à cause du cancer | SHAKE MY BEAUTY*, 2021 ; *Man Who Lost His Nose To Cancer Gets Incredible Prosthetic Nose | Body Parts*, 2022 ; *Simplest Nasal Prosthesis*, 2016). Ces vidéos sont des documentaires, que nous avons modifié pour en extraire les parties où le patient répondait à des questions, afin d'obtenir un plan presque continu, pour simuler au mieux l'utilisation du « Miroir Magique ». Nous obtenons ainsi 6 vidéos de 4 patients différents, qui font face à la caméra comme illustré dans la Figure 3.17. Ces vidéos sont assez proches de l'utilisation prévue du « Miroir Magique », elles sont donc pertinentes à utiliser pour la validation.
Afin d'avoir un point de référence pour obtenir les mesures de précision, nous avons annoté chaque image des vidéos. Pour cela nous avons choisi 3 points caractéristiques et facilement

identifiables pour chaque nez, de façon à ce que leur barycentre soit proche du centre du nez, comme illustré dans la Figure 3.18. Le barycentre peut être calculé grâce à :

$$\begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} = \begin{bmatrix} \frac{x_1 + x_2 + x_3}{3} \\ \frac{y_1 + y_2 + y_3}{3} \\ \frac{z_1 + z_2 + z_3}{3} \end{bmatrix} \tag{3.27}$$

Nous traitons ici une vidéo 2D, nous n'utiliserons donc que les coordonnées en x et en y.

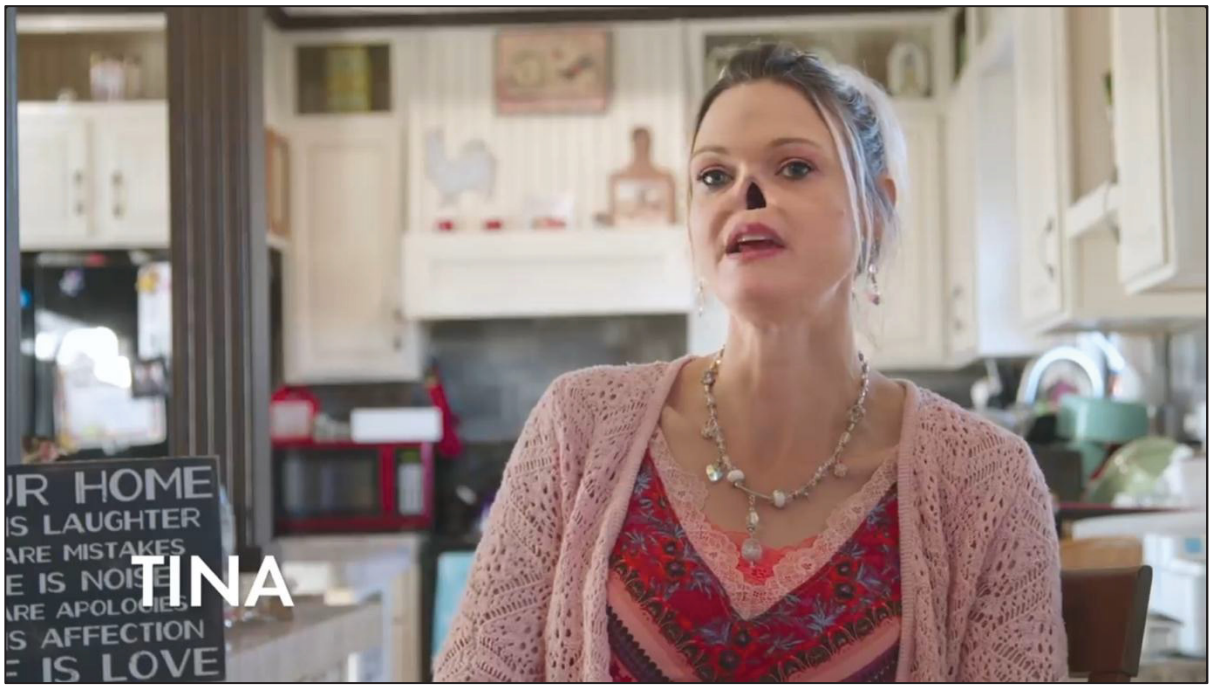

Figure 3.17 Exemple d'image présentes dans les vidéos de validation

Figure 3.18 Exemple de points de référence ainsi que leur barycentre dans une vidéo de validation

Ce point du nez est comparé à un point extrait du modèle d'épithèse qui est le barycentre de 3 points, choisis pour correspondre aux points caractéristiques choisis dans la vidéo, comme illustré dans Figure 3.19 .

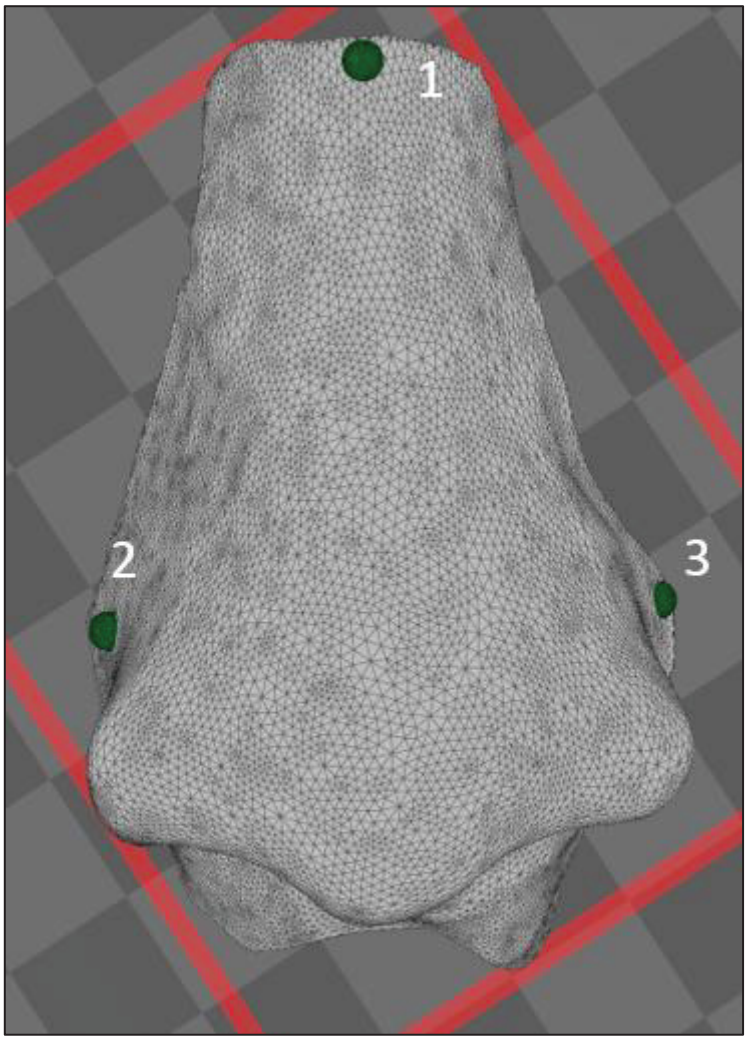

Figure 3.19 Visualisation des points choisis sur
le modèle d'épithèse

Afin de mesurer la Dérive (Engel et al., 2014), nous avons extrait une image de chaque vidéo de validation, et avons créé une vidéo de 4 secondes de cette image fixe.

Pour les méthodes de validations du filtre Minilag, qui nécessitent une vérité terrain mais ne dépend pas des images de patient, nous avons utilisé la base de donnée de l'Universidad Pública de Navarra (UPNA) (Ariz, Bengoechea, Villanueva, & Cabeza, 2016). La base de données est composée de 120 vidéos, de 10 personnes différentes. Les 12 vidéos par personne sont divisées en 6 vidéos guidées selon les différents degrés de libertés, et 6 vidéos de

mouvement libres. Les données de vérités terrains sont données en millimètres pour les translations et en degré pour les rotations, qui sont données sous la forme d'angles d'Euler.

La **précision** peut être évaluée grâce à la comparaison des barycentres du modèle d'épithèse et de la labélisation. On obtient une erreur en x, une erreur en y et une erreur totale qui est calculée comme la distance euclidienne entre les deux points :

$$Tot = \sqrt{(x_1 - x_2)^2 + (y, -y_2)^2} \tag{3.28}$$

Avec x_i et y_i les coordonnées des barycentres.

Le nombre **d'image par seconde** peut être calculé à la fin du traitement de la vidéo, en relevant le nombre d'images complètes ayant été traitées, et le temps nécessaire pour les traiter. Il est ensuite calculé par :

$$N = \frac{Nombre\ d'image}{Temps\ nécessaire} \tag{3.29}$$

La latence « motion-to-photon » peut être évaluée en calculant la moyenne du temps pris par le « Miroir Magique » entre le moment où il commence à traiter l'image et le moment où il l'affiche.

$$T = T_{Affichage} - T_{Début} \tag{3.30}$$

Les vibrations peuvent être calculées facilement grâce aux formules données par les créateurs du filtre Minilag (Song et al., 2023) :

$$J_R = \frac{1}{3n}\sum_{i=1}^{n}\left(\left(\omega_1^{i\,\prime\prime}\right)^2 + \left(\omega_2^{i\,\prime\prime}\right)^2 + \left(\omega_3^{i\,\prime\prime}\right)^2\right) \tag{3.31}$$

$$J_t = \frac{1}{3n}\sum_{i=1}^{n}\left(\left(t_1^{i\prime\prime}\right)^2 + \left(t_2^{i\prime\prime}\right)^2 + \left(t_3^{i\prime\prime}\right)^2\right) \tag{3.32}$$

Avec J_R et J_t les vibrations en rotation et translation.

Dans un premier temps la robustesse peut être mesurée grâce au calcul de la Dérive (Engel et al., 2014). Elle peut être calculé sur une vidéo de visage fixe grâce à :

$$Drift = \left\|p_t - p_{t_0}\right\| \tag{3.33}$$

Dans un second temps, elle peut être calculée grâce à une boite englobante du nez. Dans les images où un détecteur de points de repère arrive à prédire la position des points de repère du visage (autre que le nez), ils peuvent être utilisés afin de créer une boite englobante de nez. Dans un premier temps cette boite englobante est 3D, elle est obtenue grâce aux points situés sur les yeux et la bouche, comme illustré dans Figure 3.20. La profondeur de la boîte est le double de sa longueur.

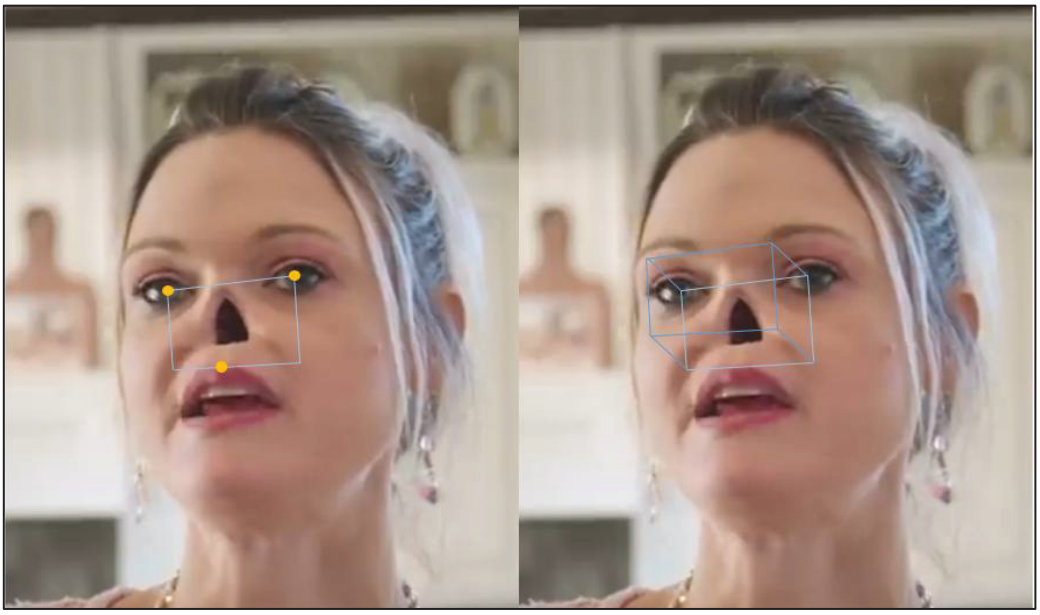

Figure 3.20 Création de la boîte englobante du nez

Une boite 2D est ensuite créée à partir des coordonnées extrêmes de la boîte précédente comme illustré dans la Figure 3.21.

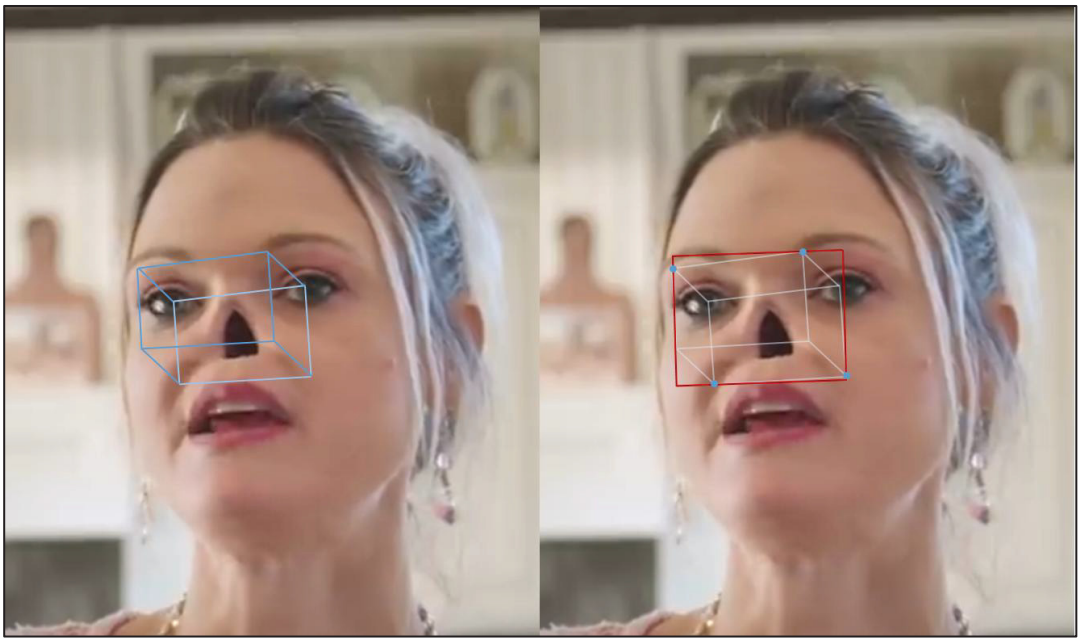

Figure 3.21 Création de la boîte englobante 2D à partir de la boîte englobante 3D

La projection orthogonale du modèle 3D d'épithèse sur le plan de l'image et la boite englobante 2D du nez sont ensuite comparées. Si au moins 95 % des points 2D du modèle d'épithèse sont compris dans la boîte englobante, on considère l'image comme « correctement alignée ». Il n'est pas possible de créer une boite englobante 2D directement depuis les points des yeux et de la bouche car lorsque l'utilisateur est de côté, la boîte serait réduite à une ligne.

Étant donné qu'il s'agit d'une application destinée à être utilisé par des patients, nous utiliserons également une validation qualitative qui consistera à observer si les vidéos produites par le « Miroir Magique » semblent adaptées à son utilisation prévue.

3.5.6 Intégration dans le projet AUDACE

Le « Miroir Magique » peut facilement s'intégrer dans le reste du projet AUDACE, en l'ajoutant à une application permettant de créer des moules d'épithèse en impression 3D comme présenté dans la Figure 3.22, Figure 3.23, et Figure 3.24.

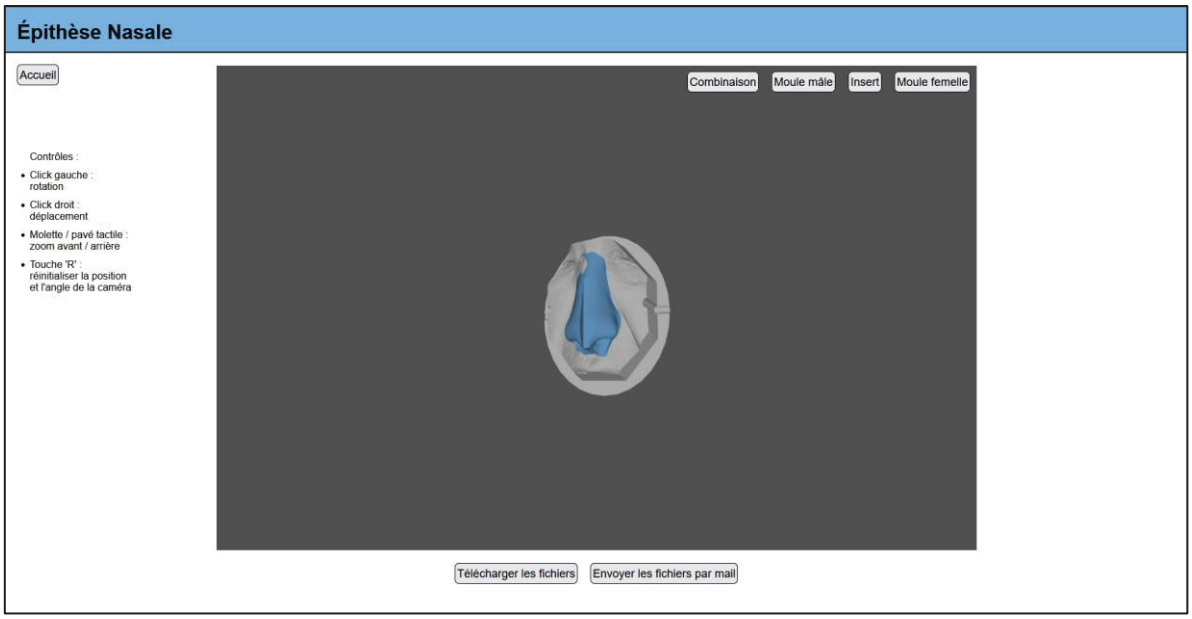

Figure 3.22 Moule mâle d'une épithèse avec son insert de nez

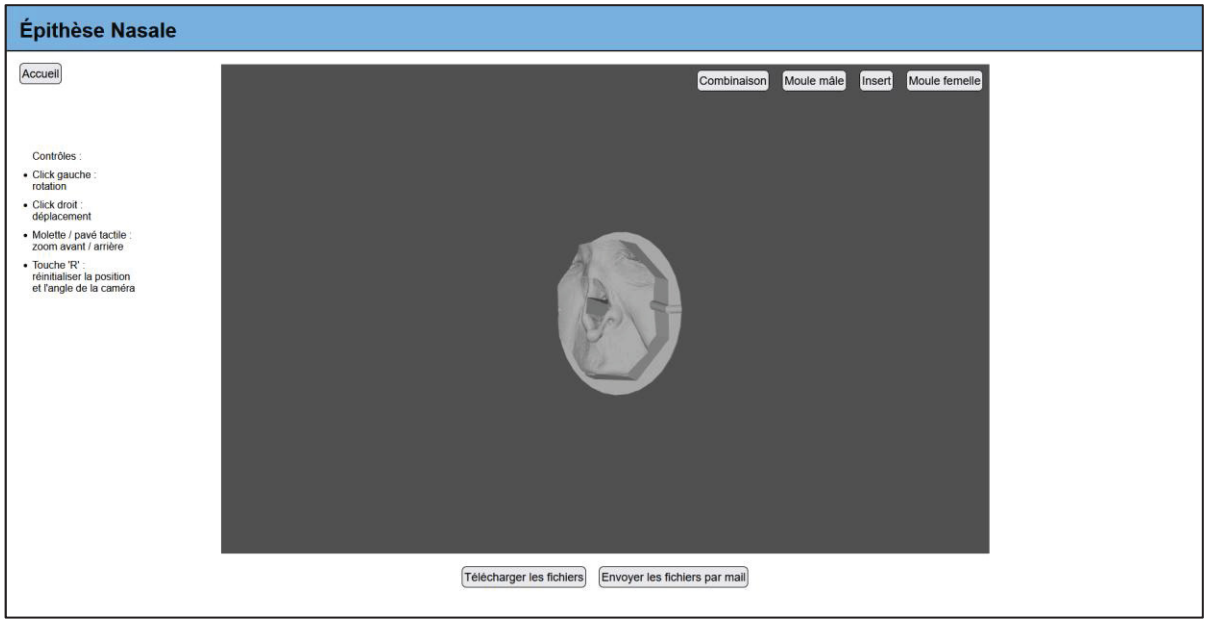

Figure 3.23 Moule mâle d'une épithèse

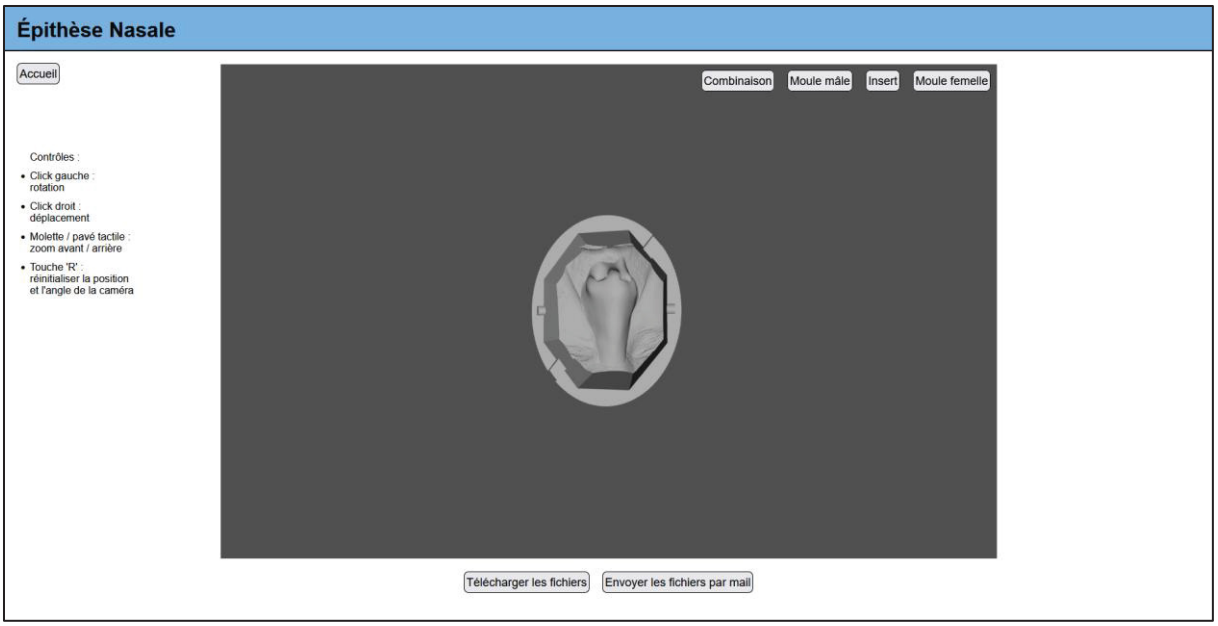

Figure 3.24 Moule femelle d'une épithèse

L'application permet par ailleurs de reconstruire un modèle 3D du nez à partir d'une photographie préopératoire du patient (Le Chartier, 2023). Cette reconstruction facilite le choix de l'épithèse en fournissant une base de référence personnalisée. Un exemple de nez généré par l'application et sa référence sont donnés dans la Figure 3.25.

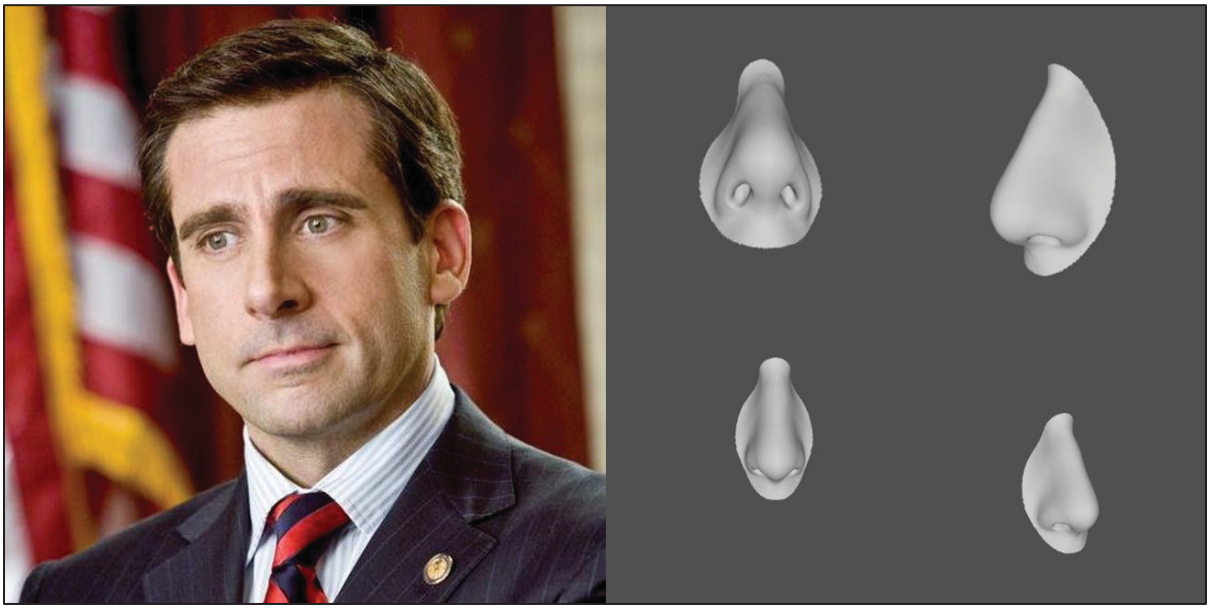

Figure 3.25 Exemple de nez généré par l'application ainsi que sa référence

En intégrant le « Miroir Magique » dans cette application, il est facilement intégré dans le reste du projet AUDACE, en complétant cette application, qui permet de créer une épithèse à partir d'une interface simple, illustrée dans la Figure 3.26 et la Figure 3.27.

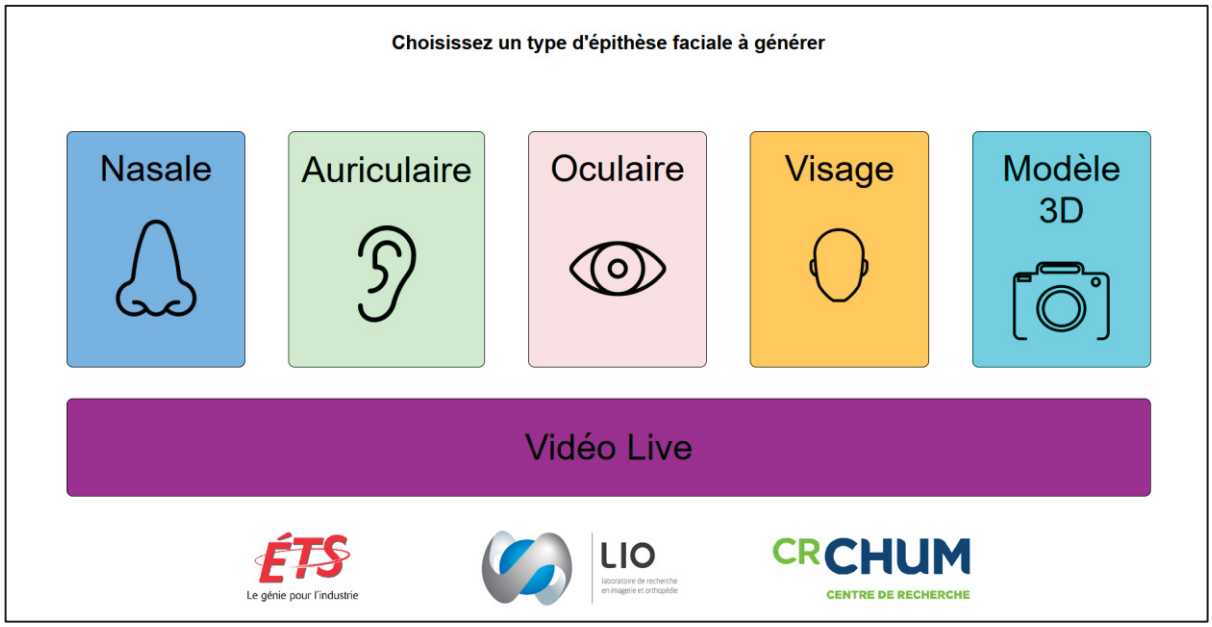

Figure 3.26 Page d'accueil de l'application

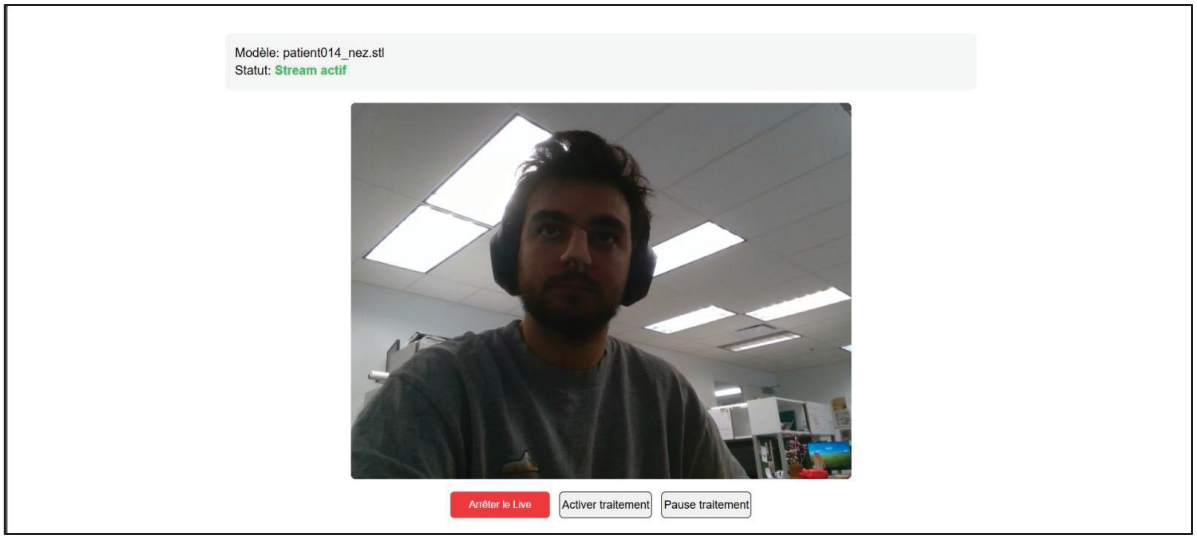

Figure 3.27 Aperçu de la fenêtre d'utilisation du Miroir Magique, dans l'application complète

CHAPITRE 4

VALIDATION EXPÉRIMENTALE

4.1 Résultats de l'expérience

La **précision** de positionnement du modèle 3D *nez_épithèse* est mesurée en calculant la différence de position entre le barycentre issu de la labélisation et celui issu du modèle d'épithèse. On obtient une différence moyenne de 18,04 pixels (px) sur les vidéos de validations, pour un maximum de 87,09 pixels de différence. Les résultats sont détaillés dans le Tableau 4.1. Les vidéos ont une taille variant de 640x360 pixels à 1920x1080 pixels, avec des visages compris entre 215x258 pixels et 405x467 pixels.

Tableau 4.1 Récapitulatif des résultats de la précision en pixels

	Moyenne (x)	Moyenne (y)	Moyenne distance euclidienne	Maximum (x)	Maximum (y)	Maximum distance euclidienne
Précision (px)	8,74	15,78	18,04	79,52	37,48	87,9

Il est également possible de mesurer la précision du modèle en soustrayant la différence de position des barycentres lors de la première image à toutes les coordonnées de position suivantes (ce qui revient à aligner les deux points). Ces résultats sont présentés dans le Tableau 4.2.

Tableau 4.2 Résultats de la précision, en alignant les points dans la première image

	Moyenne (x)	Moyenne (y)	Moyenne distance euclidienne	Maximum (x)	Maximum (y)	Maximum distance euclidienne
Précision (px)	6,28	7,39	9,70	71,76	43,12	83,72

En normalisant les résultats par la taille verticale du visage on obtient le Tableau 4.3.

Tableau 4.3 Résultats de la précision normalisés, en alignant les points dans la première image

	Moyenne (x)	Moyenne (y)	Moyenne De x et y	Maximum (x)	Maximum (y)	Maximum De x et y
Précision normalisée (%)	2,07	2,92	2,49	19,5	15,02	17,28

La moyenne des **images par secondes** qui sont capable d'être traitées par le programme est de 12 images par seconde.

Plus précisément, le temps de calcul pour chaque partie du programme, lors de l'utilisation du « Miroir Magique », hors initialisation, est présenté dans le Tableau 4.4.

Tableau 4.4 Temps de calcul pour chaque partie du programme

	Détection du visage	Création du modèle de visage	Déplacement du modèle d'épithèse	Affichage final
Temps de calcul (ms)	16	11	17	34

La moyenne de **latence** relevée est quant à elle de 23 ms.

La vibration est calculée grâce à la base de données de l'UPNA, qui donne pour les **rotations** dans l'algèbre de Lie les résultats présentés dans le Tableau 4.5. Ces résultats sont obtenus grâce à la formule du calcul de la vibration(1.21), donné par les créateurs du filtre Minilag (Song et al., 2023).

Tableau 4.5 Résultats du calcul de la vibration en rotation

	Vérité terrain	Prédictions brutes	Prédictions filtrées
Vibration (1.21)	$2{,}31 \cdot 10^{-6}$	$2{,}90 \cdot 10^{-4}$	$9{,}34 \cdot 10^{-5}$

La différence de degré entre les orientations prédites et la vérité terrain est également relevée. Ces résultats sont détaillés dans le Tableau 4.6.

Tableau 4.6 Résultats des différences d'orientation en degré

	Écart moyen	Écart maximum
X	12,38°	33,49°
Y	1,99°	6,78°
Z	9,09°	24,18°
Moyen	7,82°	21,48°

Les mesures de la vibration en rotation sont calculées grâce à la formule utilisée pour les rotations (1.22), en l'adaptant pour les translations. Les mesures sont présentées dans le Tableau 4.7.

Tableau 4.7 Résultats de la vibration en translation en millimètres

	Vérité terrain	Prédictions brutes	Prédictions filtrées
Moyenne (mm)	0,12841012	3,13777488	0,48886126

La **Dérive** est calculée sur une image fixe, on relève l'écart entre la position initiale et la position finale pour relever la cumulation des écarts, et on relève l'écart maximum et moyen depuis la position initiale. Les résultats sont présentés dans le Tableau 4.8.

Tableau 4.8 Récapitulatif des résultats de la dérive

	Dérive	Écart moyen	Écart Maximum
Dérive (pixels)	0.43	0.26	1.65

La **robustesse** du « Miroir Magique » est relevée pour chaque image à l'aide d'une boite englobante issue des points de détections des yeux et de la bouche, comme décrit dans le chapitre 1.5. Les résultats de robustesse pour chaque itération d'une vidéo de la base de données de l'UPNA est représenté dans la Figure 4.1.

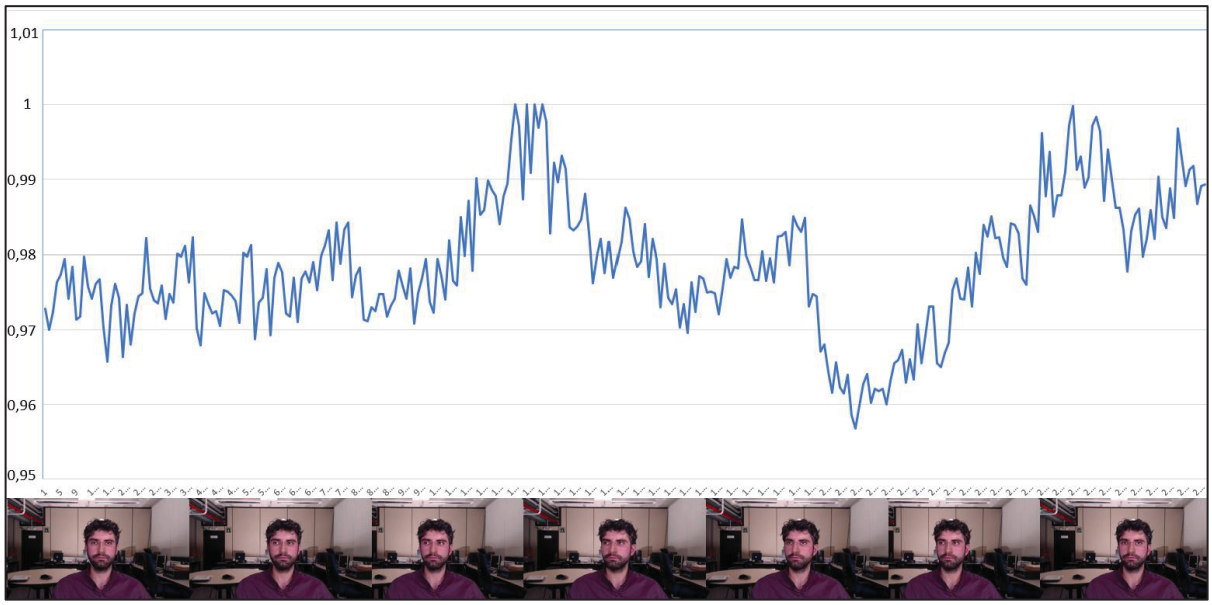

Figure 4.1 Résultats des scores de robustesse à chaque itération d'une vidéo de test, avec des exemples d'images traitées par le "Miroir Magique"

Le reste des résultats de robustesse sur les vidéos de la base de données UPNA sont en Annexe I. Les résultats issus des vidéos de validation de personnes sans nez sont présentés dans la Figure 4.2.

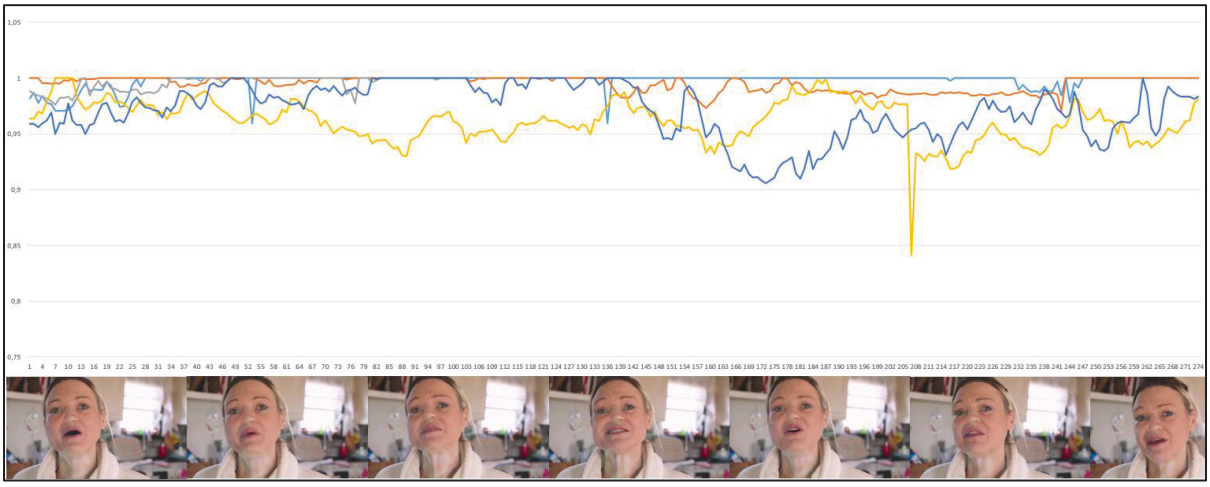

Figure 4.2 Courbes de robustesse pour les visages sans nez, coupée sur la longueur de la vidéo la plus courte

La Figure 4.3 est obtenue en enlevant les courbes produisant des valeurs extrêmes.

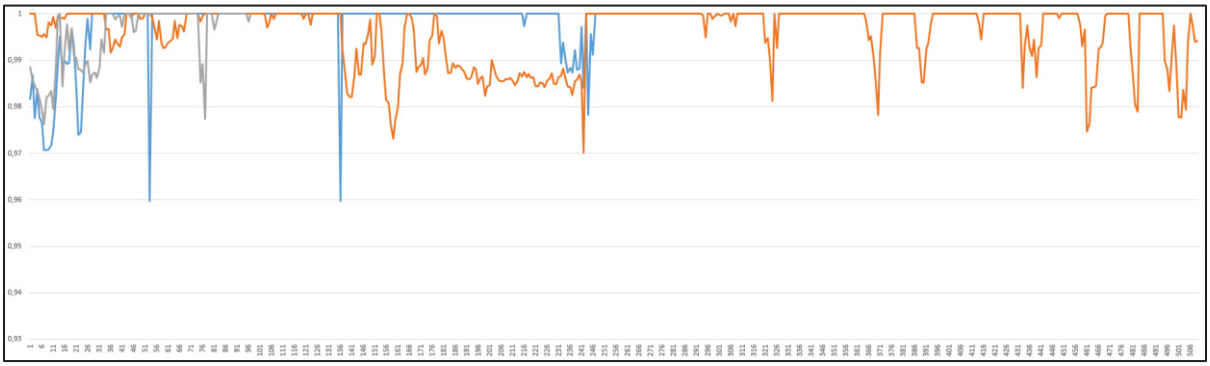

Figure 4.3 Courbes de robustesse sans les vidéos avec des pics extrêmes

On peut tracer les deux courbes contenant des extrêmes sur un seul graphique, et rassembler les pics des courbes de robustesse extrêmes dans deux catégories en fonction de leur durée, comme présenté dans la Figure 4.4.

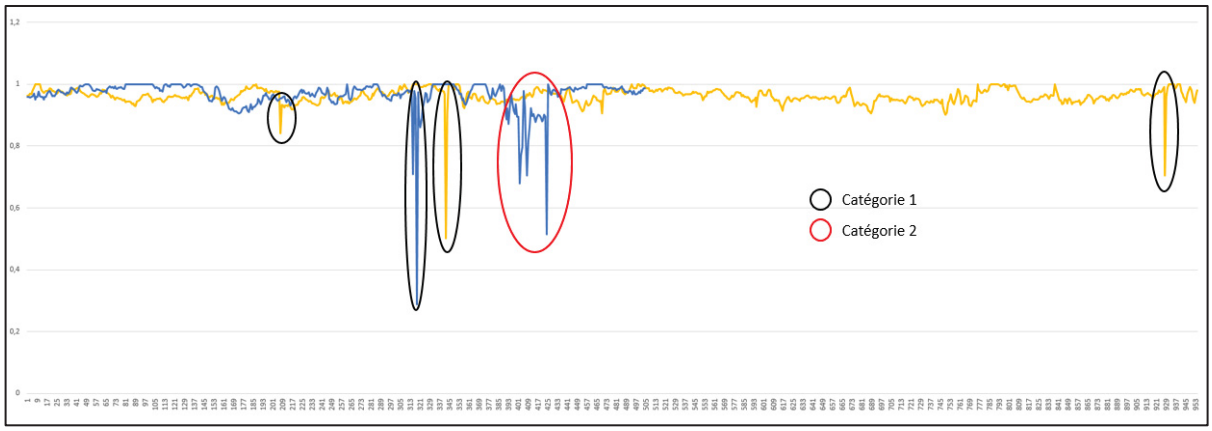

Figure 4.4 Illustration des catégories des écarts dans les courbes de robustesse

Des images où le nez n'est pas placé correctement à cause d'une occlusion sont relevées, comme illustré dans la Figure 4.5, Figure 4.6 et Figure 4.7

Figure 4.5 Photo extraite d'une vidéo de validation, la zone du nez est masquée

Figure 4.6 Exemple d'une occlusion qui empêche le modèle d'épithèse d'être placé correctement

Figure 4.7 Occlusion de la moitié du visage qui empêche le bon placement du nez

Des itérations ou le nez est bien placées sont également relevées, comme illustré dans la Figure 4.8 et la Figure 4.9.

Figure 4.8 Exemples d'itération où le programme fonctionne correctement

Figure 4.9 Exemples d'itération où le programme fonctionne correctement

CHAPITRE 5

DISCUSSION

5.1 Interprétation des résultats

Les résultats de la **précision** doivent être nuancés pour plusieurs raisons. Dans un premier temps, la labélisation des données a été effectuée à la main, ce qui induit inévitablement une source d'erreur (Dong et al., 2018). De plus il n'y a pas de garantie que les points de l'épithèse et de la vidéo correspondent parfaitement. En effet si l'on soustrait la différence de position lors de la première image à toutes les coordonnées de position suivantes (ce qui revient à aligner les deux points), de biens meilleurs résultats sont obtenus, présentés dans Tableau 4.2.

Il est à noter que les écarts maximums, présentés dans le Tableau 4.2, restent élevés par rapport aux moyennes, cela est dû à deux vidéos parmi les vidéos de validation qui ont une résolution élevée (1920x1080), et dans lesquelles sont présentent des occlusions, en l'occurrence les mains de la personne filmée qui passent devant son visage, qui décalent le nez par rapport à sa référence comme illustré dans la Figure 4.5 ce qui introduit de grandes erreurs. Le Tableau 4.3 fait d'ailleurs ressortir un écart proportionnellement bien plus petit entre les valeurs moyennes normalisées et les valeurs maximums normalisées.

Finalement, après une analyse qualitative des vidéos produites par le « Miroir Magique », on observe que le modèle d'épithèse est bien aligné avec la zone du visage qui devrait contenir le nez, ce qui confirme que le « Miroir Magique » est assez précis pour être utilisé.

Les **images par secondes** donnent un résultat un peu trop bas pour être utilisé en temps réel, de 12 images par seconde. Lors de l'utilisation en temps réel du « Miroir Magique », le rendu semble un peu saccadé, cela est dû aux différentes opérations qui doivent s'effectuer à chaque image, qui ralentissent considérablement le processus. Toutefois, les essais sur des vidéos ont tous été effectués sur un processeur. L'utilisation d'un processeur graphique permettrait d'accélérer les calculs, particulièrement pour l'affichage final, et par conséquent rendre

l'application jusqu'à 100 fois plus rapide (Hasnaine, Fouda, & Chiu, 2025), ce qui serait suffisant pour considérer qu'elle fonctionne en temps réel.

Le calcul de la **latence** indique que nous respectons bien le critère des 30 ms, ce qui ne devrait donc pas créer d'inconfort chez le patient.

Les résultats de la **vibration en rotation,** présentés dans le Tableau 4.5, et tels que donnés par la formule d'évaluation du filtre Minilag nous indique que les vibrations sont bien réduites. En effet la vibration filtrée est inférieure à la vibration brute, ce qui le rapproche bien de la vérité terrain, qui la vibration minimum possible. Étant donné que ces valeurs se trouvent dans l'algèbre de Lie, il est très difficile d'en tirer une interprétation concrète. Afin d'y remédier nous avons également extrait les écarts moyens et maximum de rotation en degré entre la vérité terrain et la prédiction du « Miroir Magique », dans le Tableau 4.6. Comme pour la précision, il existe une grande différence entre la moyenne et le maximum. Encore une fois cet écart est dû aux occlusions, qui empêchent le modèle de nez d'être placé correctement. Ce phénomène est illustré dans la Figure 4.6, qui montre bien que l'orientation du nez n'est pas alignée avec celle de la tête.

Les **vibrations en translation,** présentés dans le Tableau 4.7, représentent quant à eux la fluidité du placement du modèle d'épithèse. Il est difficile d'interpréter concrètement ce que signifient les résultats relevés, mais il est intéressant de noter qu'après le filtrage des positions, les résultats se trouvent dans le même ordre de grandeurs que les données vérité terrain, ce qui nous indique que le déplacement est bien fluide. De plus, cela correspond également aux observations faites par l'analyse qualitative des vidéos issues du « Miroir Magique » : les déplacements sont fluides, malgré des mouvements brusques des patients et le retard introduit par le filtre est assez faible pour ne pas être remarqué lors de l'utilisation du « Miroir Magique ».

Lors de la mesure de la **Dérive**, présentée dans le Tableau 4.8, un écart inférieur à un pixel est relevé, ce qui démontre bien que le modèle ne dérive pas systématiquement au cours du temps,

et qu'il est capable de rester pratiquement immobile si le patient ne bouge pas. De plus cela correspond aux observations faites après l'analyse qualitative : le modèle d'épithèse bouge très peu dans la vidéo produite par le « Miroir Magique » où le patient est immobile.

Les courbes de **robustesse** (Figure 4.2) sont plus simples à interpréter mais ont besoin d'être nuancées. Sur les visages avec nez, seuls 12 points sur 3600 sont en dessous de 0.95, avec un minimum de 0.93. Le critère de robustesse (plus de 95% du nez compris dans la boîte englobante) semble donc bien respecté. Dans les vidéos de visage avec nez en revanche le critère de robustesse n'est pas du tout respecté sur certaines images, le minimum étant 29% du nez dans la boîte englobante. Ces points extrêmes proviennent de deux vidéos. Lorsque leurs résultats sont supprimés du graphique, le « Miroir Magique » remplit le critère de robustesse sur les autres vidéos comme illustré dans la Figure 4.3.

Ces écarts extrêmes peuvent être divisés en deux catégories en fonction de leur durée comme illustré dans la Figure 4.4.

La cause de ces écarts peut très vite être identifiée en regardant les images dans lesquelles ils sont relevés. La catégorie 1 est due aux occlusions du visage, qui empêchent le modèle de nez d'être placé au bon endroit, comme illustré sur la Figure 4.7, ce qui ne respecte pas le critère de robustesse.

La seconde catégorie est due aux changements de positions entre deux images consécutives à cause du montage vidéo. En effet nous avons modifié les vidéos pour obtenir une vidéo de visage le plus continu possible, mais il reste encore des coupures qui changent la position du visage d'une image à l'autre, comme illustré dans la Figure 5.1.

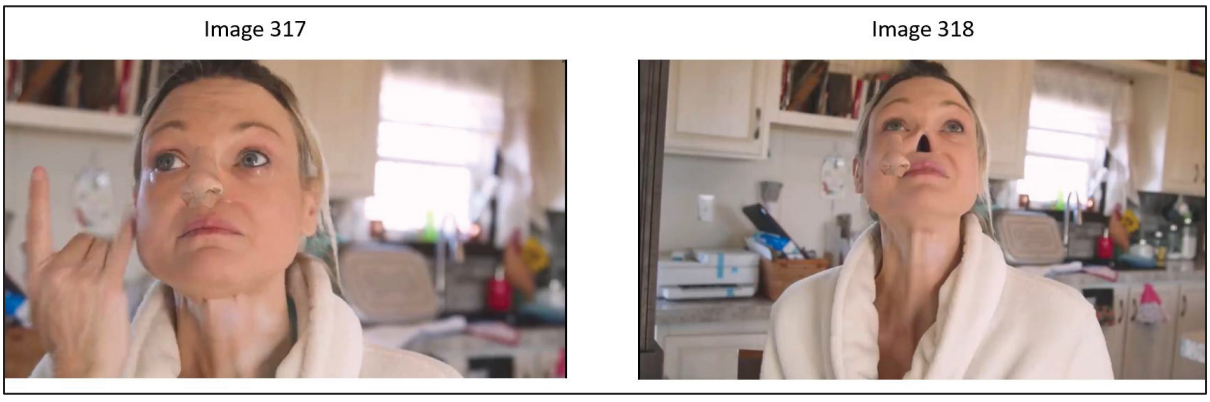

Figure 5.1 Exemple de deux images successives où le visage n'est pas au même endroit

Le filtre essaie de fluidifier le changement instantané de position, et décale donc la position du modèle d'épithèse, ce qui le fait apparaître au mauvais endroit, et ne respecte pas le critère de robustesse. Ces changements de positions et ces occlusions ne sont pas présents dans les vidéos de la base de données de l'UPNA, ce qui explique les meilleurs résultats. Les occlusions et les coupures des vidéos peuvent être simplement résolues par l'utilisation prévue pour le « Miroir Magique ». En effet cette application est prévue pour être utilisée par les patients dans un environnement contrôlé. Il est donc possible d'indiquer au patient de ne pas placer sa main sur son visage pour éviter les occlusions, et la vidéo sera continue car l'utilisation est en temps réel, ce qui résout ces deux problèmes.

Le « Miroir Magique » présente des problèmes au niveau de la robustesse et de la vitesse de calcul. Ces deux problèmes peuvent être résolus facilement, grâce à l'utilisation de processeurs graphiques, et en donnant des consignes d'utilisation aux patients. Sans ses problèmes il est possible de considérer que les résultats se rapprochent plus de ceux obtenus grâce à la base de données UPNA, qui sont assez bon pour pouvoir être utilisés en clinique.

CONCLUSION

L'objectif principal de ce projet était de développer une méthode non intrusive permettant aux patients ayant perdu leur nez de visualiser leur nouvelle apparence avec une épithèse virtuelle, en temps réel, sans recourir aux méthodes invasives traditionnelles. Face à une méthode de création d'épithèse actuellement longue, coûteuse et intrusive, nécessitant de multiples essais physiques, le projet « Miroir Magique » propose une alternative immersive simulant la découverte devant un miroir.

Pour atteindre cet objectif, nous avons combiné plusieurs technologies de l'état de l'art en reconstruction faciale et en suivi de visage. La solution développée repose sur l'utilisation conjointe de 3DDFA V3 pour la génération de texture réaliste du nez et de 3DDFA V2 pour la reconstruction rapide du visage en temps réel. Le suivi robuste du visage a été assuré par l'intégration de BlazeFace via MediaPipe, capable de détecter les visages dans des conditions variées, incluant l'absence de nez. Des algorithmes de recalage 3D, notamment l'Analyse en Composantes Principales (ACP) couplée à l'algorithme ICP, ont permis un positionnement précis des modèles d'épithèse sur le visage du patient. Enfin, l'application du filtre Minilag a garanti la fluidité des mouvements et l'élimination des vibrations, sans introduire de retard perceptible.

Les résultats expérimentaux démontrent la viabilité de l'approche proposée. Le système atteint une précision moyenne de placement de 9,70 pixels après alignement initial, avec une dérive inférieure à un pixel sur des images fixes, ce qui confirme la stabilité du suivi. La vibration des rotations et des translations a été significativement réduit grâce au filtrage, avec des écarts moyens d'orientation de 7,82° par rapport à la vérité terrain. La robustesse du système a été validée sur des vidéos de visages avec et sans nez, respectant le critère de 95% du modèle d'épithèse dans la zone d'intérêt dans la majorité des cas, sauf lors d'occlusions importantes ou de coupures vidéo non prévues dans l'utilisation réelle.

Toutefois, le système présente certaines limitations qui méritent d'être soulignées. La vitesse de traitement qui est actuellement de 4,33 images par seconde reste insuffisante pour une expérience parfaitement fluide en temps réel, bien que l'utilisation d'un processeur graphique pourrait facilement remédier à cette limitation. Les occlusions du visage et les changements

brusques de position perturbent temporairement le placement du modèle d'épithèse, mais ces situations sont évitables dans le cadre d'utilisation prévu, dans un environnement contrôlé avec des instructions appropriées au patient. La validation expérimentale a également été limitée par l'absence de base de données de patients sans nez avec vérité terrain, nécessitant l'utilisation de vidéos documentaires et une labélisation manuelle.

Malgré ces limites, le « Miroir Magique » représente une avancée significative dans l'accompagnement des patients oncologiques ayant subi une opération altérante du visage. Il offre une méthode non intrusive, rapide et immersive pour visualiser différentes options d'épithèse avant leur fabrication définitive. Cette approche s'inscrit parfaitement dans le projet AUDACE et peut être intégrée à une application plus large de création de moules d'épithèse par impression 3D, facilitant ainsi le processus de reconstruction identitaire des patients.

En conclusion, ce projet démontre qu'il est possible de combiner avec succès les technologies de reconstruction faciale, de suivi de visage et de filtrage pour créer une expérience immersive et précise destinée aux patients ayant perdu leur nez. Le « Miroir Magique » ouvre la voie à une nouvelle approche dans la création d'épithèses faciales, plus respectueuse du patient et mieux adaptée aux besoins médicaux modernes.

ANNEXE I

RÉSULTAT DE ROBUSTESSE SUR LES VIDÉOS DE UPNA

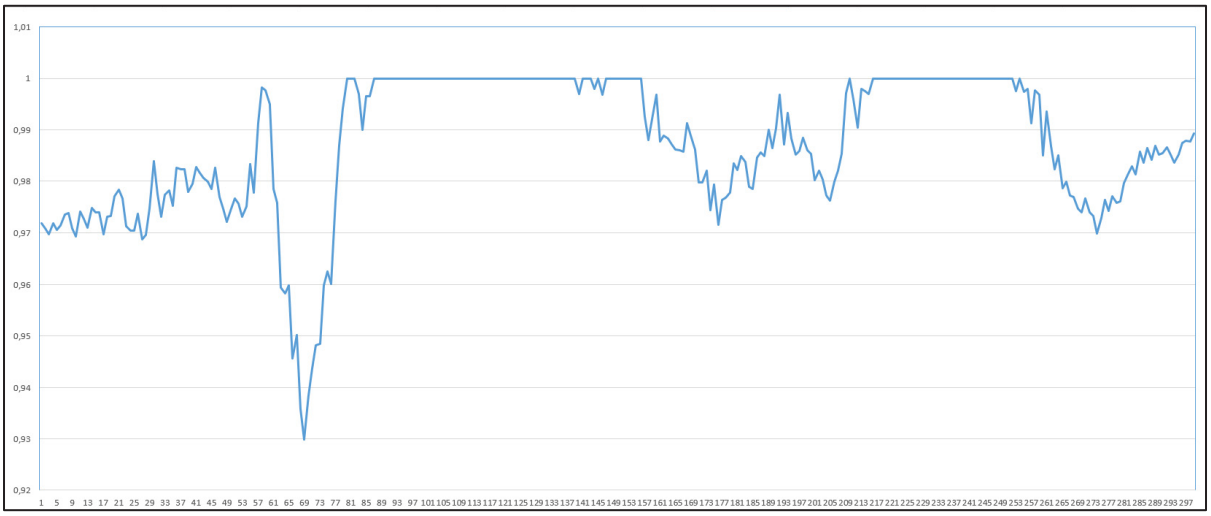

Figure-A I-1 Résultat de la robustesse pour chaque itération de la vidéo user_01_video_10

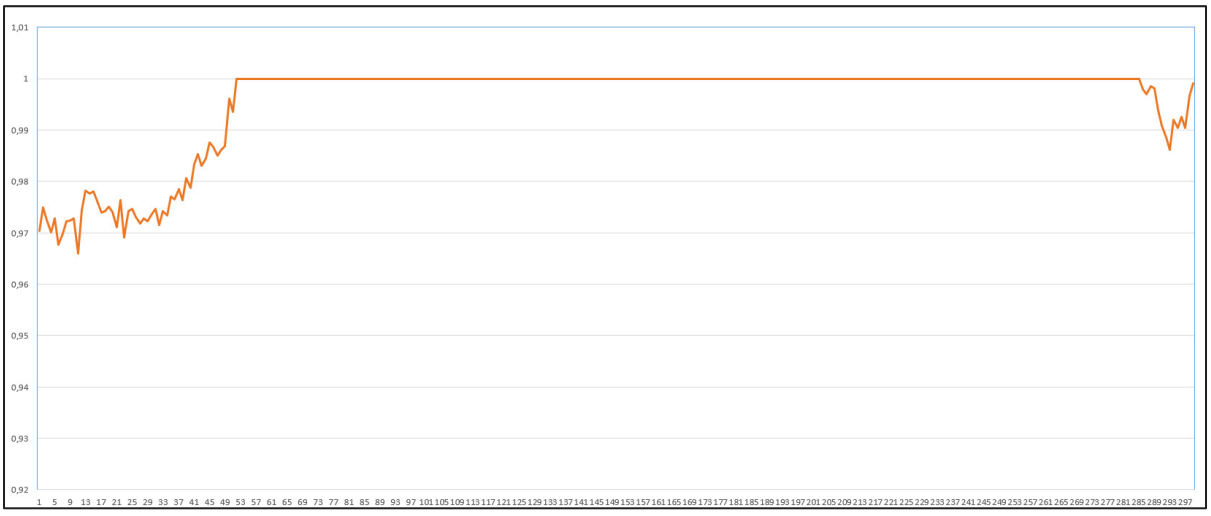

Figure-A I-2 Résultat de la robustesse pour chaque itération de la vidéo user_01_video_11

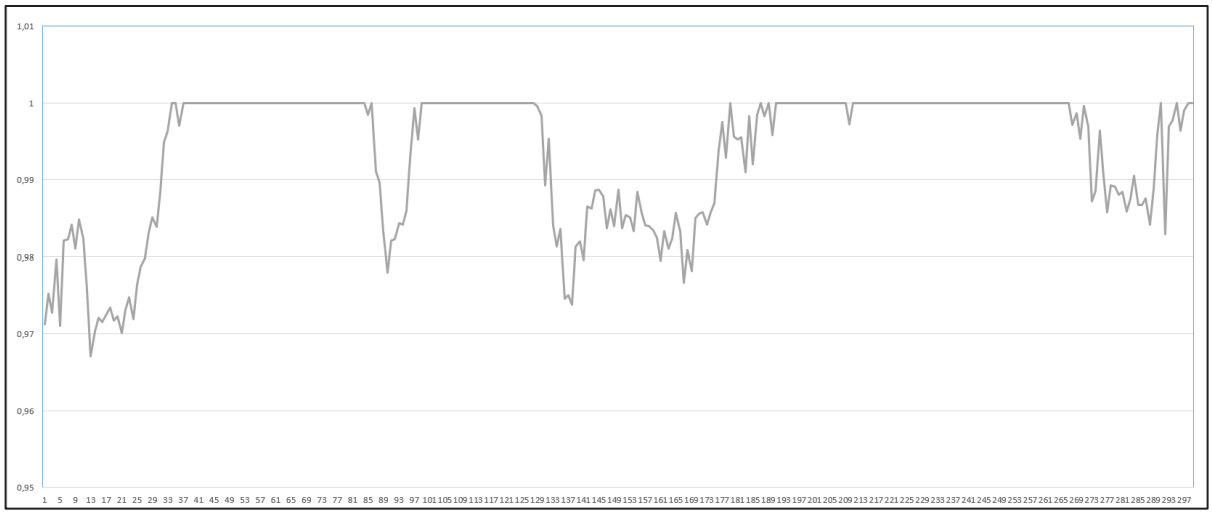

Figure-A I-3 Résultat de la robustesse pour chaque itération de la vidéo user_01_video_12

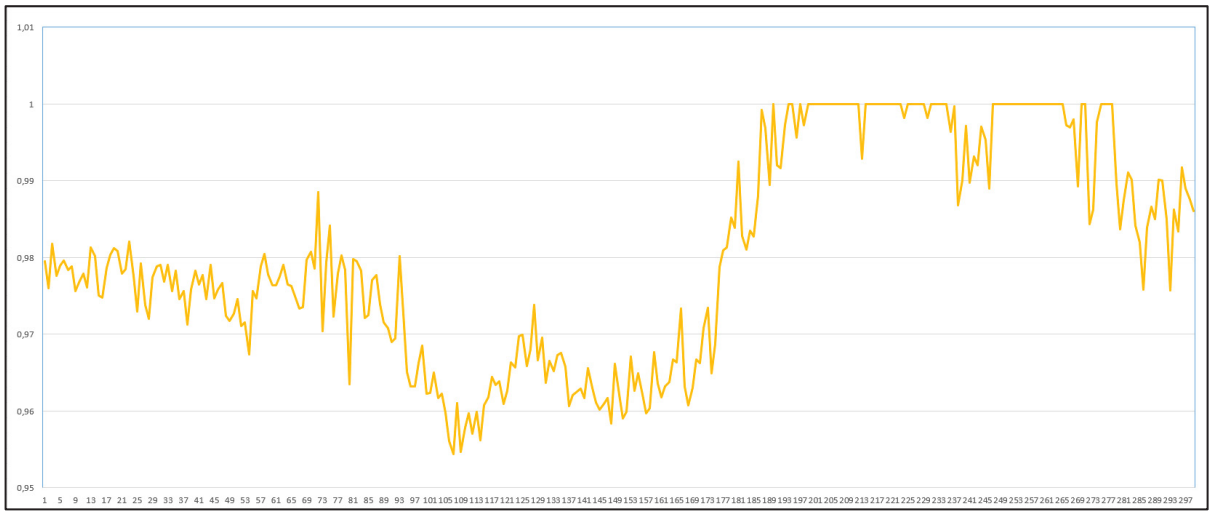

Figure-A I-4 Résultat de la robustesse pour chaque itération de la vidéo user_02_video_10

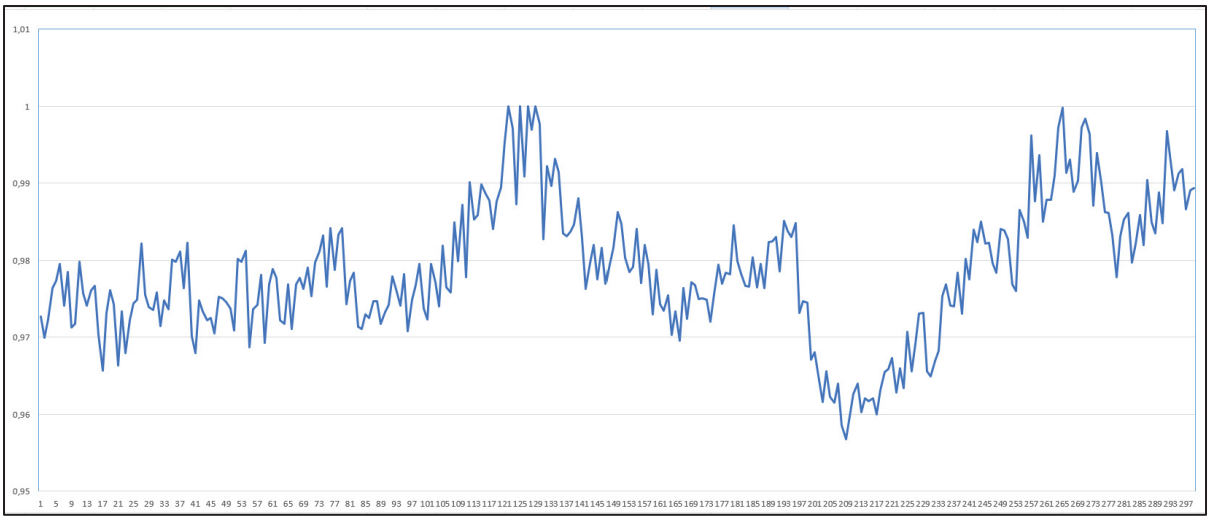

Figure-A I-5 Résultat de la robustesse pour chaque itération de la vidéo user_02_video_11

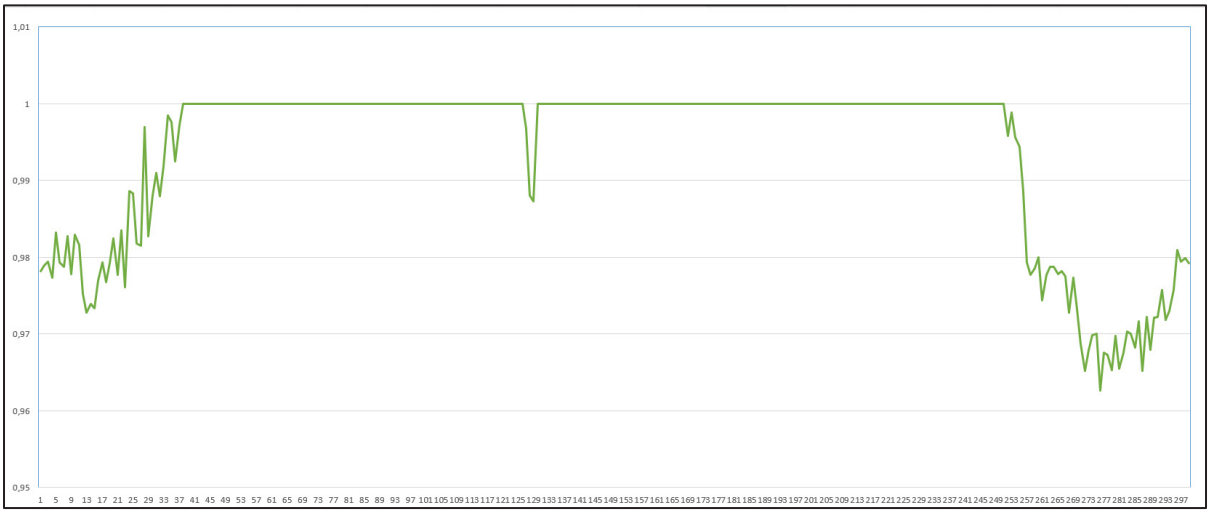

Figure-A I-6 Résultat de la robustesse pour chaque itération de la vidéo user_02_video_12

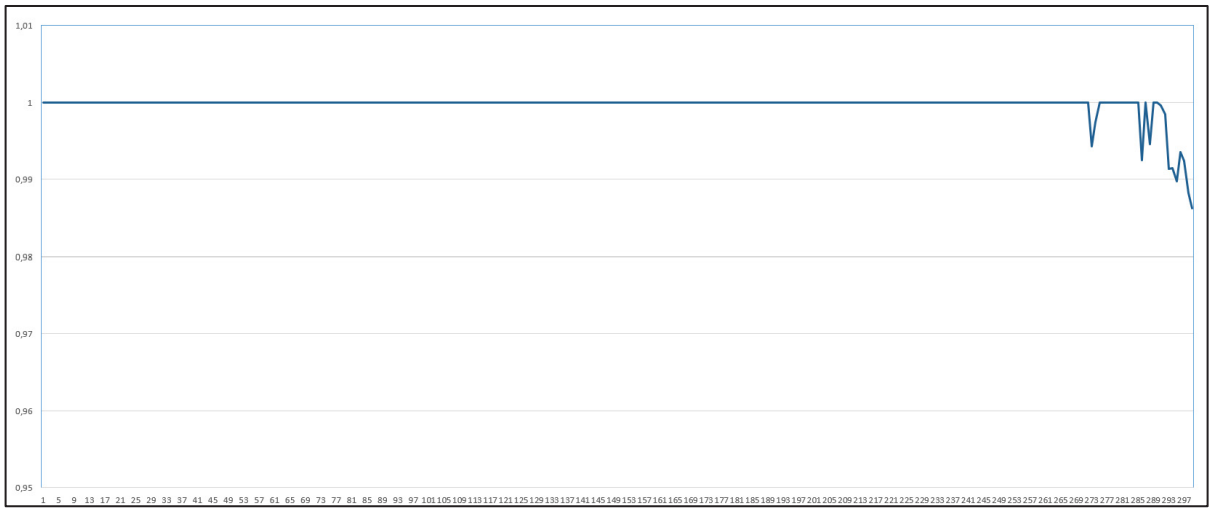

Figure-A I-7 Résultat de la robustesse pour chaque itération de la vidéo user_03_video_10

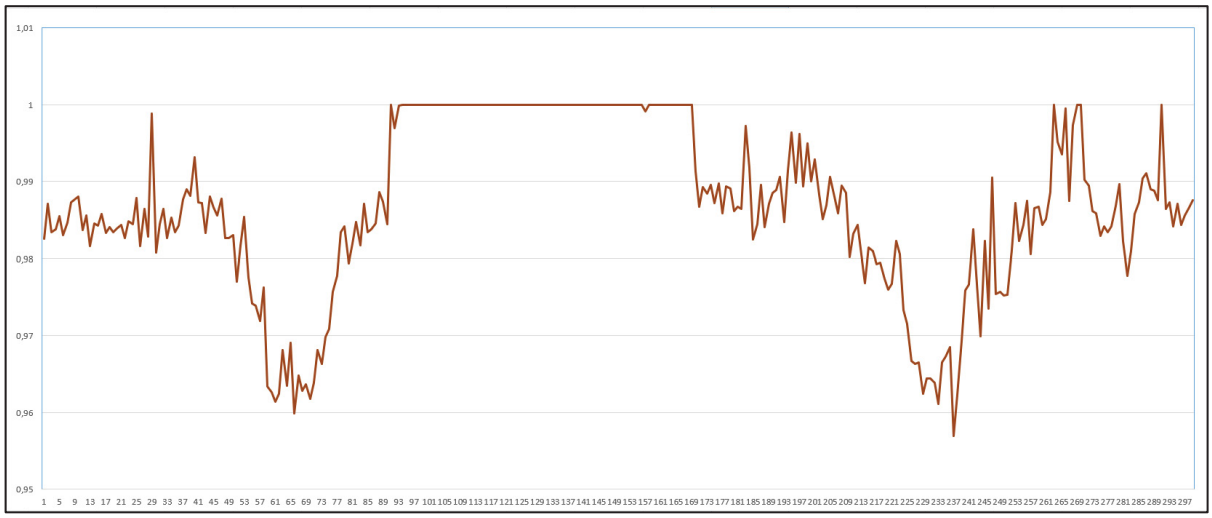

Figure-A I-8 Résultat de la robustesse pour chaque itération de la vidéo user_03_video_11

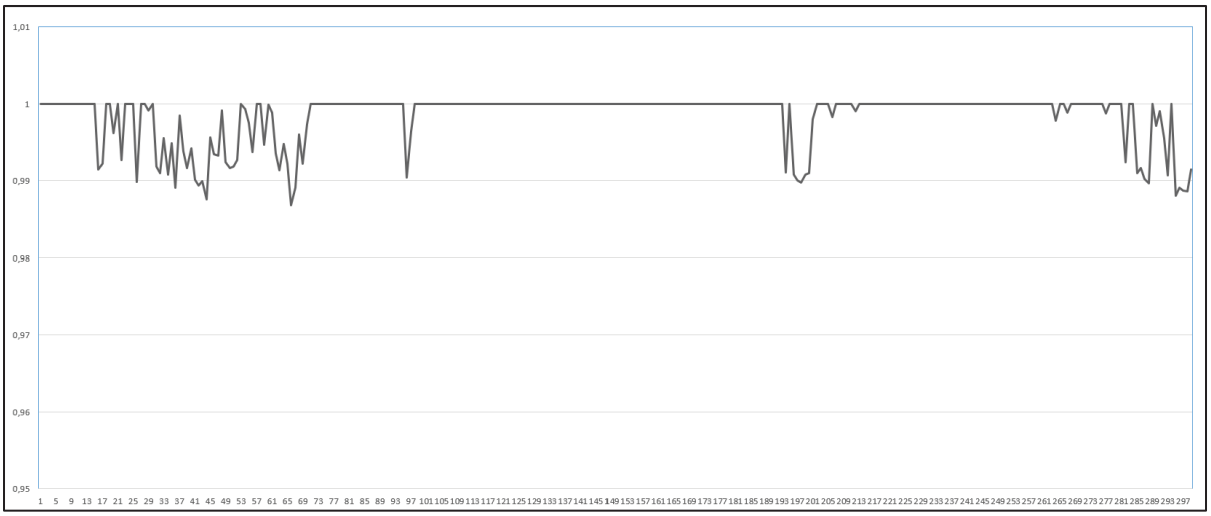

Figure-A I-9 Résultat de la robustesse pour chaque itération de la vidéo user_03_video_12

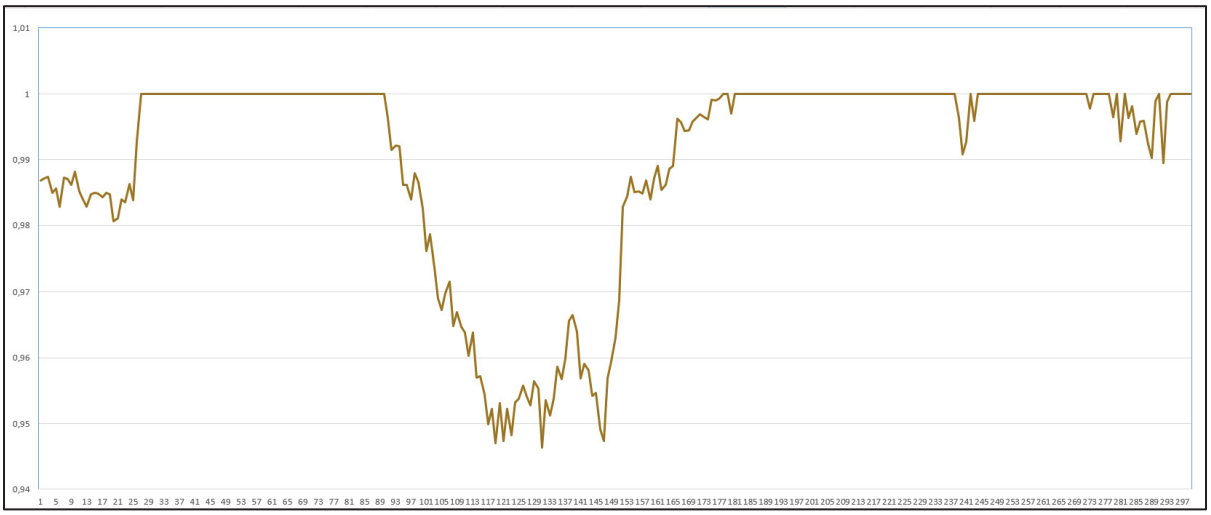

Figure-A I-10 Résultat de la robustesse pour chaque itération de la vidéo user_04_video_10

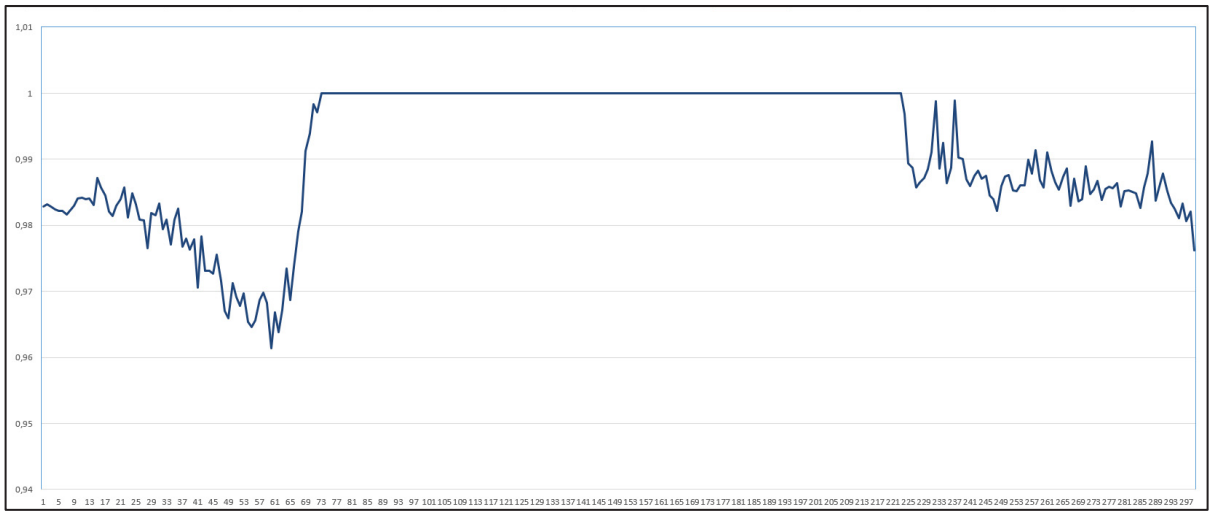

Figure-A I-11 Résultat de la robustesse pour chaque itération de la vidéo user_04_video_11

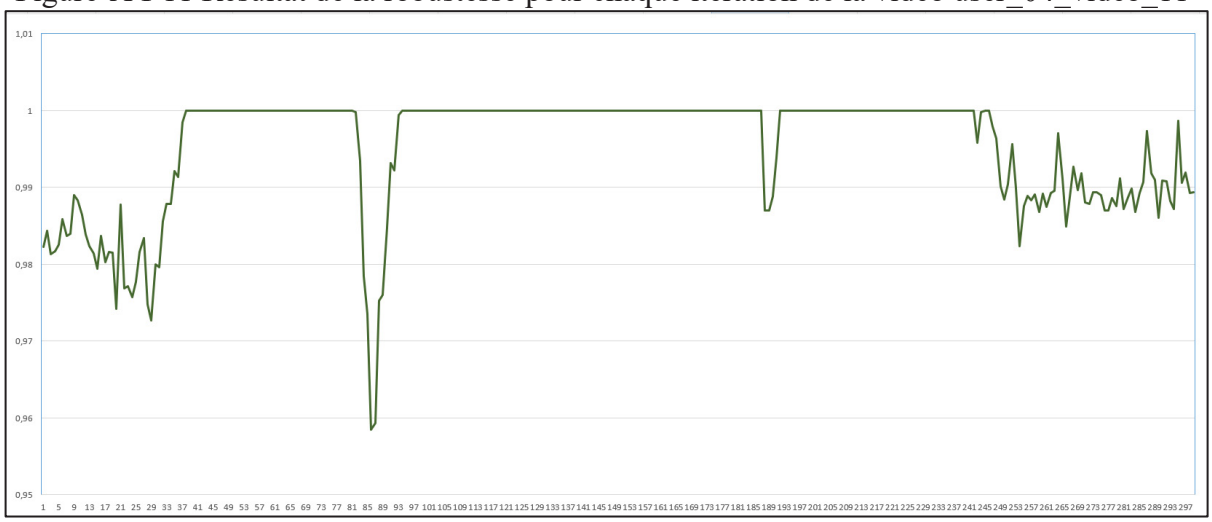

Figure-A I-12 Résultat de la robustesse pour chaque itération de la vidéo user_04_video_12

ANNEXE II

VIDÉO DE RÉSULTATS

Lien vers une vidéo de démonstration du logiciel, ainsi que des résultats :
https://youtu.be/RCpYtBUWqAA

BIBLIOGRAPHIE

After losing her nose to cancer, Becky shares her story to help others. (2012, 5 septembre). Repéré à https://www.youtube.com/watch?v=5srCCIvcWhQ

Ahuja, K., Ofek, E., Gonzalez-Franco, M., Holz, C., & Wilson, A. D. (2021). CoolMoves: User Motion Accentuation in Virtual Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(2), 1-23. https://doi.org/10.1145/3463499

Allen, B., Curless, B., & Popović, Z. (2002). Articulated body deformation from range scan data. *ACM Trans. Graph.*, *21*(3), 612-619. https://doi.org/10.1145/566654.566626

Amirgaliyev, B., Mussabek, M., Rakhimzhanova, T., & Zhumadillayeva, A. (2025). A Review of Machine Learning and Deep Learning Methods for Person Detection, Tracking and Identification, and Face Recognition with Applications. *Sensors*, *25*(5), 1410. https://doi.org/10.3390/s25051410

Ariz, M., Bengoechea, J. J., Villanueva, A., & Cabeza, R. (2016). A novel 2D/3D database with automatic face annotation for head tracking and pose estimation. *Computer Vision and Image Understanding*, *148*, 201-210. https://doi.org/10.1016/j.cviu.2015.04.009

Baker, S., & Matthews, I. (2004). Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, *56*(3), 221-255. https://doi.org/10.1023/B:VISI.0000011205.11775.fd

Bannink, T., De Ridder, M., Bouman, S., Van Alphen, M. J. A., Van Veen, R. L. P., Van Den Brekel, M. W. M., & Karakullukçu, M. B. (2024). Computer-aided design and fabrication of nasal prostheses: a semi-automated algorithm using statistical shape modeling. *International Journal of Computer Assisted Radiology and Surgery*, *19*(11), 2279-2285. https://doi.org/10.1007/s11548-024-03206-y

Batmaz, A. U., & Stuerzlinger, W. (2023). Rotational and Positional Jitter in Virtual Reality Interaction in Everyday VR. Dans A. Simeone, B. Weyers, S. Bialkova, & R. W. Lindeman (Éds), *Everyday Virtual and Augmented Reality* (pp. 89-118). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-031-05804-2_4

Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019, 14 juillet). BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. arXiv. https://doi.org/10.48550/arXiv.1907.05047

Beauchemin, S. S., & Barron, J. L. (1995). The computation of optical flow. *ACM Computing Surveys*, *27*(3), 433-466. https://doi.org/10.1145/212094.212141

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, *18*(9), 509-517. https://doi.org/10.1145/361002.361007

Besl, P. J., & McKay, N. D. (1992). Method for registration of 3-D shapes. Dans *Sensor Fusion IV: Control Paradigms and Data Structures* (Vol. 1611, pp. 586-606). SPIE. https://doi.org/10.1117/12.57955

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. Dans *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (pp. 187-194). USA : ACM Press/Addison-Wesley Publishing Co. https://doi.org/10.1145/311535.311556

Booth, J., Roussos, A., Ponniah, A., Dunaway, D., & Zafeiriou, S. (2018). Large Scale 3D Morphable Models. *Int. J. Comput. Vision*, *126*(2-4), 233-254. https://doi.org/10.1007/s11263-017-1009-7

Bouaziz, S., Tagliasacchi, A., Li, H., & Pauly, M. (2016). Modern techniques and applications for real-time non-rigid registration. Dans *SIGGRAPH ASIA 2016 Courses* (pp. 1-25). Macau : ACM. https://doi.org/10.1145/2988458.2988490

Carmack, J. (2013). Latency mitigation strategies (by John Carmack). Repéré à https://danluu.com/latency-mitigation/

Casiez, G., Roussel, N., & Vogel, D. (2012). 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. Dans *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2527-2530). Austin Texas USA : ACM. https://doi.org/10.1145/2207676.2208639

Chai, Z., Zhang, H., Ren, J., Kang, D., Xu, Z., Zhe, X., … Bao, L. (2022). REALY: Rethinking the Evaluation of 3D Face Reconstruction. Dans S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Éds), *Computer Vision – ECCV 2022* (pp. 74-92). Cham : Springer Nature Switzerland.

Chrysos, G. G., Antonakos, E., Snape, P., Asthana, A., & Zafeiriou, S. (2018). A Comprehensive Performance Evaluation of Deformable Face Tracking “In-the-Wild”. *International Journal of Computer Vision*, *126*(2), 198-232. https://doi.org/10.1007/s11263-017-0999-5

Cook, R. L., & Torrance, K. E. (1982). A Reflectance Model for Computer Graphics. *ACM Trans. Graph.*, *1*(1), 7-24. https://doi.org/10.1145/357290.357293

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6), 681-685. https://doi.org/10.1109/34.927467

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. Dans *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5202-5211). https://doi.org/10.1109/CVPR42600.2020.00525

Destruhaut, F., Vigarios, E., Andrieu, B., & Pomar, P. (2011). Regard anthropologique en Prothèse Maxillo-Faciale : entre science et conscience. *Chimères*, *75*(1), 45-56. https://doi.org/10.3917/chime.075.0045

Diao, H., Jiang, X., Fan, Y., Li, M., & Wu, H. (2024). 3D Face Reconstruction Based on a Single Image: A Review. *IEEE Access*, *12*, 59450-59473. https://doi.org/10.1109/ACCESS.2024.3381975

Dong, X., Yu, S.-I., Weng, X., Wei, S.-E., Yang, Y., & Sheikh, Y. (2018). Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors (pp. 360-368). Communication présentée au Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Repéré à https://openaccess.thecvf.com/content_cvpr_2018/html/Dong_Supervision-by-Registration_An_Unsupervised_CVPR_2018_paper.html

Engel, J., Schöps, T., & Cremers, D. (2014). LSD-SLAM: Large-Scale Direct Monocular SLAM. Dans D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Éds), *Computer Vision – ECCV 2014* (pp. 834-849). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-319-10605-2_54

Feng, Yao, Wu, F., Shao, X., Wang, Y., & Zhou, X. (2018). Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network. Dans V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Éds), *Computer Vision – ECCV 2018* (Vol. 11218, pp. 557-574). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-030-01264-9_33

Feng, Youyang, Wang, Q., & Zhang, H. (2019). Total Least-Squares Iterative Closest Point Algorithm Based on Lie Algebra. *Applied Sciences*, *9*(24), 5352. https://doi.org/10.3390/app9245352

Feng, Z.-H., Kittler, J., Awais, M., Huber, P., & Wu, X.-J. (2018). Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks. Dans *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2235-2245). https://doi.org/10.1109/CVPR.2018.00238

Filho, A. M., Laversanne, M., Ferlay, J., Colombet, M., Piñeros, M., Znaor, A., … Bray, F. (2025). The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. *International Journal of Cancer*, *156*(7), 1336-1346. https://doi.org/10.1002/ijc.35278

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, *24*(6), 381-395. https://doi.org/10.1145/358669.358692

Gao, H., Lim, M., Lin, E., & Green, R. (s.d.). Marker Based Facial Tracking Application in Communication Disorder Research.

Garrido, P., Valgaert, L., Wu, C., & Theobalt, C. (2013). Reconstructing detailed dynamic face geometry from monocular video. *ACM Trans. Graph.*, *32*(6), 158:1-158:10. https://doi.org/10.1145/2508363.2508380

Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schoenborn, S., & Vetter, T. (2018). Morphable Face Models - An Open Framework. Dans *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (pp. 75-82). https://doi.org/10.1109/FG.2018.00021

Grishchenko, I., Yan, G., Bazavan, E. G., Zanfir, A., Chinaev, N., Raveendran, K., … Sminchisescu, C. (s.d.). Blendshapes GHUM: Real-time Monocular Facial Blendshape Prediction.

Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z., & Li, S. Z. (2020). Towards Fast, Accurate and Stable 3D Dense Face Alignment. Dans A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Éds), *Computer Vision – ECCV 2020* (pp. 152-168). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-030-58529-7_10

Guo, X., Li, S., Yu, J., Zhang, J., Ma, J., Ma, L., … Ling, H. (2019, 3 mars). PFLD: A Practical Facial Landmark Detector. arXiv. https://doi.org/10.48550/arXiv.1902.10859

Guo, Y., zhang, juyong, Cai, J., Jiang, B., & Zheng, J. (2019). CNN-Based Real-Time Dense Face Reconstruction with Inverse-Rendered Photo-Realistic Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(6), 1294-1307. https://doi.org/10.1109/TPAMI.2018.2837742

Hall, B. (2015). Lie Algebras. Dans B. C. Hall (Éd.), *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction* (pp. 49-76). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-319-13467-3_3

Hasnaine, Q. R., Fouda, M. M., & Chiu, S. C. (2025). GPU vs. CPU: A Comparative Study of Performance, Architecture, and NVIDIA's Innovations in Modern Computing. Dans *2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI)* (pp. 1-6). https://doi.org/10.1109/ICMI65310.2025.11141315

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … Adam, H. (2017, 17 avril). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv. https://doi.org/10.48550/arXiv.1704.04861

Huang, X., Deng, W., Shen, H., Zhang, X., & Ye, J. (2020). PropagationNet: Propagate Points to Curve to Learn Structure Information. Dans *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7263-7272). https://doi.org/10.1109/CVPR42600.2020.00729

Huber, P., Hu, G., Tena, R., Mortazavian, P., Koppen, W. P., Christmas, W., … Kittler, J. (2016, février). A Multiresolution 3D Morphable Face Model and Fitting Framework. *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. [Proceedings Paper], SciTePress. Repéré à https://doi.org/10.5220/0005669500790086

Jackson, M., Clovin, T., Montiel, C., Bogdanova, E., Côté, C., Descoteaux, A., … Pomey, M.-P. (2023). Adopting a learning pathway approach to patient partnership in telehealth: A proof of concept. *PEC innovation*, *3*, 100223. https://doi.org/10.1016/j.pecinn.2023.100223

J'ai perdu mon nez à cause du cancer | SHAKE MY BEAUTY. (2021, 5 février). Repéré à https://www.youtube.com/watch?v=OBGU-l1ljFw

Jin, H., Liao, S., & Shao, L. (2021a). Pixel-in-Pixel Net: Towards Efficient Facial Landmark Detection in the Wild. *International Journal of Computer Vision*, *129*(12), 3174-3194. https://doi.org/10.1007/s11263-021-01521-4

Jin, H., Liao, S., & Shao, L. (2021b). Pixel-in-Pixel Net: Towards Efficient Facial Landmark Detection in the Wild. *International Journal of Computer Vision*, *129*(12), 3174-3194. https://doi.org/10.1007/s11263-021-01521-4

Kajalo, S., & Lindblom, A. (2016). The role of formal and informal surveillance in creating a safe and entertaining retail environment. *Facilities*, *34*(3-4), 219-232. https://doi.org/10.1108/F-06-2014-0055

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, *82*(Series D), 35-45.

Khabarlak, K., & Koriashkina, L. (2022). Fast Facial Landmark Detection and Applications: A Survey. *Journal of Computer Science and Technology*, *22*(1), e02-e02. https://doi.org/10.24215/16666038.22.e02

Kim, K.-S., Jang, D.-S., & Choi, H.-I. (2007). Real Time Face Tracking with Pyramidal Lucas-Kanade Feature Tracker. Dans O. Gervasi & M. L. Gavrilova (Éds), *Computational Science and Its Applications – ICCSA 2007* (pp. 1074-1082). Berlin, Heidelberg : Springer. https://doi.org/10.1007/978-3-540-74472-6_89

Kumar, A., Marks, T. K., Mou, W., Wang, Y., Jones, M., Cherian, A., … Feng, C. (2020). LUVLi Face Alignment: Estimating Landmarks' Location, Uncertainty, and Visibility Likelihood. Dans *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8233-8243). https://doi.org/10.1109/CVPR42600.2020.00826

Le Chartier, P. (2023). *GÉNÉRATION AUTOMATIQUE DE MOULES NUMÉRIQUES POUR LA CONFECTION DE PROTHÈSES FACIALES*. CRCHUM : ETS.

Lewis, J. P., Anjyo, K., Rhee, T., Zhang, M., Pighin, F., & Deng, Z. (s.d.). Practice and Theory of Blendshape Facial Models.

Li, H., Adams, B., Guibas, L. J., & Pauly, M. (2009). Robust single-view geometry and motion reconstruction. *ACM Trans. Graph.*, *28*(5), 1-10. https://doi.org/10.1145/1618452.1618521

Li, T., Bolkart, T., Black, M. J., Li, H., & Romero, J. (2017). Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics*, *36*(6), 1-17. https://doi.org/10.1145/3130800.3130813

Li, W., Lu, Y., Zheng, K., Liao, H., Lin, C., Luo, J., … Miao, S. (2020). Structured Landmark Detection via Topology-Adapting Deep Graph Learning. Dans A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Éds), *Computer Vision – ECCV 2020* (pp. 266-283). Cham : Springer International Publishing.

Li, Y., Wu, H., Wang, X., Qin, Q., Zhao, Y., wang, Y., & Hao, A. (2024, 4 juin). FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and Inpainting Guidance. arXiv. https://doi.org/10.48550/arXiv.2406.02074

Lin, C., Zhu, B., Wang, Q., Liao, R., Qian, C., Lu, J., & Zhou, J. (2021). Structure-Coherent Deep Feature Learning for Robust Face Alignment. *IEEE Transactions on Image Processing*, *30*, 5313-5326. https://doi.org/10.1109/TIP.2021.3082319

Liu, Y., Zhou, W., Yang, Z., Deng, J., & Liu, L. (2014). Globally consistent rigid registration. *Graphical Models*, *76*(5), 542-553. https://doi.org/10.1016/j.gmod.2014.04.003

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. Dans *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2* (pp. 674-679). San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., … Grundmann, M. (s.d.). MediaPipe: A Framework for Perceiving and Processing Reality.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T.-K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence*, *293*, 103448. https://doi.org/10.1016/j.artint.2020.103448

Man Who Lost His Nose To Cancer Gets Incredible Prosthetic Nose | Body Parts. (2022, 29 novembre). Repéré à https://www.youtube.com/watch?v=f41K9u0gvpM

Memon, A., Manjotho, A. A., Arain, Q. A., Sulaiman, A., Pirzada, N., Al Reshan, M. S., … Shaikh, A. (2025). Lightweight CNN-based head pose estimation using heatmaps and anthropometric facial measures. *ICT Express*, *11*(5), 914-918. https://doi.org/10.1016/j.icte.2025.06.006

Meng, Y., Chen, X., Gao, D., Zhao, Y., Yang, X., Qiao, Y., … Zheng, Y. (2022, 9 mars). 3D Dense Face Alignment with Fused Features by Aggregating CNNs and GCNs. arXiv. https://doi.org/10.48550/arXiv.2203.04643

Micaelli, P., Vahdat, A., Yin, H., Kautz, J., & Molchanov, P. (2023). Recurrence Without Recurrence: Stable Video Landmark Detection With Deep Equilibrium Models (pp. 22814-22825). Communication présentée au Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Repéré à https://openaccess.thecvf.com/content/CVPR2023/html/Micaelli_Recurrence_Without_Recurrence_Stable_Video_Landmark_Detection_With_Deep_Equilibrium_CVPR_2023_paper.html

Ming, X., Li, J., Ling, J., Zhang, L., & Xu, F. (2025). High-Quality Mesh Blendshape Generation from Face Videos via Neural Inverse Rendering. Dans A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol (Éds), *Computer Vision – ECCV 2024* (pp. 106-125). Cham : Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72897-6_7

Mitra, N. J., Gelfand, N., Pottmann, H., & Guibas, L. (2004). Registration of point cloud data from a geometric optimization perspective. Dans *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing* (pp. 22-31). New York, NY, USA : Association for Computing Machinery. https://doi.org/10.1145/1057432.1057435

Models | Shape modelling @ University of Basel. (s.d.). Repéré à https://shapemodelling.cs.unibas.ch/web/models/

Nguyen, D.-P., Nguyen, T.-N., Dakpé, S., Ho Ba Tho, M.-C., & Dao, T.-T. (2022). Fast 3D Face Reconstruction from a Single Image Using Different Deep Learning Approaches for Facial Palsy Patients. *Bioengineering*, *9*(11), 619. https://doi.org/10.3390/bioengineering9110619

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D Face Model for Pose and Illumination Invariant Face Recognition. Dans *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance* (pp. 296-301). https://doi.org/10.1109/AVSS.2009.58

Pearson, K. (1901). LIII. *On lines and planes of closest fit to systems of points in space*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572. https://doi.org/10.1080/14786440109462720

Prados-Torreblanca, A., Buenaposada, J. M., & Baumela, L. (2022). Shape Preserving Facial Landmarks with Graph Attention Networks. Dans *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press. Repéré à https://bmvc2022.mpi-inf.mpg.de/0155.pdf

Qin, Q., Li, Y., Wen, A., Zhu, Y., Gao, Z., Shan, S., … Wang, Y. (2024). Three-Dimensional Virtual Reconstruction of External Nasal Defects Based on Facial Mesh Generation Network. *Diagnostics*, *14*(6), 603. https://doi.org/10.3390/diagnostics14060603

Ranftl, A., Alonso-Fernandez, F., & Karlsson, S. (2015). Face Tracking Using Optical Flow. Dans *2015 International Conference of the Biometrics Special Interest Group (BIOSIG)* (pp. 1-5). https://doi.org/10.1109/BIOSIG.2015.7314604

Ross, D. A., Lim, J., Lin, R.-S., & Yang, M.-H. (2008). Incremental Learning for Robust Visual Tracking. *International Journal of Computer Vision*, *77*(1-3), 125-141. https://doi.org/10.1007/s11263-007-0075-7

Ruse, M. K., Calhoun, M., & Davis, B. K. (2024). Prosthetic Nasal Reconstruction. *Facial Plastic Surgery Clinics of North America*, *32*(2), 327-337. https://doi.org/10.1016/j.fsc.2023.12.002

Rusu, R. B., Blodow, N., & Beetz, M. (2009). Fast Point Feature Histograms (FPFH) for 3D registration. Dans *2009 IEEE International Conference on Robotics and Automation* (pp. 3212-3217). https://doi.org/10.1109/ROBOT.2009.5152473

Rusu, R. B., Blodow, N., Marton, Z. C., & Beetz, M. (2008). Aligning point cloud views using persistent feature histograms. Dans *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3384-3391). https://doi.org/10.1109/IROS.2008.4650967

Saito, S., Li, T., & Li, H. (2016). Real-Time Facial Segmentation and Performance Capture from RGB Input. Dans B. Leibe, J. Matas, N. Sebe, & M. Welling (Éds), *Computer Vision – ECCV 2016* (pp. 244-261). Cham : Springer International Publishing. https://doi.org/10.1007/978-3-319-46484-8_15

Shakir, S., & Al-Azza, A. (2022). Facial Modelling and Animation: An Overview of The State-of-The Art. *Iraqi Journal for Electrical and Electronic Engineering*, *18*(1), 28-37. https://doi.org/10.37917/ijeee.18.1.4

Sharma, S., & Kumar, V. (2022). 3D Face Reconstruction in Deep Learning Era: A Survey. *Archives of Computational Methods in Engineering*, *29*(5), 3475-3507. https://doi.org/10.1007/s11831-021-09705-4

Shen, J., Zafeiriou, S., Chrysos, G. G., Kossaifi, J., Tzimiropoulos, G., & Pantic, M. (2015). The First Facial Landmark Tracking in-the-Wild Challenge: Benchmark and Results. Dans *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)* (pp. 1003-1011). Santiago, Chile : IEEE. https://doi.org/10.1109/ICCVW.2015.132

Simplest Nasal Prosthesis. (2016, 6 février). Repéré à https://www.youtube.com/watch?v=MoSI24-UphY

Sirovich, L., & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *JOSA A*, *4*(3), 519-524. https://doi.org/10.1364/JOSAA.4.000519

Song, X., Xie, W., Li, J., Wang, N., Zhong, F., Zhang, G., & Qin, X. (2023). Minilag Filter for Jitter Elimination of Pose Trajectory in AR Environment. Dans *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 950-959). https://doi.org/10.1109/ISMAR59233.2023.00111

Sorkine-Hornung, O., & Rabinovich, M. (s.d.). Least-Squares Rigid Motion Using SVD.

Stauffert, J.-P., Niebling, F., & Latoschik, M. E. (2020). Latency and Cybersickness: Impact, Causes, and Measures. A Review. *Frontiers in Virtual Reality*, *1*. https://doi.org/10.3389/frvir.2020.582204

Tam, G. K. L., Cheng, Z.-Q., Lai, Y.-K., Langbein, F. C., Liu, Y., Marshall, D., … Rosin, P. L. (2013). Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid. *IEEE Transactions on Visualization and Computer Graphics*, *19*(7), 1199-1217. https://doi.org/10.1109/TVCG.2012.310

Thomas, D., & Taniguchi, R.-I. (2016). Augmented Blendshapes for Real-Time Simultaneous 3D Head Modeling and Facial Motion Capture. Dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3299-3308). Las Vegas, NV, USA : IEEE. https://doi.org/10.1109/CVPR.2016.359

Tiwari, H., Kurmi, V. K., Venkatesh, K. S., & Chen, Y.-S. (2022). Occlusion Resistant Network for 3D Face Reconstruction. Dans *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 297-306). https://doi.org/10.1109/WACV51458.2022.00037

Tomasi, C. (1991). Detection and Tracking of Point Features. Repéré à https://www.semanticscholar.org/paper/Detection-and-Tracking-of-Point-Features-Tomasi/88d55c8e7daddb0dc681494a934465b9d6630148

Vigarios, E., Destruhaut, F., Pomar, P., Dichamp, J., & Toulouse, E. (2015). *La prothèse maxillo-faciale*. Malakoff : CDP EDITIONS.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Dans *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001* (Vol. 1, p. I-I). https://doi.org/10.1109/CVPR.2001.990517

Wan, J., Liu, J., Zhou, J., Lai, Z., Shen, L., Sun, H., … Min, W. (2023). Precise Facial Landmark Detection by Reference Heatmap Transformer. *IEEE Transactions on Image Processing*, *32*, 1966-1977. https://doi.org/10.1109/TIP.2023.3261749

Wang, C.-W., & Peng, C.-C. (2021). 3D Face Point Cloud Reconstruction and Recognition Using Depth Sensor. *Sensors (Basel, Switzerland)*, *21*(8), 2587. https://doi.org/10.3390/s21082587

Wang, L., Chen, Z., Yu, T., Ma, C., Li, L., & Liu, Y. (2022). FaceVerse: A Fine-Grained and Detail-Controllable 3D Face Morphable Model From a Hybrid Dataset (pp. 20333-20342). Communication présentée au Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Repéré à https://openaccess.thecvf.com/content/CVPR2022/html/Wang_FaceVerse_A_Fine-Grained_and_Detail-Controllable_3D_Face_Morphable_Model_From_CVPR_2022_paper.html

Wang, X., Bo, L., & Fuxin, L. (2019). Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression. Dans *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 6970-6980). Seoul, Korea (South) : IEEE. https://doi.org/10.1109/ICCV.2019.00707

Wang, Z., Zhu, X., Zhang, T., Wang, B., & Lei, Z. (2024). 3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation. Dans *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1672-1682). Seattle, WA, USA : IEEE. https://doi.org/10.1109/CVPR52733.2024.00165

Welch, G., & Bishop, G. (1995). *An Introduction to the Kalman Filter*. USA : University of North Carolina at Chapel Hill.

Wijaya, S., Famuji, T., Mumin, M., Safitri, Y., Tristanti, N., Abdennasser, D., … Al-Sabur, R. (2025). Trends and Impact of the Viola-Jones Algorithm: A Bibliometric Analysis of Face Detection Research (2001-2024), *1*, 33-42. https://doi.org/10.59247/sjer.v1i1.8

Wilmott, J. P., Erkelens, I. M., Murdison, T. S., & Rio, K. W. (2022). Perceptibility of Jitter in Augmented Reality Head-Mounted Displays. Dans *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 470-478). Singapore, Singapore : IEEE. https://doi.org/10.1109/ISMAR55827.2022.00063

Wood, E., Baltrušaitis, T., Hewitt, C., Johnson, M., Shen, J., Milosavljević, N., … Valentin, J. (2022). 3D Face Reconstruction with Dense Landmarks. Dans S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Éds), *Computer Vision – ECCV 2022* (pp. 160-177). Cham : Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-19778-9_10

Wu, C.-Y., Xu, Q., & Neumann, U. (2021). *Synergy between 3DMM and 3D Landmarks for Accurate 3D Facial Geometry*. (S.l.) : (s.n.). https://doi.org/10.48550/arXiv.2110.09772

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., & Zhou, Q. (2018). Look at Boundary: A Boundary-Aware Face Alignment Algorithm. Dans *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2129-2138). Salt Lake City, UT, USA : IEEE. https://doi.org/10.1109/CVPR.2018.00227

Wu, Y., & Ji, Q. (2015). Robust Facial Landmark Detection Under Significant Head Poses and Occlusion. Dans *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 3658-3666). https://doi.org/10.1109/ICCV.2015.417

Wu, Y., & Ji, Q. (2019). Facial Landmark Detection: A Literature Survey. *International Journal of Computer Vision*, *127*(2), 115-142. https://doi.org/10.1007/s11263-018-1097-z

Xie, J., Wan, J., Shen, L., & Lai, Z. (2021). Think About Boundary: Fusing Multi-level Boundary Information for Landmark Heatmap Regression. Dans *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). https://doi.org/10.1109/IJCNN52387.2021.9534427

Xu, Z., Li, B., Yuan, Y., & Geng, M. (2021). AnchorFace: An Anchor-based Facial Landmark Detector Across Large Poses. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(4), 3092-3100. https://doi.org/10.1609/aaai.v35i4.16418

Yin, S., Wang, S., Chen, X., Chen, E., & Liang, C. (2020). Attentive One-Dimensional Heatmap Regression for Facial Landmark Detection and Tracking. Dans *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 538-546). New York, NY, USA : Association for Computing Machinery. https://doi.org/10.1145/3394171.3413509

Zafeiriou, S., Chrysos, G. G., Roussos, A., Ververas, E., Deng, J., & Trigeorgis, G. (2017). The 3D Menpo Facial Landmark Tracking Challenge. Dans *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)* (pp. 2503-2511). Venice, Italy : IEEE. https://doi.org/10.1109/ICCVW.2017.16

Zhang, J., Yao, Y., & Deng, B. (2022). Fast and Robust Iterative Closest Point. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3450-3466. https://doi.org/10.1109/TPAMI.2021.3054619

Zheng, Q., Deng, J., Zhu, Z., Li, Y., & Zafeiriou, S. (2022). Decoupled Multi-task Learning with Cyclical Self-Regulation for Face Parsing. Dans *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4146-4155). New Orleans, LA, USA : IEEE. https://doi.org/10.1109/CVPR52688.2022.00412

Zhu, M., Shi, D., Zheng, M., & Sadiq, M. (2019). Robust Facial Landmark Detection via Occlusion-Adaptive Deep Networks. Dans *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3481-3491). https://doi.org/10.1109/CVPR.2019.00360

Zhu, X., Liu, X., Lei, Z., & Li, S. Z. (2019). Face Alignment in Full Pose Range: A 3D Total Solution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(1), 78-92. https://doi.org/10.1109/TPAMI.2017.2778152

Zhuang, Y., Chen, L., Wang, X. S., & Lian, J. (2007). A Weighted Moving Average-based Approach for Cleaning Sensor Data. Dans *27th International Conference on Distributed Computing Systems (ICDCS '07)* (pp. 38-38). https://doi.org/10.1109/ICDCS.2007.83