

Estimation de mesures anthropométrique avec une seule image

par

Michael BURON YUEN

MÉMOIRE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
COMME EXIGENCE PARTIELLE À L'OBTENTION DE LA MAÎTRISE
AVEC MÉMOIRE EN GÉNIE LOGICIEL
M. Sc. A.

MONTRÉAL, LE 31 AOÛT 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Michael Buron Yuen, 2022



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

Carlos Vazquez, directeur de mémoire

Département de génie logiciel et des technologies de l'information à l'École de technologie supérieure

M. Éric Paquette, président du jury

Département de génie logiciel et des technologies de l'information à l'École de technologie supérieure

M. Sheldon Andrew, membre du jury

Département de génie logiciel et des technologies de l'information à l'École de technologie supérieure

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

LE 5 AOÛT 2022

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

J'aimerais remercier le directeur de mémoire, Dr Carlos Vazquez pour sa confiance et son support. Monsieur Vazquez m'a accompagné au long du baccalauréat et de la maîtrise afin d'assurer le développement de mes aptitudes au cours de mon parcours. Il a su me conseiller et m'outiller afin d'accomplir ce mémoire de façon efficace et complète. De la vision par ordinateur à l'intelligence artificielle, nos échanges ont permis de pousser les problèmes et leurs solutions plus loin que je ne l'aurais pensé.

Estimation de mesures anthropométriques avec une seule image

Michael BURON YUEN

RÉSUMÉ

Dans cette recherche, nous proposons une méthode pour obtenir des mesures anthropométriques d'un sujet à partir d'une seule image en mettant à l'échelle un avatar morphologique et normalisé du sujet dans le monde réel. Pour acquérir le facteur d'échelle, nous exploitons la distance à laquelle le sujet est placé par rapport à la caméra. Nous proposons donc une méthode pour estimer la distance entre la caméra et le sujet à partir d'une image monoculaire sans restreindre l'orientation de la tête. La distance inter pupillaire du sujet est utilisée comme point d'ancrage pour estimer la taille des pixels dans le plan du sujet et la distance à la caméra est récupérée à partir de cette mesure. L'écart inter pupillaire dans l'image est estimé en utilisant une approche basée sur l'apprentissage qui détecte la tête du sujet et estime l'orientation de la tête par rapport à l'axe optique de la caméra. Elle peut également être utilisée afin d'estimer la distance pupillaire à partir de 2 images dans le cas où cette information n'est pas disponible. Les approches proposées sont validées en comparant les résultats obtenus avec un système de capture de mouvement OptiTrack. Il est démontré que l'utilisation de l'algorithme proposé permet une estimation plus robuste de la distance indépendamment de la rotation de la tête. Il est également démontré que l'estimation de l'écart pupillaire est possible en utilisant seulement deux images. Enfin, la recherche fournit des résultats concluants en générant un avatar à l'échelle qui peut être utilisée pour des mesures anthropométriques.

Mots-clés: Distance inter pupillaire, rotation de la tête, mesure de la distance, mesures anthropométriques, mesures monoculaires

Estimation of Anthropometric Measurements from a Single Image

Michael BURON YUEN

ABSTRACT

In this research we propose a method to obtain anthropometric measurements of a subject from a single image by scaling a morphological and normalized avatar of the subject to the real world. To acquire the scaling factor, we leverage the distance at which the subject is placed from the camera. We therefore propose a method to estimate the distance between the camera and the subject from a monocular image without restricting the orientation of the head. The inter pupillary distance of the subject is used as an anchor to estimate the size of pixels in the plane of the subject and the distance to the camera is recovered from this measurement. The inter pupillary distance in the image is estimated by using a learning-based approach that detects the head of the subject and estimates the orientation of the head with respect to the optical axis of the camera. It also proposes a method to estimate the pupillary distance from two images in case this information is not available. The proposed approaches are validated by comparing the results obtained with an OptiTrack motion-capture system. It is demonstrated that using the proposed algorithm allows a more robust estimation of the distance regardless of the head rotation. It also shows that the estimation of the pupillary distance is possible using only two images. Lastly, the research provides successful results on providing a scaled avatar that can be used for anthropometric measurements.

Keywords: Inter pupillary distance, head orientation, distance measurements, anthropometric measurements, monocular image measurements

TABLE DES MATIÈRES

INTRODUCTION	13
Contexte	13
Problématique	14
CHAPITRE 1 RÉVISION DE LA LITTÉRATURE.....	17
1.1 Génération d'un avatar.....	17
1.1.1 Recherches antérieures.....	17
1.1.1.1 Ajustement d'un avatar surfacique normalisé.....	17
1.1.1.2 Génération de l'avatar surfacique	22
1.1.1.3 Comparatif des techniques	24
1.1.2 Estimation de pose	24
1.1.3 Récapitulatif l'obtention d'un avatar	28
1.2 Estimation de mesure à l'échelle à partir d'une image	29
1.2.1 Estimation de mesures à trois dimensions	29
1.2.2 Calibration du système.....	31
1.2.2.1 Calibration par objet de référence.....	32
1.2.2.2 Calibration active	32
1.2.2.3 Calibration par modèle planaire.....	33
1.2.2.4 Calibration par intelligence artificielle	33
1.2.2.5 Résumé des méthodes de calibrations.....	34
1.2.3 Outil d'analyse du visage à partir d'une image	35
1.2.4 Récapitulatif de l'obtention de mesure à l'échelle.....	39
1.3 Récapitulatif de la revue de littérature	39
CHAPITRE 2 MÉTHODOLOGIE.....	41
1.1 Ajustement d'un avatar à une image.....	43
2.1 Ajustement d'un avatar à une distance	44
2.2 Estimation de la distance pupillaire en utilisant deux images	50
CHAPITRE 3 EXPÉRIMENTATION.....	53
3.1 Contexte	54
3.2 Calibration.....	55
3.3 Protocole d'acquisition des données	56
CHAPITRE 4 RÉSULTATS ET DISCUSSION	61
4.1 Ajustement d'un avatar à partir d'une image.....	61
4.2 Ajustement de l'avatar à partir de la distance inter pupillaire	67
4.2.1 Angle de rotation.....	67
4.2.2 Estimation de la distance pupillaire corrigée dans l'image	70
4.2.3 Distance du sujet	75
4.2.4 Dimension du sujet	80
4.3 Estimation de la distance inter pupillaire.....	83

VIII

CHAPITRE 5	CONSIDÉRATIONS ÉTHIQUES	87
CONCLUSION	89
ANNEXE I	DÉMARCHE AFIN D'OBTENIR UNE MESURE À PARTIR DE DEUX IMAGES ET DEUX DISTANCES CONNUES	91
ANNEXE II	GRILLE DE MESURE D'UN SUJET	92
ANNEXE III	PROTOCOL D'ACQUISITION DE DONNÉS	93
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES.....		99

LISTE DES TABLEAUX

	Page
Tableau 4.1 Résultat de $ \Delta\theta $	69
Tableau 4.2 Erreur de PDP en fonction du lacet de la tête	72
Tableau 4.3 Erreur de PDP , $ \theta $	73
Tableau 4.4 Estimation de la distance en fonction de la rotation	77
Tableau 4.5 Erreur de la distance en fonction du lacet de la tête, $ \theta \leq 40^\circ$	78
Tableau 4.6 Erreur d'estimation de mesure anthropométrique.....	81
Tableau 4.7 PD mesuré comparé à l'estimé à ΔD de 2m	83
Tableau 4.8 Erreur d'estimation de PD en fonction de ΔD	85
Tableau 4.9 Erreur d'estimation de PD en fonction de ΔD - suite.....	86

LISTE DES FIGURES

	Page
Figure 1.1	Exemple de morphologie possible du modèle SMPL.....19
Figure 1.2	Exemple de position du modèle SMPL-X21
Figure 1.3	Processus SMPLify-X.....22
Figure 1.4	Exemple de résultat de PIFu23
Figure 1.5	Marqueurs du corps et pied de OpenPose.....26
Figure 1.6	Marqueurs de main de OpenPose27
Figure 1.7	Marqueurs du visage de OpenPose.....27
Figure 1.8	Marqueurs OpenFace36
Figure 1.9	Caractérisation de l'orientation de la tête37
Figure 1.10	Technique de détection de marqueur du visage.....38
Figure 1.11	Précision de l'estimation de la rotation de la tête38
Figure 2.1	Survol de la méthodologie proposée.....42
Figure 2.2	Méthodologie d'obtention d'un avatar à partir d'une image43
Figure 2.3	Calcul de <i>Ds</i>45
Figure 2.4	Vue de dessus du système.....46
Figure 2.5	Méthodologie d'estimation de PD50
Figure 3.1	Environnement de capture de données54
Figure 3.2	Schéma de damier55
Figure 3.3	Mesures anthropométriques partie 157
Figure 3.4	Mesures anthropométriques partie 2.....58
Figure 4.1	Sujet à la position initiale.....62
Figure 4.2	Détection de la pose sur sujet.....63

Figure 4.3	Superposition de l'avatar sur le sujet.....	64
Figure 4.4	Avatars générés.....	65
Figure 4.5	Courbe empirique du lacet du Sujet 10.....	67
Figure 4.6	Estimation du lacet du sujet 10.....	68
Figure 4.7	Erreur de $ \theta $ de 10 Sujet.....	69
Figure 4.8	Erreur de PD_P en fonction du lacet de la tête.....	71
Figure 4.9	Erreur de PD_P , $ \theta \leq 40^\circ$	73
Figure 4.10	Estimation de la distance en fonction de la rotation.....	76
Figure 4.11	Erreur de la distance en fonction du lacet de la tête, $ \theta \leq 40^\circ$	78
Figure 4.12	Comparatif de l'erreur d'estimation de distance entre PD et PD moyen.....	80
Figure 4.13	Avatar sur sujet 2.....	82
Figure 4.14	Erreur d'estimation de PD en fonction de ΔD	84

LISTE DES SYMBOLES ET UNITÉS DE MESURE

UNITÉS ET SYMBOLES DE BASE

Cm	Centimètre (unité de longueur)
Mm	Millimètre (unité de longueur)
moy.	Moyenne

ACCRONYME

<i>PD</i>	Distance pupillaire d'un sujet
------------------	--------------------------------

INTRODUCTION

Contexte

Rétrécir l'écart entre le monde virtuel et le monde réel est un problème que plusieurs recherches tentent de résoudre depuis plusieurs années. Une importante étape à la résolution de ce problème est d'être capable d'observer, d'analyser, et de manipuler une représentation numérique précise d'un humain. Acquérir cette donnée peut être grandement bénéfique dans différents secteurs tels que la médecine, le divertissement, la vente, etc.

Par exemple, lors de situation d'urgence, pouvoir automatiquement évaluer la grandeur et les mesures d'un sujet lorsque celui entre à l'urgence pourrait permettre d'identifier rapidement le traitement nécessaire sur des brûlures ou autre blessure physique. Avec des réseaux de neurones et des classificateurs, il serait potentiellement possible d'identifier le niveau d'urgence du traitement et ainsi catégoriser automatiquement la priorité du traitement. En y ajoutant d'autres variables telles que les mesures anthropométriques et le poids du sujet, cette solution pourrait suggérer la dose d'une prescription au docteur pour un sujet dédié. Il serait aussi possible d'utiliser un simulateur afin de mieux comprendre les effets reliés à la pose d'une prothèse ou de personnaliser l'implant à un individu en particulier.

Les jeux vidéo à réalité virtuelle incluent souvent une numérisation partielle ou complète de l'être humain dans le monde virtuel (Desai, Desai, Ajmera, & Mehta, 2014). Ceci permet une plus grande immersion dans le jeu vidéo ou l'activité entreprise telle qu'un sport. Une représentation précise d'un sujet permet un meilleur effet d'immersion. Par exemple, adéquatement représenter un dinosaure dans un jeu à l'ère de Jurassique ou bien la grosseur d'une voiture dans un jeu de course par rapport à l'avatar. Il peut aussi être utilisé par des professionnels tels que des joueurs de golf ou des sportifs olympiques afin de pratiquer leurs professions dans un monde virtuel (Govil, You, & Neumann, 2000).

Un détaillant de vêtement pourrait investir dans un écran interactif qui serait placé dans un centre d'achat. Cet écran permettrait de naviguer au travers l'inventaire du magasin et

d'essayer virtuellement le vêtement et la grandeur de celui-ci sur un avatar représentatif du client incitant ainsi non seulement l'achat d'article, mais créant un point de vente potentielle à très faible coût. Ceci pourrait aussi diminuer l'inventaire nécessaire dans les détaillants puisque les clients pourraient faire livrer directement chez eux le matériel commandé. Ils pourraient faire la totalité des essais dans le confort de leurs maisons. Il diminuerait aussi les retours potentiels dus à des mauvaises grandeurs ou insatisfactions du produit.

L'analyse de l'être humain et de ses mesures anthropométriques peut être aussi utilisée dans la surveillance et la sécurité. Connaître les dimensions d'un suspect permet de mieux cibler la liste des potentiels suspects. Pouvoir obtenir automatiquement ces informations augmenterait la rapidité de distributions d'informations.

Problématique

Tous ces cas d'utilisation possibles suscitent une grande attention à l'égard du problème de l'obtention d'une représentation numérique 3D précise de l'être humain. Ce mémoire tente de répondre à la problématique en se concentrant sur l'utilisation d'une seule image comme entrée afin de générer un avatar 3D numérique, échelonné, amovible et morphologiquement représentatif d'un humain. Une seule image d'un sujet est récupérable dans plusieurs systèmes et contextes. De plus, une méthodologie développée sur une seule image peut aussi être appliquée lorsque plusieurs images sont disponibles. C'est pour cela que la recherche se concentre sur une seule image comme entrée.

Ce problème est vaste et comporte de nombreuses variables. Ainsi, ce mémoire contraint certaines variables afin de proposer une preuve de concept initiale qui peut être itérée et améliorée par la suite.

Les contraintes appliquées sont :

- Aucune occlusion entre le sujet et la caméra;
- L'entièreté du sujet est visible sur l'image;
- Il y a un seul sujet présent sur l'image;
- Il n'y a aucun problème de qualité ou netteté de l'image.

Aucune recherche trouvée au moment de ce mémoire ne répond directement à la problématique utilisant une seule image. La majorité des recherches existantes requiert soit plusieurs images ou bien l'avatar n'est pas amovible et à l'échelle (Alldieck, Magnor, Xu, Theobalt, & Pons-Moll, 2018; Saito et al., 2019). En revanche, deux volets de recherches se rapprochent d'une solution:

- Obtention d'un avatar à partir d'une image;
- Obtention de mesure à l'échelle dans une image.

Une hypothèse est émise qu'il est possible de combiner ces deux domaines d'études afin de constituer une seule méthodologie qui apportera une solution unique à la problématique. Plus précisément qu'il est possible d'obtenir un avatar à 3 dimensions et de le mettre à l'échelle en utilisant une mesure de références.

En analysant les recherches qui permettent d'obtenir des mesures anthropométriques à partir d'un humain (Barron & Kakadiaris, 2000), il est possible d'identifier que ceux-ci calibrent leur système composé d'une caméra et d'un sujet humain, récupèrent les marqueurs de visage et les utilisent en conjonction avec des facteurs relationnels théoriques de l'être humain afin de calculer les diverses mesures. Cependant, il n'y a pas de représentation numérique 3D et donc pas toutes les mesures peuvent être prises dépendant de l'angle de la caméra. Lorsque le même exercice est fait sur les recherches sur la génération d'avatar (Pavlakos et al., 2019), l'estimation d'une pose ainsi que l'utilisation de cette pose afin de générer ou manipuler un avatar sont les étapes génériques de la méthodologie. Cependant, cet avatar n'est pas à

l'échelle. Afin de valider l'hypothèse, il faut analyser chacune des composantes de ces recherches et comprendre les différentes possibilités de chacune.

Ce mémoire contribue trois points importants à la solution soit :

- Une méthodologie afin d'obtenir un avatar à l'échelle à partir d'une seule image;
- Une amélioration sur la méthodologie d'estimation de la distance d'un sujet à une caméra en utilisant la distance inter pupillaire et considérant la rotation du visage;
- Une méthodologie afin d'obtenir la distance inter pupillaire en utilisant deux images et la différence de distance entre celles-ci.

Le reste de ce document est décortiqué comme suit. Une revue de la littérature des deux systèmes composant l'hypothèse est faite afin de bien comprendre les avancées existantes et leurs contraintes. Puisqu'il nous faut connecter ces deux systèmes, il faut s'assurer que les entrées et les sorties sont compatibles. Ensuite, la méthodologie proposée est expliquée en utilisant les composantes qui répondent le mieux au besoin. Une expérience et ses résultats seront finalement présentés afin de démontrer l'efficacité des contributions à la problématique. En sommes, une conclusion récapitulera les contributions de cette recherche et les futures améliorations possibles.

CHAPITRE 1

RÉVISION DE LA LITTÉRATURE

La revue de littératures fait une première analyse des deux parties de l'hypothèse : la génération d'un avatar et l'estimation de mesures à l'échelle à partir d'une image. Elle cherche à identifier les avantages et les inconvénients des différentes méthodologies ainsi que de comprendre les composantes nécessaires dans chacun des systèmes. Bien les comprendre est nécessaire afin de générer un système unifié et fonctionnel qui permettra d'obtenir un avatar à l'échelle.

1.1 Génération d'un avatar

Cette section explore et compare les différentes méthodologies qui permettent d'obtenir une représentation digitale à trois dimensions d'un humain. Elle identifie aussi que l'estimation de la pose est nécessaire dans plusieurs recherches. Par conséquent, un survol des différentes recherches d'estimation de la pose est accru à cette section.

1.1.1 Recherches antérieures

Deux grandes catégories peuvent être identifiées dans l'obtention d'un avatar représentatif d'un sujet: la génération de l'avatar surfacique sans utilisation d'un modèle normalisé et l'ajustement d'un avatar surfacique normalisé. Un survol de chaque technique est fait afin de générer un comparatif.

1.1.1.1 Ajustement d'un avatar surfacique normalisé

L'ajustement d'un avatar surfacique tire profit d'un modèle surfacique normalisé et pré-généré. Ce modèle est composé de paramètres qui définissent et contraignent la façon dont il est possible de modifier le modèle afin que la représentation finale reste dans le plausible. Une des complexités de cette méthodologie est qu'il faut non seulement bien représenter l'humain en termes de positions possibles (pose), mais aussi adéquatement représenter la surface et la

dynamique de la peau (morphologie). Une série d'algorithmes basés sur les réseaux de neurones est utilisée afin d'estimer la déformation du modèle générique pour le faire correspondre avec le sujet sur l'image via un recalage 3D à 2D. Les réseaux de neurones sont entraînés sur plusieurs bases de données de numérisation d'humain tel que CEASAR (Robinette et al., 2002). Plusieurs auteurs tentent de répondre à cette problématique de recalage du modèle 3D aux images, telles que Dyna (Pons-Moll, Romero, Mahmood, & Black, 2015), BlendSCAPE (Hirshberg, Loper, Rachlin, & Black, 2012), MPII (Pishchulin, Wuhler, Helten, Theobalt, & Schiele, 2017). Cependant, un modèle nommé SMPL (Loper, Mahmood, Romero, Pons-Moll, & Black, 2015) ressort en particulier dans la littérature pour sa simplicité, précision, et efficacité.

Le modèle (SMPL) est un modèle surfacique 3D normalisé et linéaire de la morphologie d'un être humain (Alldieck et al., 2018). Supporté par un réseau de neurones et des régressions linéaires, SMPL est une tentative de représenter l'être humain en trois dimensions. Il est composé de points 3D et permet de représenter une grande variété de postures et morphologies de l'être humain en respectant les contraintes de la morphologie humaines telles que l'orientation et l'intersection des membres ainsi que l'angle de rotation des articulations. Le modèle est basé sur des paramètres de déformations qui appliquent une transformation sur le modèle afin de simuler une morphologie différente. Il est ensuite possible de positionner le modèle dans la pose désirée en utilisant des marqueurs de pose 2D. Différentes positions et morphologies sont données en exemple dans la Figure 1.1.



Figure 1.1 Exemple de morphologie possible du modèle SMPL
Tirée de Loper (2015, p.1)

La limitation de SMPL est qu'il ne permet pas de replacer les expressions faciales, les pieds, et les mains. Pour ce faire, les auteurs de SMPL ont poussé plus loin leur modèle en développant une version améliorée dénommée SMPL-X (Pavlakos et al., 2019). Ils approfondissent aussi l'utilisation des contraintes de positionnement du corps humain. Ils tentent ainsi de limiter le nombre de positionnements impossibles de l'avatar. Trois versions du modèle spécialisés aux différents genres (mâle, femelle, unisexe) ont été développés afin d'optimiser la précision d'après le sexe lorsque possible. Étant un avatar normalisé en trois dimensions, il offre l'opportunité de reproduire la morphologie en ajustant une série de paramètres. Le positionnement est possible grâce au squelette qui est intégré au modèle. Le maillage à trois dimensions est composé de 10 475 sommets et 54 articulations. Le modèle est défini par l'équation (1).

$$\mathbf{M}(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\psi}) = \mathbf{W}(\mathbf{T}_p(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\psi}), \mathbf{J}(\boldsymbol{\beta}), \boldsymbol{\theta}, \boldsymbol{\omega}) \in \mathbb{R}^{10475 \times 3} \quad (1)$$

$\boldsymbol{\beta}$: Morphologie du corps

$\boldsymbol{\psi}$: Expression faciale

$\boldsymbol{\theta}$: Pose du visage, des mains, et du corps

\mathbf{M} : Modèle SMPL-X

\mathbf{T}_p : Correction initiale du modèle de base SMPL

\mathbf{J} : Régression linéaire des points 3D

\mathbf{W} : Fonction linéaire de mélange de points

$\boldsymbol{\omega}$: Poids des points dans la fonction linéaire \mathbf{T}_p

La Figure 1.2 démontre des exemples de différentes poses et expressions du même modèle. Elle démontre qu'il est possible de varier les paramètres de configuration de SMPL-X afin de bien recaler la morphologie des mains, des pieds, ainsi que de l'expression faciale pour qu'elle suive celle du sujet. SMPL-X est grandement utilisé dans plusieurs recherches comme étant une référence de valeur théorique afin de tester les solutions proposées (Saito et al., 2019; Zheng, Yu, Wei, Dai, & Liu, 2019).



Figure 1.2 Exemple de position du modèle SMPL-X
Tirée de Pavlakos (2019, p. 2)

Le défi reste tout de même de trouver les bons paramètres afin de recaler le modèle à une image. Les auteurs de SMPL-X proposent un algorithme dénommé SMPLfy-X (Pavlakos et al., 2019) qui introduit la possibilité d'estimer les paramètres nécessaires afin d'adapter SMPL-X au sujet présent dans une image. Il utilise des marqueurs 2D du visage, des mains, des pieds, ainsi que le corps humain obtenu préalablement afin de positionner le modèle approximativement. Par la suite, ils utilisent un réseau de neurones ainsi qu'une caméra fictive afin d'estimer itérativement les paramètres de déformations (la pose, l'expression faciale, la morphologie) afin de mieux cadrer le modèle à l'image. Les sorties de l'algorithme sont le modèle ajusté et la translation de la caméra fictive d'ajustement. Le processus est visible via la Figure 1.3 :

1. Capture d'une image d'entrée
2. Estimation des marqueurs (faciale, corps, mains, pieds)
3. Repositionnement du modèle du corps à partir des marqueurs du corps
4. Estimation des paramètres de morphologies et application des marqueurs faciaux, mains, et pieds.

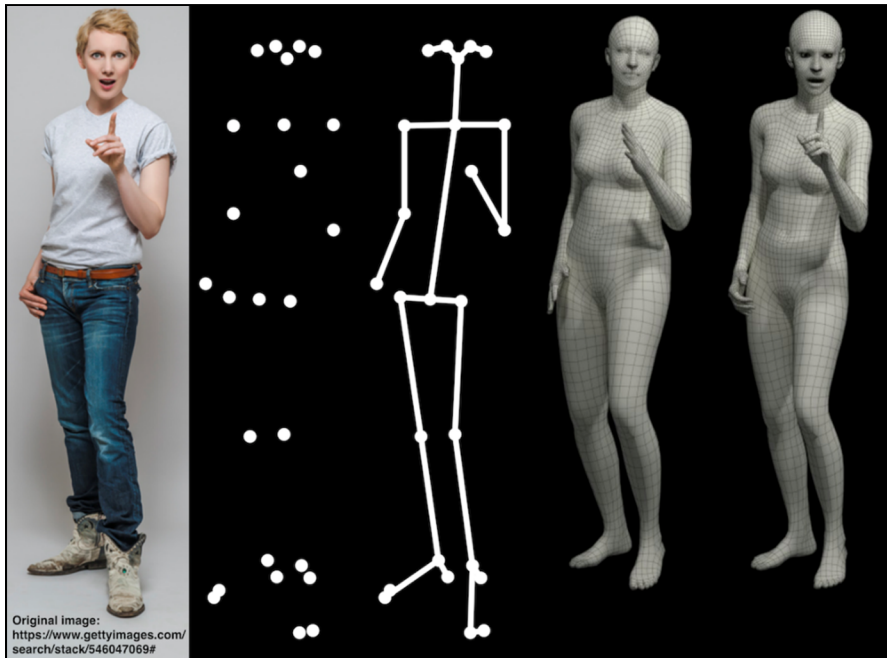


Figure 1.3 Processus SMPLify-X
 Tirée de Pavlakos (2019, p. 2)

1.1.1.2 Génération de l'avatar surfacique

La génération d'un avatar surfacique est basée sur la génération directe de l'avatar sans passer par un modèle générique (R. Li et al., 2020; Zheng et al., 2019). L'enlignement de pixel implicite (PIFu) présenté par Saito (Saito et al., 2019) est une méthode qui est utilisé dans plusieurs recherches (Lazova, Insaftudinov, & Pons-Moll, 2019; Ma et al., 2020; Xu, Wang, Ceylan, Mech, & Neumann, 2019). PIFu utilise premièrement un réseau de neurones afin de segmenter l'humain de l'image. Par la suite, il utilise un réseau de neurones entraîné afin d'estimer le champ scalaire à trois dimensions de l'être humain (Saito et al., 2019). Il génère l'avatar en utilisant la technique *marching cube* (Lorensen & Cline, 1987) qui consiste à parcourir le champ scalaire et d'isoler huit points à la fois afin de déterminer et générer les polygones qui ont besoin d'être créés. Il est aussi possible d'utiliser un système à plusieurs caméras afin d'améliorer le résultat. Un exemple à partir d'une seule image est démontré dans la Figure 1.4. Des chercheurs tel que (R. Li et al., 2020) ont amélioré cette méthode en donnant la possibilité d'exécuter l'algorithme en temps réel. Pour ce faire, au lieu d'utiliser un champ

scalaire linéaire, ils utilisent un encodage binaire basé sur un *octree* (Meagher, 1982) afin d'améliorer la rapidité et la précision. De plus, cette technique permet de créer une seule surface fermée au lieu de créer plusieurs surfaces qui sont par la suite reliées ensemble. Ceci permet d'obtenir environ 15 images par seconde.

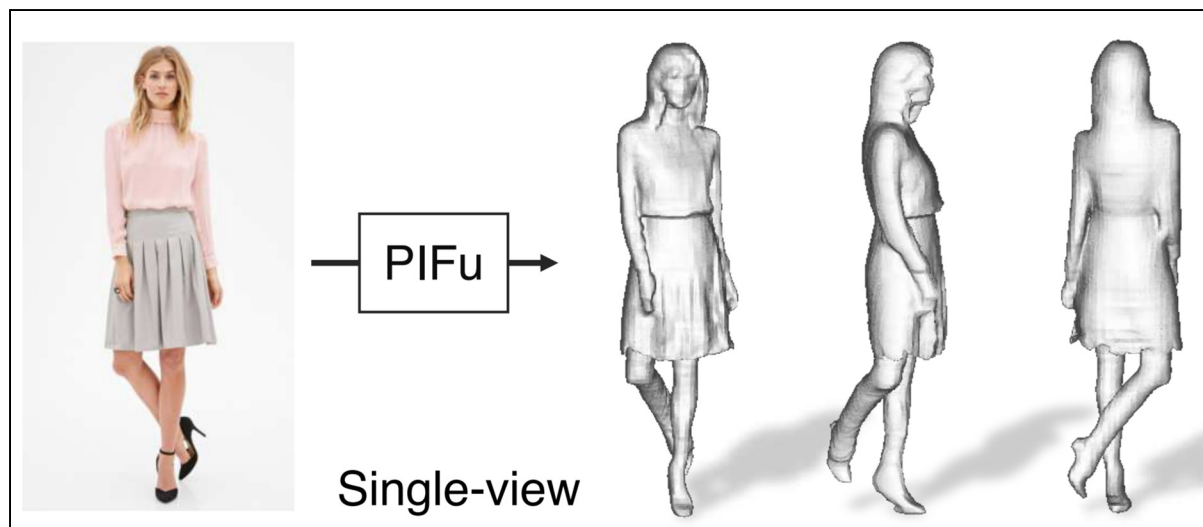


Figure 1.4 Exemple de résultat de PIFu
Tiré de Saito (2019, p.1)

Un autre exemple de génération complète de l'avatar est DeepHuman (Zheng et al., 2019) et DeepPose (Alp Güler, Neverova, & Kokkinos, 2018). L'algorithme proposé par les auteurs est séparé en trois parties : un encodeur d'image, un système neuronal sur le volume, et un bloc de polissage des normales. Ce système est principalement basé sur le projet BodyNet (Pavlakos et al., 2019). BodyNet est un algorithme proposé en 2019 qui émet une solution similaire, mais avec de plus faibles résultats lorsque comparé à DeepPose. Ces algorithmes utilisent SMPL (Loper et al., 2015) afin d'augmenter leurs banques de données d'entraînement en générant des données par l'ajustement du modèle.

1.1.1.3 Comparatif des techniques

Chacune des deux types de méthodologies comportent leurs lots d'avantages et d'inconvénients. La technique de génération complète de l'avatar apporte de meilleurs résultats de recalage que des techniques de recalage d'avatar tel que la technique utilisée par SMPLify-X (R. Li et al., 2020) ou DensePose (Alp Güler et al., 2018; R. Li et al., 2020; Saito et al., 2019). Cependant, cette comparaison est basée sur une technique volumétrique inclusive. Par conséquent, le volume mesuré inclut les vêtements et les accessoires portés par le sujet. De plus, il fournit un résultat surfacique non articulé. En d'autres mots, bien que le recalage soit adéquat sur une image, si l'objectif est d'avoir des mesures du sujet, celles-ci seront biaisées.

La technique d'ajustement d'un avatar surfacique normalisée permet d'avoir une représentation de l'humain en omettant les vêtements ou autres accessoires du sujet. La présence et l'utilisation d'un squelette, de sommets et de jonctions permettent aussi de faire des mesures systématiques. Les modèles ont aussi une dynamique de peau qui permet d'avoir une surface plus représentative du vrai sujet. Une fois les paramètres du modèle ajustés au sujet, il est possible de les manipuler en changeant sa pose et l'expression faciale tout en conservant la morphologie du sujet. Cependant, puisque la méthode utilise un modèle normalisé, le recalage n'est pas aussi précis que son antagoniste.

Malheureusement, aucune des deux techniques ne permet d'avoir un avatar à l'échelle. Toute solution ci-dessus énumérée génère un avatar qui est relatif à l'image fournie sans se soucier des réelles mesures anthropométriques.

1.1.2 Estimation de pose

La manipulation d'un avatar surfacique normalisé décrit dans la section 1.1.1.2 nécessite les marqueurs de poses 2D du corps, du visage, des pieds et des mains. Cette section explore les recherches existantes qui répondent à ce besoin.

Plusieurs algorithmes sont disponibles afin d'obtenir une pose (Ji et al., 2020). Il y en a qui permettent de détecter et d'estimer la pose d'une seule personne (Alp Güler et al., 2018; Cao, Hidalgo, Simon, Wei, & Sheikh, 2018) et d'autres qui permettent d'estimer la pose de plusieurs sujets présents dans une seule image (Andriluka, Roth, & Schiele, 2010). Puisque cette recherche se concentre sur l'obtention de mesure, seule la détection d'un seul individu sera prise en considération. Il existe deux tendances dans l'estimation de poses. Les algorithmes *top-down* détectent et isolent premièrement le sujet du reste de l'image avant d'estimer les marqueurs (Andriluka et al., 2010; Cao et al., 2018; Johnson & Everingham, 2010) tandis que les algorithmes *bottom-up* estiment les marqueurs dans l'image et tente ensuite de les regrouper (Cheng et al., 2020; M. Li, Zhou, Li, & Liu, 2018). Bien que le problème soit présent depuis longtemps dans le domaine de la recherche en vision par ordinateurs, l'estimation des marqueurs de la main et des pieds est un défi qui fut résolu seulement dans les dernières années (Andriluka et al., 2010; Belagiannis & Zisserman, 2017; Hidalgo et al., 2019; Pavlakos et al., 2019; Simon, Joo, Matthews, & Sheikh, 2017).

Une approche directe est l'utilisation des paramètres, des dépendances et des contraintes du corps humain (Andriluka et al., 2010; Johnson & Everingham, 2010; Karlinsky & Ullman, 2012; Newell, Yang, & Deng, 2016). Dans ces articles, les auteurs utilisent les contraintes physiques et kinésiques du corps humain afin d'estimer la pose dans une image. Certaines recherches utilisent un modèle en arbre structuré qui met en relation les différents membres du corps et l'espace occupé par ceux-ci d'après l'image (Andriluka et al., 2010; Johnson & Everingham, 2010). D'autres recherches prennent ce modèle d'arbre et ajoutent du post traitement tel que la gestion de la symétrie et de l'occlusion des membres (Karlinsky & Ullman, 2012; Newell et al., 2016) en utilisant une technique de détection de contour (Al-Amri, Kalyankar, & Khamitkar, 2010). Finalement, les recherches les plus récentes tentent de résoudre la problématique en combinant les concepts ci-dessus mentionnés et l'intelligence artificielle afin d'améliorer la précision des algorithmes (Belagiannis & Zisserman, 2017; Chu et al., 2017; Newell et al., 2016; Tompson, Goroshin, Jain, LeCun, & Bregler, 2015). Plus précisément, différents types de réseau de neurones tel que les réseaux de neurones convolutifs sont utilisés. Quelques exemples sont discutés ci-dessous.

Deepcut (Pishchulin et al., 2016) est une proposition top-down qui utilise un système de réseau de neurones. Cette recherche permet de récupérer les marqueurs de corps. Cependant, les occlusions sont problématiques, ce qui augmente grandement l'erreur potentielle dans l'analyse de la pose. De plus, le modèle entraîné est basé sur une résolution très basse (Andriluka, Pishchulin, Gehler, & Schiele, 2014). Par conséquent, la précision de la détection de pose sur une images à grande résolution pourrait être problématique et imprécise.

OpenPose (Cao et al., 2018) est l'algorithme suggéré par SMPLify-X (Pavlakos et al., 2019) pour obtenir les marqueurs 2D de la pose permettant la reconstruction du modèle surfacique. OpenPose propose un algorithme qui fait l'estimation de tous les paramètres du corps, ce qui inclut les pieds, le corps, les mains ainsi que le visage. Les Figure 1.5, Figure 1.6, Figure 1.7 démontrent ces marqueurs.

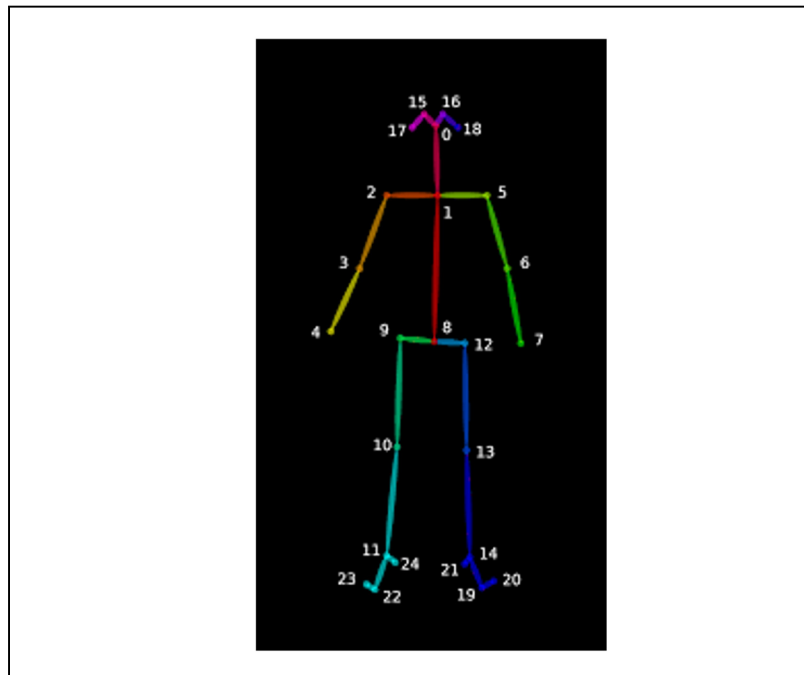


Figure 1.5 Marqueurs du corps et pied de OpenPose
Tirée de Cao (2018, p. 3)

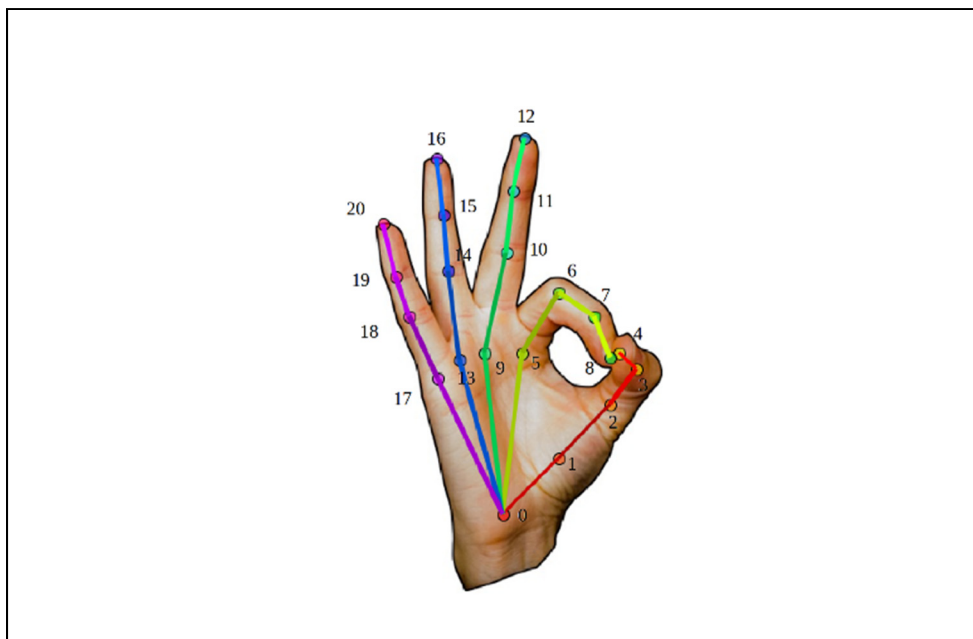


Figure 1.6 Marqueurs de main de OpenPose
Tiré de Simon (2017, p.4)

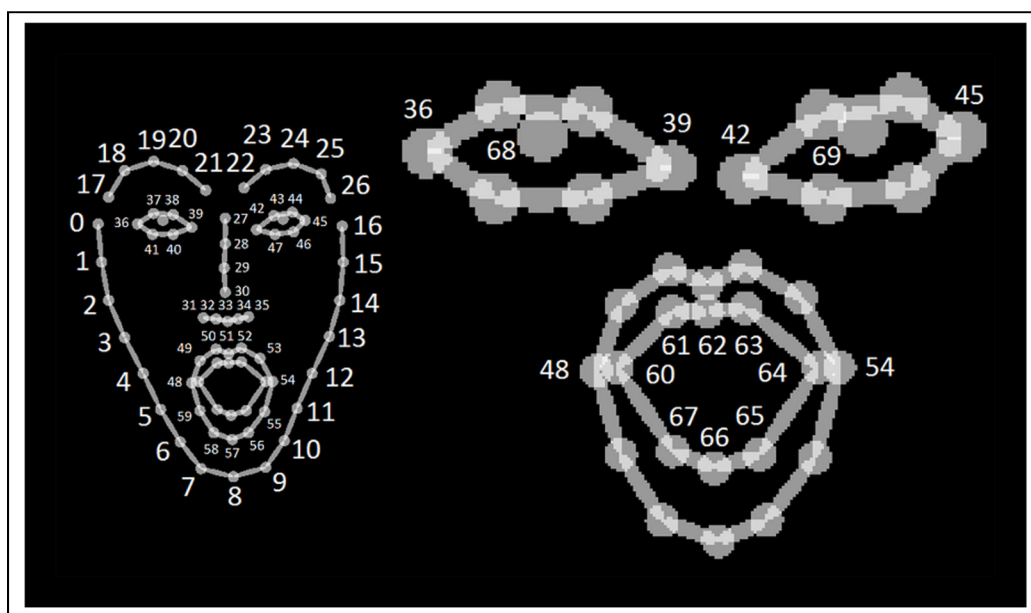


Figure 1.7 Marqueurs du visage de OpenPose
Tirée de Cao (2018, p. 3)

OpenPose combine plusieurs réseaux de neurones afin d'isoler les différentes parties du corps et d'identifier les marqueurs recherchés tels que les mains, les pieds, le corps, et le visage. Accessible et disponible à tous, les auteurs exposent leur code au public¹. De plus, il permet d'estimer la pose d'un être humain dans 1 image à 22 images par seconde lorsqu'un seul sujet est présent dans l'image.

Single-Network Whole-Body Pose Estimation (Hidalgo et al., 2019) offre une itération d'amélioration sur OpenPose. Il offre deux alternatives au modèle initial. La première a une capacité de traitement environ huit fois plus rapide tandis que la deuxième est légèrement plus lente, mais gagne en précision. L'algorithme réussit à obtenir des marqueurs plus précis en utilisant un seul algorithme neuronal afin de générer tous les marqueurs du corps incluant les mains et les pieds. Ceci augmente grandement la qualité des marqueurs puisque les informations des autres marqueurs peuvent être utilisées comme contraintes. Ils gèrent mieux les occlusions, les embrouillages, et les images à basses résolutions.

Finalement, il y a *Whole-Body Human Pose Estimation in the Wild* (Jin et al., 2020) qui propose un algorithme nommé ZoomNet qui est un réseau de neurones à convolution qui utilise l'architecture hiérarchique du corps humain afin de détecter les marqueurs du corps, du visage, et des mains. Celui-ci est validé en utilisant la même base de données que les autres algorithmes proposés par les auteurs de OpenPose (COCO-Whole Body)(Jin et al., 2020). En rétrospective, bien que la solution donne des résultats plus précis, l'algorithme et le modèle ne sont pas disponibles.

1.1.3 Récapitulatif sur l'obtention d'un avatar

Dans cette section, il a été démontré qu'il existe deux types de méthodologies afin d'obtenir un avatar normalisé. La première est la manipulation d'un avatar prégénéré et la deuxième est la génération complète de l'avatar normalisé. La manipulation d'un avatar requiert comme

¹ <https://github.com/CMU-Perceptual-Computing-Lab/openpose>

entrée des marqueurs de poses qui peut être obtenue de plusieurs façons. Cependant, OpenPose (Jin et al., 2020) est la recherche qui ressort du lot. Dans les deux cas, il est possible de répondre à la première partie de l'hypothèse qui est d'avoir un avatar normalisé. La génération complète comporte un avantage de précision, mais inclut les vêtements ce qui affecterait les mesures prises sur celui-ci une fois à l'échelle. La manipulation de l'avatar comporte tente de repositionner un avatar sans vêtements et accessoires ce qui peut rendre les mesures plus rapprochées des vraies mesures anthropométriques une fois mis à l'échelle.

1.2 Estimation de mesure à l'échelle à partir d'une image

Ayant maintenant analysé les différents systèmes qui permet d'obtenir un avatar ainsi que les estimateurs de poses qui sont nécessaires dans certain de ces systèmes, il nous faut comprendre comment obtenir une mesure à trois dimensions à partir d'une seule camera afin de mettre à l'échelle l'avatar. Cette section analyse d'abord les diverses méthodes d'estimation à trois dimensions. Il sera identifié que la méthode favorable pour cette recherche nécessitera la calibration du système et un outil d'estimation de marqueurs. Conséquemment, un survol de chacun est fait.

1.2.1 Estimation de mesures à trois dimensions

L'estimation de mesure à trois dimensions peut être subdivisées en trois grandes catégories : les méthodes actives, méthode d'utilisation d'information connue (passive) et les méthodes d'intelligence artificielle.

Les méthodes actives utilisent plusieurs dispositifs ou plusieurs entrées afin de générer une représentation numérique en 3D d'un sujet présent dans une scène. Ces systèmes sont basés sur la projection d'information visuelle connue (motif, illumination, IR, laser) et la récupération de la déformation produite dans celle-ci utilisant des systèmes calibrés (Banks, Read, Allison, & Watt, 2012; Kytö, Nuutinen, & Oittinen, 2011). Par exemple, la stéréoscopie (Banks et al., 2012; Kytö et al., 2011) consiste à l'utilisation de plusieurs caméras calibrées à différents angles. En créant une carte de disparité entre deux images au même instant, il est

possible d'estimer la position à trois dimensions des pixels. La lumière structurée est une méthodologie similaire, mais qui utilise un projecteur et une caméra afin de faire une comparaison entre l'image projetée et l'image capturée (Chou, Solli, & Jalali, 2008). Une autre méthodologie fréquemment utilisée est le temps de vol d'un laser et d'une caméra calibrée aussi nommé *LIDAR* (Xuanwu, Guijin, Chenbo, & Qingmin, 2014). Le laser dans l'image capturé est analysé afin de créer une carte de profondeur. En contrepartie, toutes ces recherches requièrent plusieurs équipements pas toujours disponibles à portée de main. Plus grand est le nombre d'équipements nécessaires, plus le cout de déployer le tout en production est élevé.

Les méthodes passives s'appuient sur l'utilisation de donnée connue dans la scène afin d'estimer des mesures (Guan, 2009; Koh & Kamiyama, 2019; Yang et al., 2014). Les dispositifs utilisés dans ces systèmes doivent être préalablement calibrés. Par exemple, (Guan, 2009) réussit à obtenir la grandeur d'un sujet en utilisant seulement une caméra. Pour ce faire, l'auteur utilise une combinaison des proportions générales du visage de l'être humain, la distance pupillaire, ainsi que les relations proportionnelles statistiques entre la grandeur du sujet et le visage. Bien que ces techniques fonctionnent bien lorsque le sujet est proportionné parmi la moyenne des humains en morphologie, ceux-ci rencontrent de grandes difficultés lorsque la morphologie n'est pas dans la moyenne. De plus, ces mesures sont seulement valides si le sujet est perpendiculaire à l'axe optique sans quoi il y a un impact sur la précision de la mesure. (Guan, 2009) rencontre ce problème lorsque le visage du sujet n'est pas perpendiculaire à l'axe optique de la caméra. Il est aussi important de noter qu'il existe une grande relation entre la précision de la mesure de référence et la précision de l'estimation.

D'autres systèmes tentent de résoudre le problème en utilisant l'intelligence artificielle et les réseaux de neurones afin de déterminer des mesures précises dans le monde à trois dimensions à partir d'une seule image (Bertoni, Kreiss, & Alahi, 2019; X. Chen et al., 2016; Haseeb, Guan, Ristic-Durrant, & Gräser, 2018; Tang & Li, 2020; Yu & Oh, 2021; Y. Zhang, Li, Zhao, & Yu, 2020; Zhu & Fang, 2019). Par exemple, DisNet (Haseeb et al., 2018) utilise un réseau de neurones de trois couches utilisant six caractéristiques afin d'estimer la distance d'un humain par rapport à une caméra et obtient une précision acceptable sur des sujets à plus de six mètres.

(Tang & Li, 2020; Y. Zhang et al., 2020; Zhu & Fang, 2019) utilise un CNN pour identifier et classifier un objet dans une image et ensuite une série de réseaux de neurones pour déduire la distance à laquelle il se trouve. (Yu & Oh, 2021) montre une approche similaire où ils font une estimation initiale approximative de la distance et fournissent ensuite l'objet à un prédicteur spécialisé sur une plage déterminée pour obtenir la distance. Les résultats de ces deux recherches montrent des résultats optimaux à une distance de 10m de la caméra et plus on s'éloigne de cette distance, plus la précision se détériore. (Bertoni et al., 2019; X. Chen et al., 2016) utilisent la détection de la pose du sujet, une caméra calibrée, et un réseau neuronal de neuf couches pour estimer la distance. Bien que cette recherche propose une méthode intéressante pour placer le sujet par rapport à la caméra, son résultat moyen dans les meilleurs scénarios est une erreur de localisation moyenne de 0,86m ce qui est au-dessus des autres recherches mentionnées précédemment.

En résumé, chacune des méthodes décrites ci-dessus ont leurs avantages et leurs inconvénients. La majorité calibre leur système réduisant ainsi les inconnus ce qui augmente la précision des résultats. La prochaine section décrit les différentes méthodes de calibration.

1.2.2 Calibration du système

Indépendant du type de systèmes utilisés, la calibration permet de diminuer les inconnus et obtenir des mesures à trois dimensions (Qi, Li, & Zhenzhong, 2010). La calibration d'un système comportant une caméra est faite en deux parties. Il y a la calibration des caractéristiques internes de la caméra (paramètre intrinsèque) telle que la distance focale de la caméra, le centre de la caméra aussi nommé le décalage du point principal et la distorsion (Simek, 2013). La deuxième partie est la calibration de la caméra dans la scène visible (paramètres extrinsèques). (Qi et al., 2010) résume clairement les différentes possibilités de calibration d'une seule caméra. Il existe quatre grandes catégories dans lesquelles nous pouvons caractériser entièrement les paramètres d'une caméra : le calibrage à l'aide d'un objet de référence, l'autocalibrage par vision active, la méthode du modèle planaire et l'intelligence artificielle.

1.2.2.1 Calibration par objet de référence

L'étalonnage à partir d'un objet de référence déterminé et précis utilise un objet de référence et des stratégies d'optimisation afin de déterminer l'étalonnage. L'objet consiste habituellement de deux à trois faces orthogonales (Banks et al., 2012; Nayar & Gupta, 2012; Tsai, 1987; Yin, Wang, Shi, & Liao, 2014). Les données d'entrée de cette méthode sont de nombreuses images à différentes positions tout en étant toujours en mesure de voir l'objet de référence 3D. Bien qu'il existe de nombreuses mises en œuvre de ce type de méthodologie, deux se distinguent : la méthode en deux étapes (Faugeras, 1993) et la méthode biplan (Qi et al., 2010). La méthode en deux étapes consiste à utiliser les équations linéaires du modèle de sténopé pour trouver les paramètres initiaux de la calibration en excluant la distorsion de la caméra. Elle utilise ensuite une optimisation statistique sur tous les paramètres, y compris la distorsion optique, afin de minimiser l'erreur. La méthode biplan s'appuie sur le fait que la source lumineuse est stable pendant la calibration et utilise la différence de luminosité entre les images. Elle détermine la ligne de visée des points connus sur un objet de référence. En utilisant le triangle formé par les points sur les images et l'objet de référence et en utilisant l'interpolation par spline cubique, il est possible d'estimer la calibration. Cette catégorie a généralement une bonne précision. Cependant, une grande dépendance de cette méthodologie est la précision de l'objet de référence et un objet précis peut être très coûteux pouvant aller jusqu'à plusieurs milliers de dollars. La précision de l'objet et sa spécification de mesure à un impact direct sur la précision de la calibration.

1.2.2.2 Calibration active

La méthode active est basée sur le contrôle du mouvement de la caméra au lieu de contrôler l'objet de référence. Cette technique utilise plusieurs images. Elle trouve des points qui peuvent être trouvés dans les images multiples et applique un algorithme d'optimisation statistique afin de trouver la matrice intrinsèque et ensuite la matrice extrinsèque. Par exemple, il existe des techniques n'utilisant aucune référence telle que la technique de point de fuite (Caprile & Torre, 1990; W. Chen & Jiang, 1991) qui analyse le point de fuite de l'image de la caméra afin de déterminer les directions orthogonales dans l'image. Une technique similaire est la

calibration d'une caméra capturant plusieurs images de la même scène en déplaçant seulement la caméra. Une correspondance entre des points caractéristiques dans les différentes images est faite et un modèle statistique est utilisé afin de trouver les paramètres intrinsèques. Le calcul est beaucoup plus lourd sur ce type d'algorithme puisque très peu de contraintes existent. De plus, il dépend fortement de la recherche de points dans la scène et de l'assurance que ces points peuvent être trouvés dans les autres images. De plus, ces points doivent être bien identifiés. Un déplacement ou l'absence de l'un de ces points injecterait une erreur dans le calibrage. Elle est largement utilisée lorsque l'environnement ne peut être contrôlé.

1.2.2.3 Calibration par modèle planaire

La méthode du modèle planaire est celle qui s'appuie sur un plan 2D sur lequel figure un motif connu et précis, tel qu'une matrice de points ou de carrés (Hartley & Zisserman, 2003; Z. Zhang, 2000, 2002). En utilisant l'homographie et la minimisation des erreurs, il est possible d'analyser le déplacement du motif entre les images et d'identifier les matrices de calibration. La précision des mesures du modèle et la couverture du plan à deux dimensions sont directement liées à la précision de la calibration (Sturm & Maybank, 1999). Ce calibrage est précis et ne nécessite pas d'équipement ou de manipulation intensive. Cependant, il nécessite plus de 2 images pour obtenir des résultats utilisables et ne peut être automatisé sans travail supplémentaire.

1.2.2.4 Calibration par intelligence artificielle

Il n'est pas surprenant que certains chercheurs aient également tenté de calibrer une caméra dans un système utilisant des réseaux neuronaux. Les diverses méthodes sont rapides, simples et obtiennent une bonne précision (Raza, Rehman, Lee, & Choi, 2019; Sturm & Maybank, 1999). Cependant, l'entraînement de ces réseaux est un processus très complexe. Il peut facilement tomber dans l'optimisation locale, être lent à apprendre, ou simplement ne pas converger en fonction des entrées. C'est pourquoi ils ne sont pas souvent utilisés (Raza et al., 2019; Sturm & Maybank, 1999).

1.2.2.5 Résumé des méthodes de calibration

Toutes ces méthodes permettent d'obtenir à la fois la calibration intrinsèque et extrinsèque d'une ou plusieurs caméras. La calibration par intelligence artificielle a été écartée pour sa complexité, le temps d'entraînement, et qu'il n'y existe pas une solution publique avec le code source disponible. La méthode active a aussi été écartée puisque nous avons seulement une image dans le système. La calibration des paramètres intrinsèques par objet de référence est très difficile et n'obtient pas la meilleure précision lorsqu'une seule caméra est disponible dans le système. La calibration par modèle planaire est la méthode qui semble le mieux répondre au besoin de cette recherche. Cependant, le modèle doit être présent dans la scène afin de calibrer les paramètres intrinsèques et extrinsèques et ceci ne peut être garanti. Il est seulement possible d'utiliser cette méthode afin de calibrer les paramètres intrinsèques préalablement puisque ceux-ci ne changent pas. La méthode par objet de référence doit donc être utilisée afin d'obtenir les paramètres extrinsèques.

Les articles (Deák, Kainz, Michalko, & Jakab, 2017; Guan, 2009) démontrent la possibilité d'utiliser la distance pupillaire comme objet de référence afin d'accomplir cette calibration. La distance pupillaire est calculée en mesurant la différence entre le centre des deux pupilles (Coon, 2016; Dodgson, 2004). Cependant, afin de pouvoir procéder à cette mesure, le sujet doit regarder à l'infini. Advenant le cas où celui-ci regarderait à un point convergeant plus proche que l'emplacement physique de la caméra, la mesure serait faussée puisque le centre de la pupille ne serait pas au centre de l'œil en soi. Un des grands avantages d'utiliser la distance pupillaire est que c'est une mesure qui a une très petite variation au sein des humains, qui est stable, et qui est facilement visible dans une image prise de front (Dodgson, 2004). La distance moyenne chez les femmes est de 61.7mm et chez les hommes 64mm. L'écart-type chez les femmes c'est de 3.6mm tandis que chez les hommes est seulement de 3.4mm. Il est donc possible d'utiliser la distance pupillaire moyenne si la vraie mesure n'est pas à portée de main. Puisque la mesure réelle de la distance pupillaire requiert d'être comparée à la distance pupillaire sur une image afin de calibrer le système, l'obtention de la distance pupillaire dans une image sera nécessaire.

1.2.3 Outil d'analyse du visage à partir d'une image

Le visage humain est un sujet qui intéresse grandement les chercheurs des dernières années (Martinez & Valstar, 2016; Sariyanidi, Gunes, & Cavallaro, 2015). Il est un vecteur de communication important (De la Torre et al., 2015; Ekman, Friesen, O'Sullivan, & Scherer, 1980). Dans le contexte de cette recherche, avoir précisément la distance pupillaire dans une image permet de calibrer le système tel que défini dans la section 1.2.2. Cette section se concentre sur l'analyse des différentes options d'outils qui permettent d'obtenir les marqueurs du visage ainsi que tout autre paramètre du visage qui pourrait être pertinent dans cette calibration.

Quatre paramètres principaux ont été identifiés par plusieurs recherches comme étant les paramètres caractérisant l'état d'un visage dans une image : les marqueurs de visage, la pose de la tête, l'expression du visage, et le point de regard des yeux (Albiol, Monzo, Martin, Sastre, & Albiol, 2008; Baltrušaitis, Robinson, & Morency, 2016; De la Torre et al., 2015).

Les marqueurs de visage définissent la forme et l'emplacement du visage et de ses composantes tel que le nez, les yeux, et la bouche. La quantité et la précision de ces marqueurs sont ce qui définit la qualité d'un algorithme. Un exemple de ces algorithmes d'estimation est OpenFace (Baltrušaitis et al., 2016) et ces marqueurs sont présentés dans la Figure 1.8.

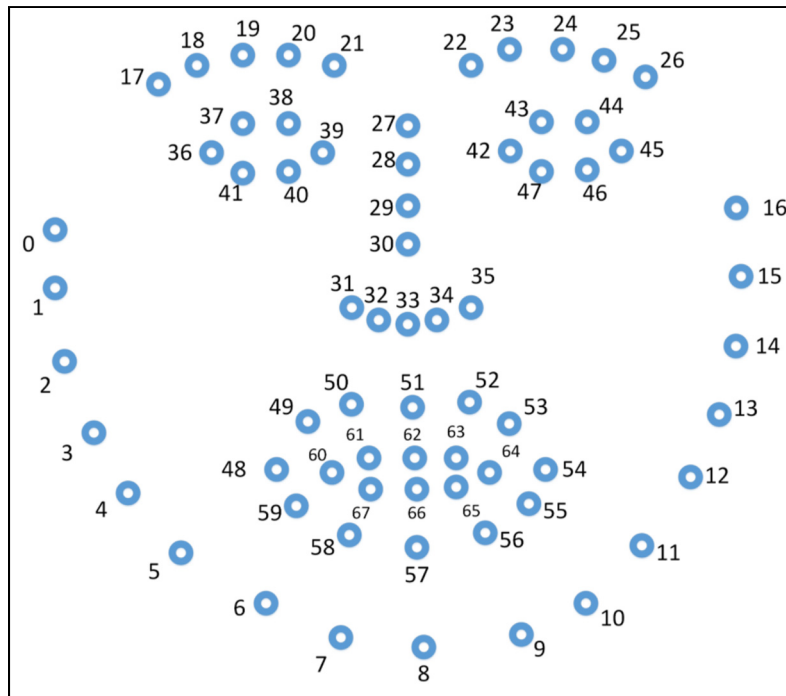


Figure 1.8 Marqueurs OpenFace
Tiré de Baltrusaitis, Zadeh, Lim, & Morency (2018, p.1)

La pose de la tête représente l'orientation à trois dimensions de la tête du sujet soit le tangage (« pitch »), le lacet (« yaw »), et le roulis (« roll ») (Arcoverde et al., 2014; Gudet, 2015). Le tangage représente la rotation autour de l'axe transversal reliant les deux oreilles. Le lacet est la rotation horizontale autour de l'axe vertical suivant le cou jusqu'au-dessus de la tête. Finalement, le roulis est la rotation autour de l'axe longitudinal traversant le nez du sujet.

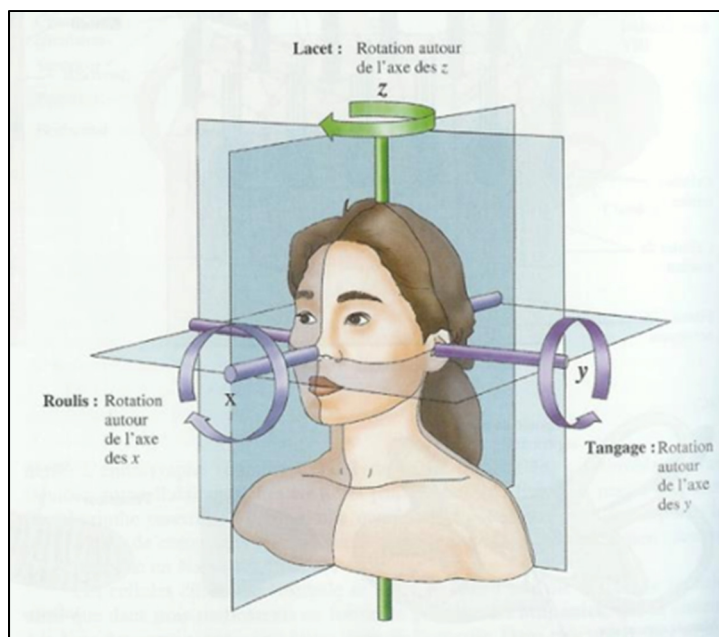


Figure 1.9 Caractérisation de l'orientation de la tête
Tirée de Gudet (2015, p.22)

Le point de regard ainsi que l'expression du visage sont tous deux des paramètres qui peuvent être utiles dans certaines circonstances lorsqu'on tente d'analyser ce que le sujet regarde où quelle émotion il ressent. Cependant, puisque cette recherche se concentre entièrement sur l'obtention de mesures, ils ne seront pas des facteurs analysés.

Il y a plusieurs boîtes à outil disponibles afin d'obtenir ces informations qui démarrent toutes de l'estimation des marqueurs du visage. Les recherches attaquent ce problème de différente façon. (Baltrušaitis et al., 2016; Baltrušaitis, Zadeh, Lim, & Morency, 2018) font une revue de ceux-ci, ce qui permet de faire une élimination rapide des diverses méthodes qui ne répondent pas aux requis, soit la position de la tête, les marqueurs du visage et le logiciel afin de pouvoir l'utiliser dans la recherche. Ceci est représenté dans la Figure 1.10 par (Baltrušaitis et al., 2018).

Tool	Approach	Landmark	Head pose	Expression	Gaze	Train	Test	Binary	Real-time	Free
COFW[13]	RCPR[13]	✓				✓	✓		✓	✓
FaceTracker	CLM[57]	✓	✓				✓	✓	✓	✓
dlib [37]	[35]	✓				✓	✓		✓	✓
Cehra	[5]	✓	✓					✓	✓	✓
Menpo [2]	AAM, CLM, SDM ¹	✓				✓	✓		2	✓
CFAN [77]	[77]	✓						✓	✓	✓
[73]	Reg. For [73]	✓	✓			✓	✓	✓	✓	✓
TCDCN	CNN [81]	✓	✓					✓	✓	✓
WebGazer.js	[54]				✓		✓	✓	✓	✓
EyeTab	[71]				✓	N/A	✓	✓	✓	✓
OKAO	unknown	✓	✓	✓	✓			✓		
Affdex	unknown	✓	✓	✓				✓	✓	
Tree DPM [85]	[85]	✓				✓	✓			✓
OpenPose [15]	Part affinity Fields [15]	✓	✓			✓	✓	✓	✓ ³	✓
CFSS [83]	CFSS [83]	✓				✓	✓			✓
iCCR [56]	iCCR [56]	✓						✓	✓	✓
LEAR	LEAR [46]	✓						✓	✓	✓
TAUD	TAUD [33]			✓				✓		✓
OpenFace	[8], [7]	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenFace 2.0	[70], [75], [78]	✓	✓	✓	✓	✓	✓	✓	✓	✓

*Les références sont de provenance de l'article original.

Figure 1.10 Technique de détection de marqueur du visage
Tiré de Baltrušaitis (2018, p.2)

Les mêmes auteurs démontrent une comparaison de la précision de l'estimation de la rotation de la tête en roulant diverses méthodologies sur la base de données ICT-3DHP (Baltrušaitis, Robinson, & Morency, 2012). Cette base de données contient 10 sujets avec plus de 14 000 images à différentes positions. Les résultats présentés dans la Figure 1.11 démontrent que OpenFace 2.0 est l'outil qui réussit à atteindre la meilleure précision sur la rotation de la tête.

Method	Yaw	Pitch	Roll	Mean
Reg. forests [25]	7.2	9.4	7.5	8.0
CLM-Z [10]	5.1	3.9	4.6	4.6
CLM [57]	4.8	4.2	4.5	4.5
Cehra [5]	13.9	14.7	10.3	13.0
OpenFace	3.6	3.6	3.6	3.6
OpenFace 2.0	3.1	3.5	3.1	3.2

*Les références sont de provenance de l'article original.

Figure 1.11 Précision de l'estimation de la rotation de la tête
Tiré de Baltrušaitis (2018, p.5)

1.2.4 Récapitulatif de l'obtention de mesure à l'échelle

Dans cette section il a été démontré qu'il est possible d'obtenir des mesures à trois dimensions à partir d'une seule caméra. Cependant, une calibration est nécessaire. Pour ce faire, considérant nos contraintes, la calibration à modèle planaire peut être utilisée afin d'obtenir les paramètres de la caméra. La calibration par objet de référence peut être utilisée en utilisant la distance pupillaire de l'humain afin de placer la caméra dans la scène (calibration extrinsèque). Cependant, il nous faut aussi la distance pupillaire dans l'image. Il a été identifié que l'outil le plus efficace pour l'analyse du visage est OpenFace 2.0. Ceci démontre que la deuxième partie de l'hypothèse est possible.

1.3 Récapitulatif de la revue de littérature

L'hypothèse initiale est qu'il est possible de combiner une technique d'obtention d'avatar avec une technique d'obtention de mesure à trois dimensions afin d'obtenir un avatar à l'échelle. Pour ce faire, une analyse des deux différents sujets a été faite.

Il a été démontré qu'il était possible d'obtenir un avatar à partir d'une seule image. Bien qu'une génération complète soit plus précise que le remplacement d'un modèle préexistant, le modèle préexistant contient des points connus qui peuvent être utilisés afin d'obtenir des mesures. Celui-ci permet de faire abstraction des accessoires (vêtement, souliers, etc.) sur un sujet en plus de pouvoir le déplacer dans d'autre position si besoin grâce à son squelette. C'est pour cela que ce type de modèle est utilisé dans notre recherche. Puisque SMPL-X (Pavlakos et al., 2019) est le modèle qui contient le plus de détails anthropométriques (visage, corps, pieds, mains) et qu'il est la référence au moment d'écriture de ce mémoire, c'est celui-ci qui sera utilisé et tenté d'être mis à l'échelle.

La revue de littérature a aussi démontré qu'il était possible d'obtenir des mesures à trois dimensions d'un sujet dans un système comprenant seulement une caméra. Plus précisément, en pré-calibrant les paramètres intrinsèques de la caméra et en utilisant la distance pupillaire du sujet, des mesures à trois dimensions peuvent être extraites d'une seule image. Cependant,

les recherches existantes démontrent que ce type de système a une faiblesse à la rotation du visage du sujet. Cette recherche devra donc tenter de trouver une solution à ce problème aussi.

CHAPITRE 2

MÉTHODOLOGIE

Tel que le démontre la revue de la littérature, diverses méthodes peuvent être établies afin de répondre à la problématique d'obtention de mesure anthropométrique d'un humain. Les recherches démontrent la possibilité d'obtenir des mesures à l'échelle et un avatar à partir d'une seule image et d'une caméra calibrée. Cette recherche se concentre donc sur la connexion et l'amélioration des méthodologies existantes afin d'obtenir une représentation numérique en 3D d'un sujet humain sur laquelle des mesures anthropométriques peuvent être prises. Ce problème peut être décortiqué en trois sous-objectifs :

- Ajuster un avatar à une image
- Ajuster l'avatar à l'échelle 3D en utilisant une mesure dans l'image
- Trouver la mesure réelle utilisée dans l'image

Le premier objectif étant d'obtenir un avatar normalisé à partir d'une image est répondu en utilisant la méthode d'ajustement d'avatar normalisée de SMPLify-X (Pavlakos et al., 2019) puisque celle-ci est la référence présentement en modèle à trois dimensions ayant un squelette.

Le deuxième objectif est en fait d'accomplir la calibration du système et ensuite mettre l'avatar à l'échelle en utilisant cette calibration. Tel que décrit dans la section 1.2.2, il est possible d'utiliser une calibration par objet de référence si nous avons préalablement fait la calibration intrinsèque. Plus précisément, la distance pupillaire peut être utilisée comme mesure de référence afin d'accomplir cette calibration. Cependant, la corrélation entre la distance pupillaire réelle et celle mesurée dans une image comporte des erreurs dues à la rotation de la tête telle que démontre (Deák et al., 2017; Guan, 2009). Afin de répondre à la deuxième partie du problème, nous tentons donc de corriger cette erreur en prenant en compte la rotation de la tête avant d'appliquer l'échelle à l'avatar normalisé. Un résumé de la méthodologie proposée pour les deux premiers objectifs est présenté dans la Figure 2.1.

Enfin, tel que décrit dans la section 1.2.2, la précision de la mesure de référence a un impact direct sur la calibration et donc la précision des résultats obtenus grâce à cette calibration. Conséquemment, afin de minimiser cette erreur en plus de fournir une méthodologie dans un cas où la distance pupillaire du sujet n'est pas disponible, nous proposons une méthode afin de l'obtenir.

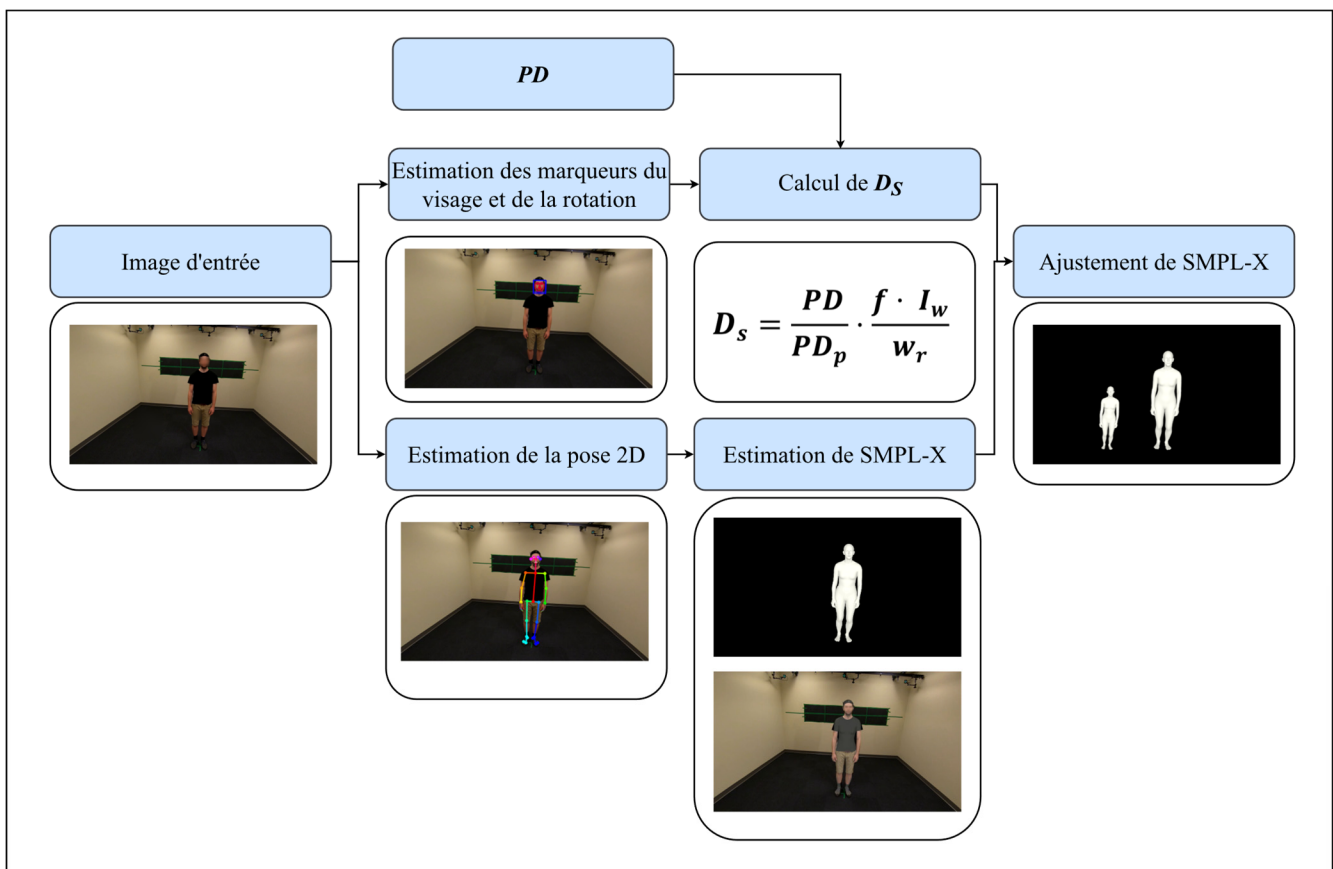


Figure 2.1 Survol de la méthodologie proposée

1.1 Ajustement d'un avatar à une image

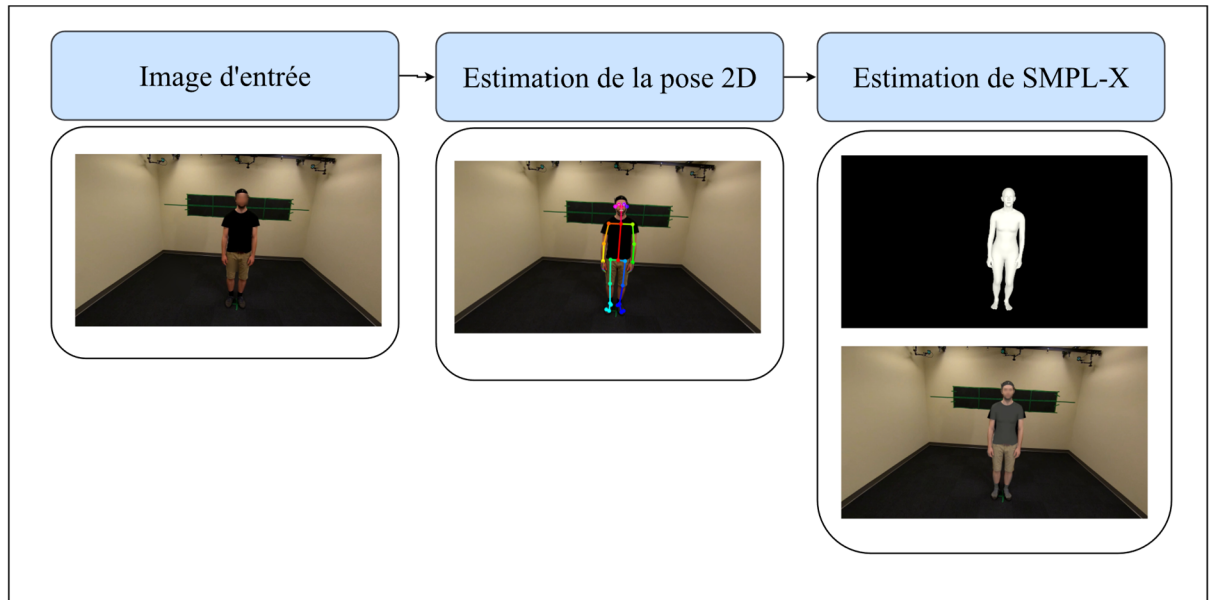


Figure 2.2 Méthodologie d'obtention d'un avatar à partir d'une image

Sous les contraintes du problème, le système à un sujet S et une caméra C_1 . Si $I = \{I(n), n \in \Lambda \subset \mathbb{N}^2\}$ est une image contenant le sujet dont le visage et le corps sont visibles et $I_s = \{I_s(x), x \in C_1 \subset \mathbb{R}\}$ est l'image I sur le capteur de la caméra. Rappelons aussi que l'objectif final est d'obtenir un avatar à l'échelle afin de pouvoir prendre des mesures sur celui-ci. Identifions cet avatar à l'échelle comme étant A_E . A_E est donc l'ensemble de sommets et de joints de l'avatar à l'échelle. Cependant, tel qu'identifié dans la revue de littérature, il n'y a aucune méthodologie qui permet d'obtenir A_E directement. L'hypothèse de cette recherche est donc qu'il est possible de passer par un avatar obtenu à partir d'une image (A_n) et de le mettre à l'échelle en utilisant un facteur de correspondance (α) tel que démontré dans l'équation (2).

$$A_E = A_n \cdot \alpha \quad (2)$$

A_E : Avatar à l'échelle

A_n : Avatar normalisé

α : Facteur d'échelle

Il faut identifier une méthode qui nous permette d'obtenir A_n . Puisque le but est d'ajuster et mettre à l'échelle cet avatar afin de prendre des mesures, il serait préférable d'avoir des points connus. Observant les différentes revues de littérature faites, il est possible d'identifier que le modèle SMPL-X (Loper et al., 2015) est celui qui répond le mieux aux besoins. Il fournit non seulement des résultats précis et morphologiquement représentatif, mais il contient un squelette qui peut être utilisé à des fins de mesure. Cependant, cette méthodologie requiert de trouver les paramètres nécessaires afin de reproduire la morphologie et la pose du sujet. Pour simplifier le processus, SMPLify-X (Pavlakos et al., 2019) a été sélectionné afin de générer la morphologie. SMPLify-X a aussi besoin des marqueurs de positionnement et OpenPose 2.0 (Cao et al., 2018) a été choisi afin de les obtenir puisque les résultats des auteurs sont les plus précis d'après (Cao et al., 2018; Pavlakos et al., 2019). Plus la précision de la pose sera grande, mieux sera le recalage du modèle sur le sujet. Cette étape de la méthodologie qui permet d'obtenir A_n peut être observé dans la Figure 2.2

2.1 Ajustement d'un avatar à une distance

Ayant maintenant A_n , il nous faut trouver α dans l'équation (2). Laissons $S \subset I$ être un ensemble de pixels de l'image I définissant le sujet lorsqu'il est debout devant la caméra et $D_S \in \mathbb{R}$ être la distance du centre optique de la caméra C_1 jusqu'au point central séparant les centres pupillaires de chaque œil sur la surface à trois dimensions S_{3D} représentant le sujet. Il est alors possible de définir α comme étant la distance à laquelle le sujet est (D_S) sur la distance à laquelle le sujet devrait être afin d'obtenir l'avatar A_n . Définissons A_n comme étant une un ensemble de point à trois dimensions $[x,y,z]$. Si nous prenons le plan lorsque $z=0$ alors il existe au moins une image où le sujet peut être positionné afin de recréer ce plan sur la caméra. Puisque le modèle est normalisé et que chaque unité sur le plan représente une unité du monde réel, la seule distance à laquelle ceci est possible est la distance focale de la caméra ($f \in \mathbb{R}$). Il est donc possible de définir l'avatar à l'échelle tel que l'équation (3) le démontre.

$$\alpha = \frac{D_s}{f}$$

$$A_E = A_n \cdot \frac{D_s}{f} \quad (3)$$

A_E : Avatar à l'échelle

A_n : Avatar normalisé

D_s : Distance du sujet à la caméra

f : Distance focale

Assumons que l'axe vertical de la tête est parallèle à la normale du sol et que le regard du sujet est principalement orienté vers la caméra C_1 sans être nécessairement parallèle à l'axe optique de la caméra C_1 . Assumons aussi que la caméra C_1 est calibrée et n'a pas de distorsion radiale. La distance focale calibrée f de C_1 est et il nous faut trouver D_s . La méthode proposée est démontrée dans la Figure 2.3 et détaillée ci-dessous tel que le système dans la Figure 2.4.

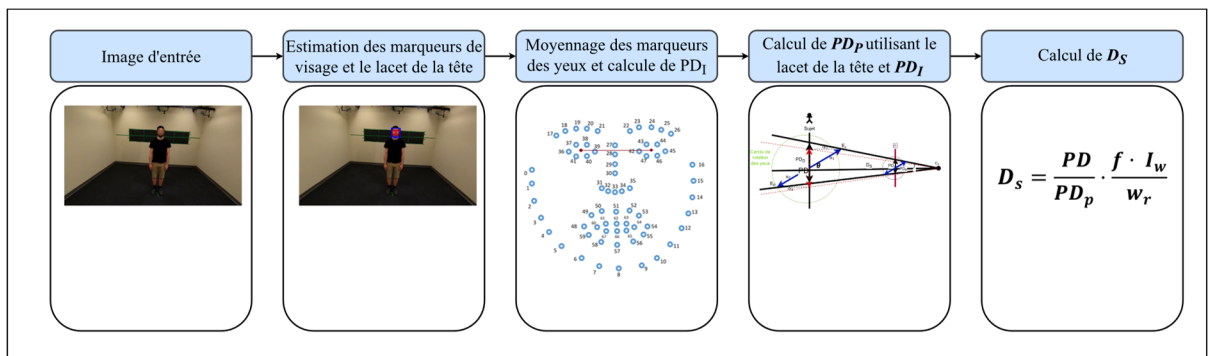


Figure 2.3 Calcul de D_s

Il est possible d'utiliser le modèle sténopé afin de trouver la distance à laquelle le sujet est situé. Dans un système contenant une caméra et un objet, il est possible d'estimer la grosseur ou la distance de l'objet en utilisant une mesure connue qui est présente dans l'image telle que définie dans la section 1.2.2 (Faugeras, 1993; Gremban, Thorpe, & Kanade, 1988). Assumons que le sujet est présent dans l'image I et qu'il regarde à l'infini. La distance pupillaire peut alors être utilisé comme mesure où $PD = \|E_l - E_r\|$, avec $E_l, E_r \in S_{3D}$ étant le centre

pupillaire gauche et droite respectivement de la surface 3D tel que prescrit par les recherches identifiées dans la section 1.2.2. Cette assumption est nécessaire puisque la distance pupillaire est calculée à partir du centre de l'iris (Dodgson, 2004). Advenant le cas où le sujet regarderait à un point convergeant entre la caméra et le sujet, les iris ne seraient pas au centre de l'œil ce qui changerait la mesure de la distance pupillaire. Définissons le système tel que le démontre la Figure 2.4 dans lequel le cercle de gauche représente la rotation des yeux du sujet et que $PD = a_1 + a_2$.

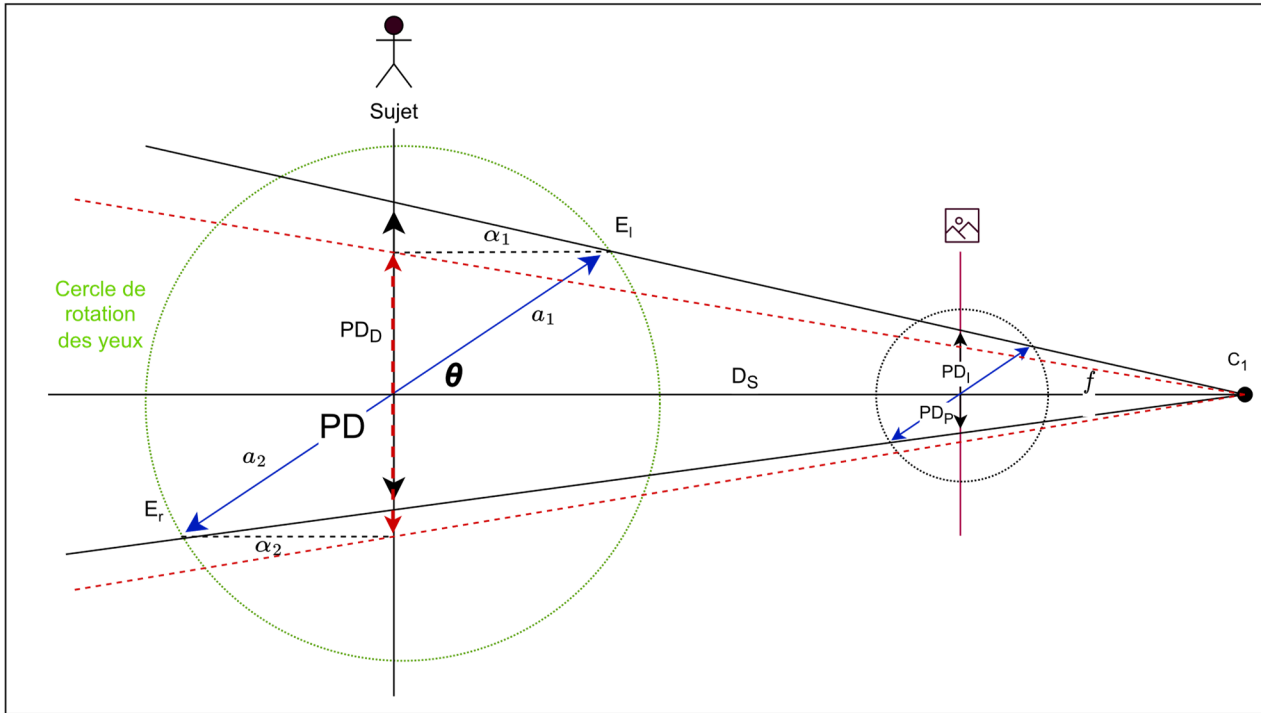


Figure 2.4 Vue de dessus du système

Définissons $PD_I = \|E_l^P - E_r^P\| \in I$ comme étant la distance pupillaire en pixel mesuré sur l'image I sans distorsion et $PD_s = \|E_l^S - E_r^S\| \in I_s$ comme étant la projection de PD sur le capteur de C_1 . Si nous assumons que $\alpha_1 = \alpha_2$ (visage symétrique), alors la projection de PD sur le plan parallèle au plan d'image (PD_D) et passant par le point central entre les deux yeux peut être défini par $PD = PD_D \cdot \cos(\theta)$, θ étant l'orientation horizontale de la tête (le lacet). Ceci est démontré ci-dessous (équation (4)).

Dans le monde **3D**:

$$E_l = (X_1, Y_1, Z_1) \text{ \& } E_r = (X_2, Y_2, Z_2)$$

Dans le plan de l'image **I**:

$$E_l^p = \left(f \frac{X_1}{Z_1}, f \frac{Y_1}{Z_1}\right) \text{ \& } E_r^p = \left(f \frac{X_2}{Z_2}, f \frac{Y_2}{Z_2}\right)$$

$$PD = f \sqrt{\left(\frac{X_1}{Z_1} - \frac{X_2}{Z_2}\right)^2 + \left(\frac{Y_1}{Z_1} - \frac{Y_2}{Z_2}\right)^2}$$

Si **XY** représente un point dans une image:

$$PD = f \left(\frac{XY_1}{Z_1} - \frac{XY_2}{Z_2}\right)$$

$$XY_1 = \frac{PD_D}{2} \cos(\theta) \text{ \& } XY_2 = \frac{PD_D}{2} \cos(\theta)$$

$$Z_1 = D_s - \frac{PD_D}{2} \sin(\theta) \text{ \& } Z_2 = D_s + \frac{PD_D}{2} \sin(\theta)$$

$$PD = f \frac{PD_D}{2} \cos(\theta) \left(\frac{1}{D_s - \frac{PD_D}{2} \sin(\theta)} + \frac{1}{D_s + \frac{PD_D}{2} \sin(\theta)} \right)$$

Si $D_s \gg \frac{PD_D}{2} \sin(\theta)$:

$$PD = f \frac{PD_D}{2} \cos(\theta) \left(\frac{1}{D_s} + \frac{1}{D_s} \right)$$

$$PD = f \frac{PD_D}{D_s} \cos(\theta) \text{ \& } PD = PD_D \cdot \cos(\theta) \quad (4)$$

Si f est la focale de la caméra C_I alors l'équation de sténopé peut être définie tel que démontré dans l'équation (5).

$$\frac{PD_s}{f} = \frac{PD_D}{D_s}$$

$$D_s = \frac{PD \cdot \cos(\theta) \cdot f}{PD_s} \quad (5)$$

θ : Angle de rotation du visage du sujet

PD_s : Distance pupillaire sur le capteur de la caméra

Si $S_r = (w_r, h_r) \in \mathbb{R}^2$ est la longueur et hauteur du capteur de la caméra C_I et assumons que le capteur a des pixels carrés. De plus, si $I = (I_w, I_h) \in \mathbb{N}^2$ est la largeur et la hauteur de l'image I . La mesure pupillaire du sujet S sur le capteur de la caméra C_I (PD_s) peut être calculée en utilisant une règle de trois tel que démontré dans l'équation (6).

$$PD_s = \frac{w_r \cdot PD_I}{I_w} \quad (6)$$

PD_s : Distance pupillaire sur le capteur de la caméra

w_r : longueur du capteur de la caméra

PD_I : Distance pupillaire sur l'image

I_w : Longueur de l'image en pixel

Lorsque nous combinons l'équation (5) et l'équation (6) nous obtenons l'équation (7) qui s'applique pour n'importe quel système ayant une seule caméra tant qu'un sujet s'y retrouve avec les assumptions faites ci-dessus.

$$D_s = \frac{PD \cdot \cos(\theta)}{PD_I} \cdot \frac{f \cdot I_w}{w_r} \quad (7)$$

Définissons \mathbf{PD}_p comme étant les inconnus de l'équation (7) telle que le démontre l'équation (8).

$$\mathbf{PD}_p = \frac{\mathbf{PD}_I}{\cos(\theta)} \quad (8)$$

Il est donc possible de simplifier l'équation (7) à l'équation (9).

$$\mathbf{D}_s = \frac{\mathbf{PD}}{\mathbf{PD}_p} \cdot \frac{f \cdot I_w}{w_r} \quad (9)$$

\mathbf{PD}_p : La distance pupillaire dans l'image divisée par le cosinus de la rotation du visage

Ainsi, \mathbf{A}_E est trouvé en combinant les équations (3) et (9) tel que démontré dans l'équation (10).

$$\mathbf{A}_E = \mathbf{A}_N \cdot \frac{\mathbf{PD}}{\mathbf{PD}_p} \cdot \frac{f \cdot I_w}{w_r} \cdot f \quad (10)$$

Tel que d'écrits dans l'équation (9), il faut trouver \mathbf{PD}_p . Pour ce faire, il faut obtenir \mathbf{PD}_I et l'angle de rotation de la tête du sujet à partir d'une image sans distorsion. L'outil OpenFace (Baltrušaitis et al., 2016) a été utilisé afin d'accomplir la tâche. Bien que OpenPose aurait pu être utilisé puisqu'il est déjà utilisé pour l'estimation de la pose pour la première partie du problème, OpenFace offre une plus grande précision sur les marqueurs du visage (Baltrušaitis et al., 2016) et est spécialisé sur le visage. De plus, contrairement à OpenPose, OpenFace fournit l'information de l'orientation de la tête qui est nécessaire afin de calculer \mathbf{PD}_p . L'estimation de l'emplacement de \mathbf{E}_r^P et \mathbf{E}_l^P est défini comme étant la moyenne de tous les marqueurs entourant l'œil respectif. Ainsi, \mathbf{PD}_I peut être calculé en calculant la distance entre les deux points 2D comme le démontre l'équation (11) si $\mathbf{E}_r^P = (x_1, y_1) \in \mathbb{N}^2$ et $\mathbf{E}_l^P = (x_2, y_2) \in \mathbb{N}^2$.

$$PD_I = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (11)$$

Utilisant une image sans distorsion et le lacet de la tête, il est possible de calculer PD_p avec l'équation (8). Une fois obtenu et connaissant la distance pupillaire de PD , il est possible d'utiliser l'équation (9) afin de trouver D_s . Advenant le cas ou PD n'est pas connu et vu la mince variation entre les humains sur la distance pupillaire, il est aussi possible d'utiliser la moyenne de chacun des genres.

2.2 Estimation de la distance pupillaire en utilisant deux images

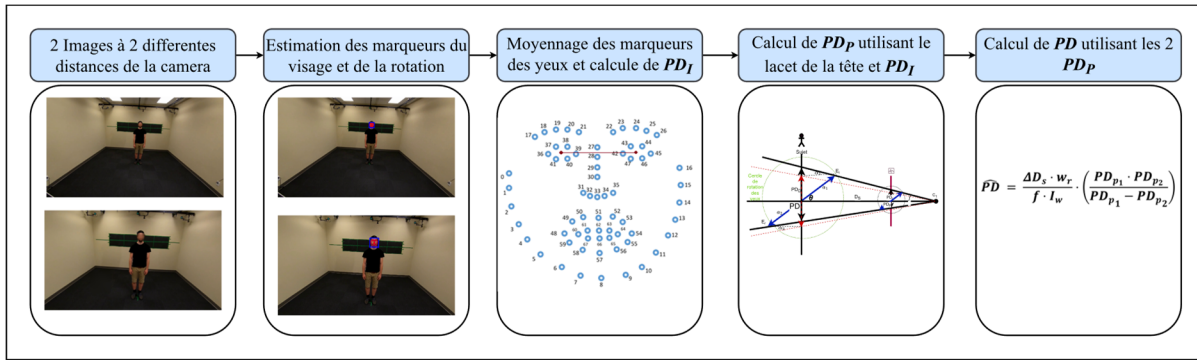


Figure 2.5 Méthodologie d'estimation de PD

Il a été démontré dans la section 1.2.2 que la précision de l'objet de référence a un impact immédiat sur la précision de la calibration et donc la mesure finale prise en utilisant cette calibration. Dans notre cas, la référence est la distance inter pupillaire et la mesure désirée est D_s . Bien que la stabilité de la distance pupillaire permette d'utiliser la moyenne basée sur le sexe de l'individu, avoir la mesure précise augmenterait la précision de la mesure de la distance. Nous proposons donc une méthodologie afin d'obtenir cette mesure comme démontré dans la Figure 2.5. Cependant, puisque ceci n'était pas le problème initial, la contrainte d'une seule image dans le système a été changé à deux images.

Advenant qu'il existe deux images à deux distances connues $I = \{I_i(\mathbf{n}), \mathbf{n} \in \Lambda \subset \mathbb{N}^2\}, i \in [1,2]$, l'équation (9) peut être utilisé afin de représenter chacune des images et des distances

auquel elles ont été prises (D_{s1} , D_{s2}), et des distances pupillaires sur les images (PD_{p1} , PD_{p2}). Tel que démontré dans l'annexe I, il est possible de mettre en relation les deux équations afin d'obtenir l'équation (12).

$$\widehat{PD} = \frac{\Delta D_s \cdot w_r}{f \cdot I_w} \cdot \left(\frac{PD_{p1} \cdot PD_{p2}}{PD_{p1} - PD_{p2}} \right) \quad (12)$$

\widehat{PD} : Estimation de la distance pupillaire

ΔD_s : Différence de distance du sujet à la caméra entre les deux images

w_r : La largeur du capteur de la caméra

f : Distance focale de la camera

I_w : La largeur de l'image

PD_{p1} : La distance pupillaire dans l'image 1 divisée par le cosinus de la rotation du visage à l'image 1

PD_{p2} : La distance pupillaire dans l'image 2 divisée par le cosinus de la rotation du visage à l'image 2

Cette équation démontre qu'il est possible d'utiliser deux images et la différence de distance entre ces deux images afin d'obtenir la distance pupillaire d'un individu.

En pratique, il est possible d'utiliser le même outil que la section 2.1, OpenFace 2.0 (Baltrusaitis et al., 2018). Cet outil nous permet d'obtenir E_r^P et E_l^P de chacune des images ainsi que θ . Ainsi, il est possible de calculer PD_{p1} et PD_{p2} . Le dernier inconnu est ΔD_s .

Bien que ΔD_s peut être facile à obtenir de façon statique tel que marquer deux mesures sur le sol et calculer la différence, il serait bien plus efficace de l'obtenir à l'aide d'algorithme automatique. Deux outils disponibles dans la majorité des appareils mobiles intelligents de 2020 sont le gyroscope et l'accéléromètre. Tel que décrit dans (Kowalczyk & Merta, 2015; Muset & Emerich, 2012; Pang & Liu, 2001; Seifert & Camacho, 2007), il est possible de

déterminer la distance parcourue en utilisant les données fournies par ces appareils. Pour ce faire, les articles démontrent qu'il faut combiner la double intégration de l'accélération ainsi que l'orientation du gyroscope. En contrepartie, les articles utilisent aussi plusieurs filtres ou réseau de neurones afin de stabiliser les signaux puisque les mesures sont généralement très bruitées ce qui crée une erreur exponentielle due à la double intégrale. Par conséquent, seule la méthode statique sera utilisée dans ce mémoire. Ainsi, il est possible d'utiliser soit la vraie distance pupillaire, la distance pupillaire moyenne, ou bien la distance pupillaire estimée afin de calculer le facteur d'échelle.

CHAPITRE 3

EXPÉRIMENTATION

Afin de valider les méthodologies proposées, il nous faut une base de données avec les critères suivants :

- Des images dans lesquelles le sujet est complètement visible prisent à différentes rotations de la tête et à différentes distances de la caméra
- Les mesures anthropométriques de ce sujet
- La distance entre la caméra et le sujet
- La distance inter pupillaire du sujet
- Le lacet de la tête par rapport à l'axe optique de la caméra

Au moment de l'écriture de cette mémoire, aucune base de données qui respecte ces critères n'a pas été trouvée. Par conséquent, afin de tester l'équation (9) et l'équation (12), des expériences ont été complétées ce qui a permis de générer une base de données. Celle-ci comporte 10 sujets ayant différentes morphotypes et caractéristiques physiques. La distance pupillaire, le lacet de la tête ainsi que la distance du sujet à la caméra ont été mesurés. Tous les sujets sont des adultes.

3.1 Contexte

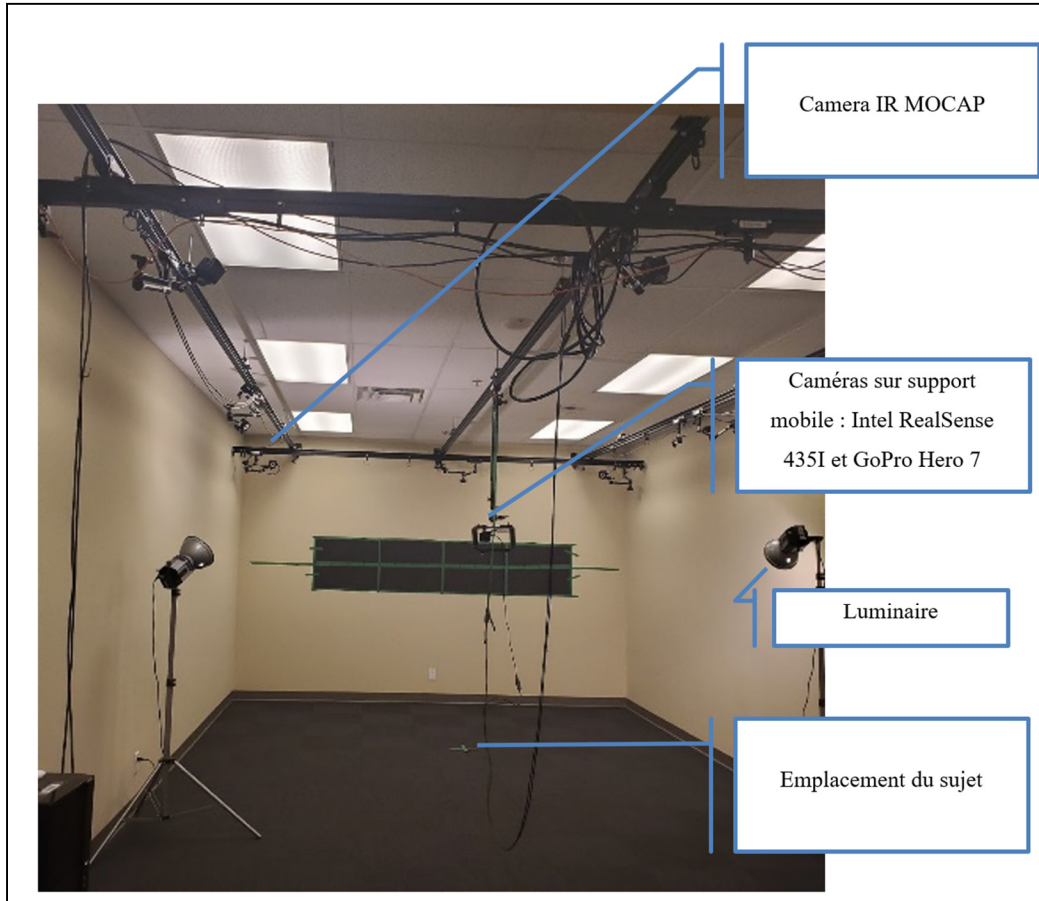


Figure 3.1 Environnement de capture de données

Un système de capture à trois dimensions de haute précision dénommée Optitrack a été utilisé afin d'obtenir les mesures théoriques de chacune des étapes (Furtado, Liu, Lai, Lacheray, & Desouza-Coelho, 2019). Une caméra infrarouge a été installée à chaque coin des rails ainsi qu'au milieu de chacun des rails pour un total de huit caméras infrarouges. Un rail a été installé dans le centre d'un rectangle de six mètres par quatre mètres. La position initiale du sujet était à deux mètres depuis le centre du côté de quatre mètres de l'OptiTrack. Un support à caméra amovible a été installé sur le rail central. Sur ce support, une caméra GoPro Hero 7 ainsi qu'une caméra Intel RealSense D435I ont été installées. Les deux caméras ont été calibrées en utilisant la calibration de cadres (Scaramuzza, Martinelli, & Siegwart, 2006; Sturm & Maybank, 1999).

La caméra Intel permet non seulement de capturer une image en haute définition (1980×1080), mais aussi les données d'un accéléromètre et d'un gyroscope intégrés. Le support fut ajusté afin que les caméras soient à la hauteur des yeux du sujet. De plus, l'angle des caméras fut ajusté afin de s'assurer d'avoir la totalité du sujet dans l'image initiale. Deux luminaires ont été ajoutées de chaque côté de la scène afin de s'assurer une bonne illumination et d'avoir l'image la plus nette que possible pour les algorithmes lors de l'analyse de l'image. Le sujet devait enfiler un casque ayant des marqueurs optiques pouvant être détectés par le système. Il y avait un marqueur placé de chaque côté de la tête au-dessus de la tempe et un jeu de 5 marqueurs sur le front du sujet.

3.2 Calibration

La calibration des caméras a été accomplie en utilisant un modèle à deux dimensions connue. Plus précisément, un schéma en damier a été utilisé. La Figure 3.2 est un exemple de ce type de schéma. Chaque carré mesurait 2,5 cm et couvrait une feuille entière de format lettre. 35 images ont été prises à différentes positions, en essayant de couvrir tout le champ de vision de la caméra et de couvrir différentes distances de la caméra. Enfin, la boîte à outils de vision par ordinateur de Matlab (Corke, 2005) a été utilisée pour calculer les paramètres intrinsèques des caméras.

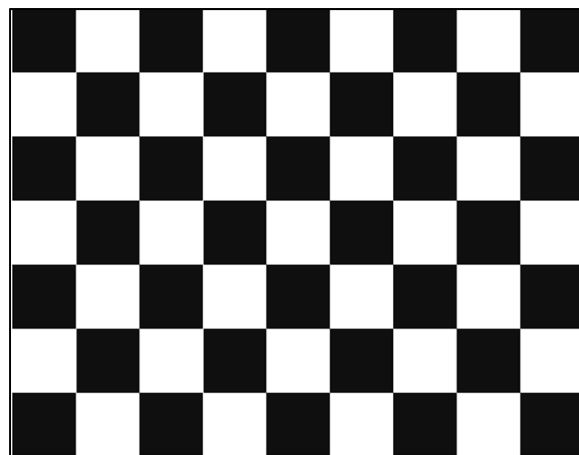


Figure 3.2 Schéma de damier

3.3 Protocole d'acquisition des données

La première étape préalable à l'expérimentation était de mesurer le sujet et les différentes mesures anthropométriques définies dans l'annexe III. Les mesures sont aussi représentées par les mesures ayant un numéro dans la Figure 3.3 et Figure 3.4 en plus de la grandeur du sujet. Ces mesures ont été choisis en suivant le guide défini dans (Norton & Olds, 1996) puisque ce sont les mesures anthropométriques les plus utilisées. La méthodologie utilisée afin de prendre les mesures suit aussi (Norton & Olds, 1996) en utilisant un rouleau à mesurer de couture. Ces mesures agissent comme étant les mesures théoriques afin de valider A_E . Par la suite, l'objectif était de mesurer la distance pupillaire à trois répétitions en utilisant la méthode suggérée par les optométristes (Dodgson, 2004). La médiane est la mesure retenue comme étant la mesure théorique de PD .

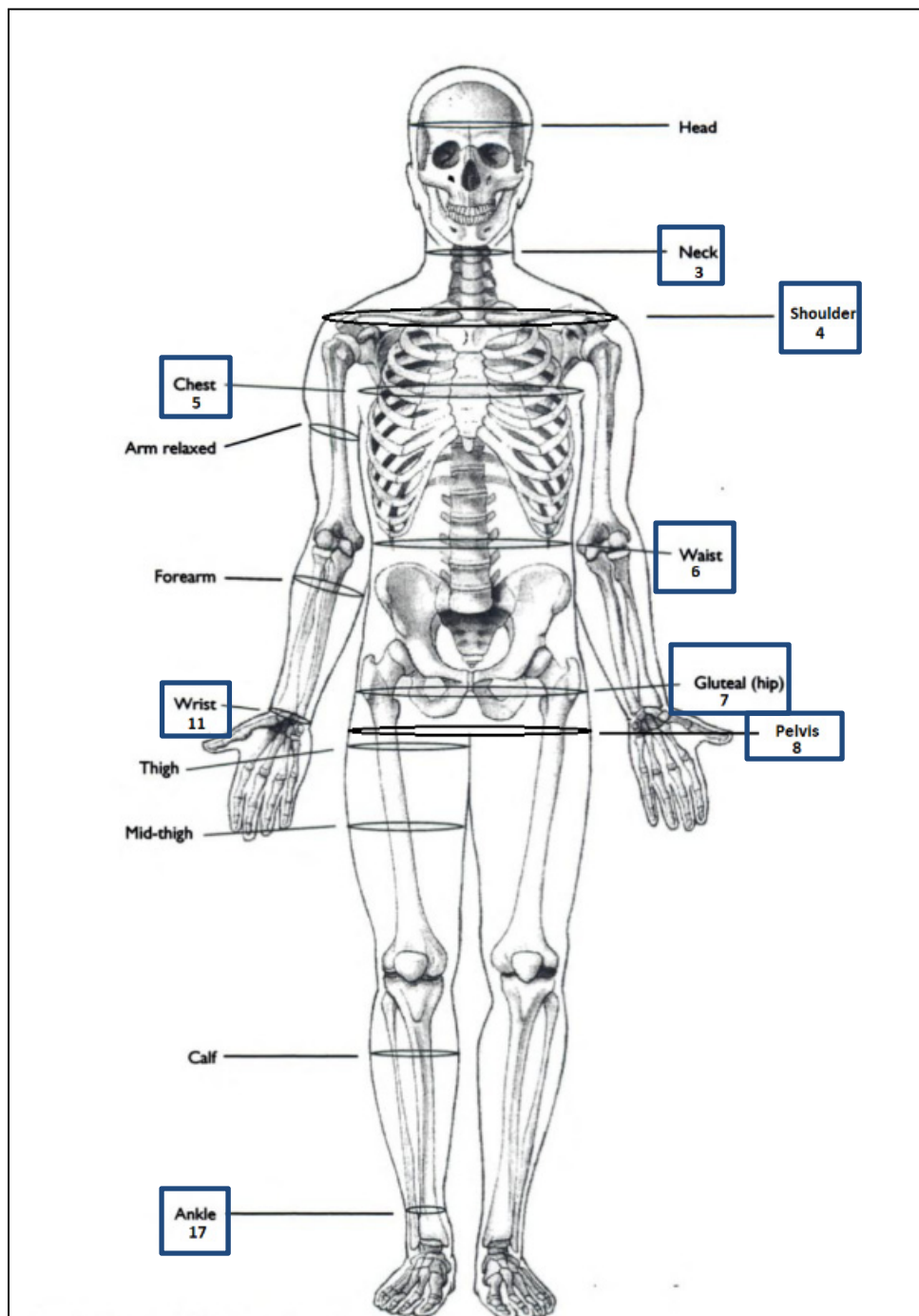


Figure 3.3 Mesures anthropométriques partie 1
Tiré de Norton & Olds (1996, p.57)

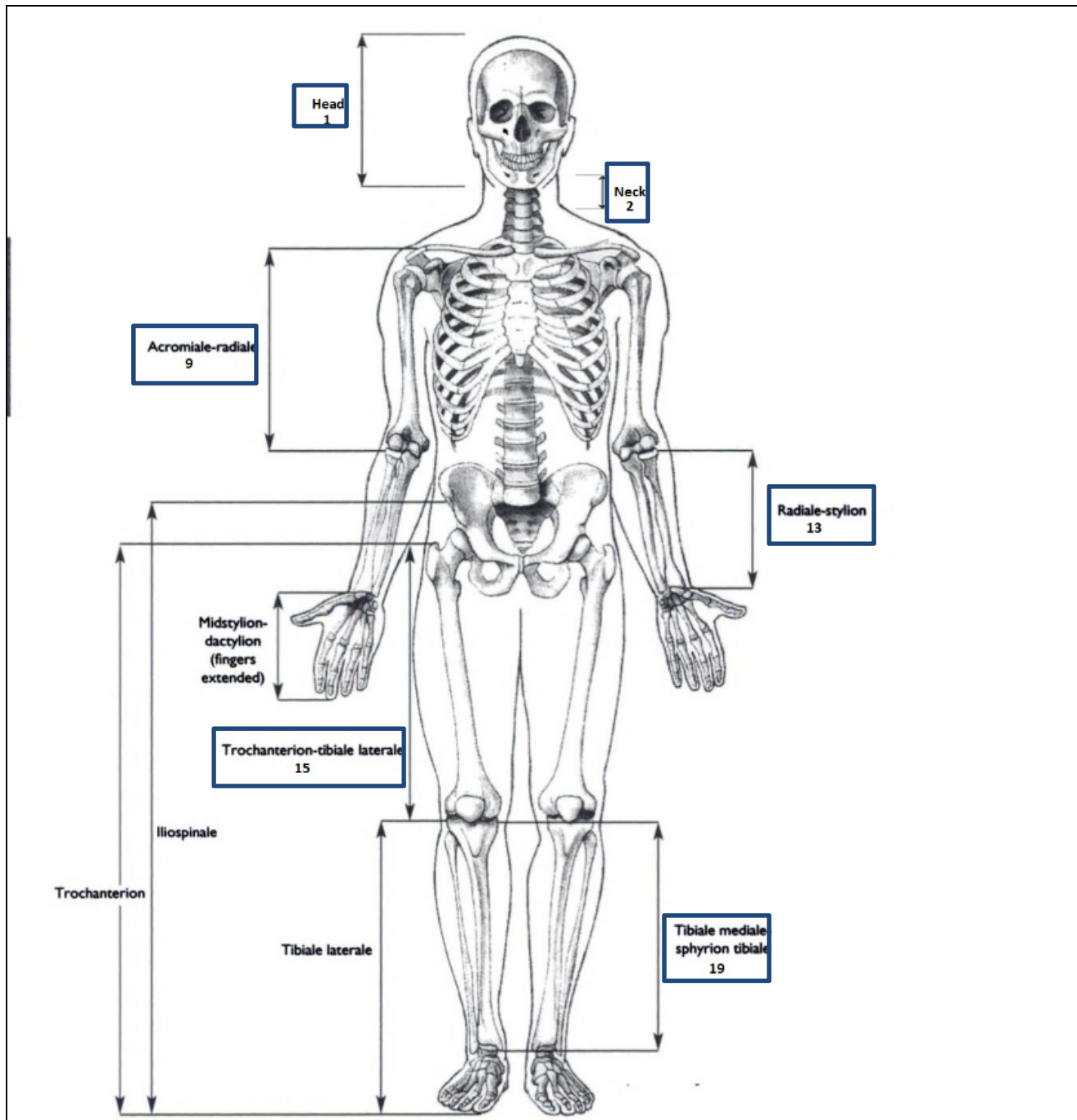


Figure 3.4 Mesures anthropométriques partie 2
Tiré de Norton & Olds (1996, p.64)

Le sujet pouvait ensuite se placer sur la marque de départ de 2m. À cet instant, il est possible de générer l'avatar SMPL-X et comparer la morphologie du sujet à celle de SMPL-X.

Une fois tous les systèmes de capture démarrés (OptiTrack, Intel RealSense D435I, GoPro Hero 7), l'expérimentation peut être définie en deux segments : rotation de la tête à une distance constante et déplacement de la caméra à différentes distances.

Pour le premier segment, le sujet faisait un pas vers l'avant et un pas vers l'arrière afin de synchroniser les appareils de capture. Il/elle devait ensuite effectuer une rotation de la tête sur l'axe vertical le plus loin que possible de façon à recouvrir la totalité du champ de vision de celui-ci tout en gardant sa position immobile et la vision droite devant. Chacune des images dans le vidéo accompagné des données Mocap est utilisée comme une entrée indépendante dans notre base de données. Celles-ci permettent de vérifier la rotation de la tête que OpenFace permet d'obtenir contre les mesures théoriques du OptiTrack. Vu l'importance de l'angle dans l'équation (9), il est important que celui-ci soit le plus précis possible. Cette étape démontrera l'erreur avec lequel cette expérimentation commence. Cette étape nous permet aussi de faire un compte rendu de l'amélioration pratique de PD_p et de l'estimation de la distance utilisant l'équation (9) avec et sans l'utilisation de l'angle d'inclination de la tête. Il est aussi possible d'utiliser cette base de données afin de générer l'avatar à l'échelle en utilisant la méthodologie proposée. Ainsi, utilisant MeshLab (Cignoni et al., 2011), il est possible de mesurer les dimensions anthropométriques décrites dans l'ANNEXE II et de les comparer aux mesures théoriques prises antérieurement.

Dans la deuxième section de l'expérimentation, le support à caméra sera déplacé de façon à se rapprocher du sujet à une distance de 0,5m jusqu'à une distance de 3,5m. Il est possible d'utiliser deux images dans le temps du vidéo et de prendre les positions respectives déterminées par le système OptiTrack afin de calculer PD en utilisant l'équation (12). Cela nous permet de comparer PD avec sa valeur théorique mesurée. Cependant, cette comparaison est limitée par la précision de la règle utilisée lors de la mesure de la distance pupillaire (0,5mm).

CHAPITRE 4

RÉSULTATS ET DISCUSSION

Dans ce chapitre, nous présenterons et analyserons les résultats de la génération de l’avatar normalisé, la précision de l’estimation de l’angle, la précision de l’estimation de la rotation, l’estimation de PD_p , l’estimation de la distance du sujet, l’estimation des mesures anthropométriques et l’estimation de l’écart pupillaire.

4.1 Ajustement d’un avatar à partir d’une image

Les 10 sujets de l’expérimentation peuvent être observés dans la Figure 4.1. Pour des questions de confidentialité, leurs visages furent anonymisés. Les Figure 4.2, Figure 4.3, et Figure 4.4 représentent respectivement la pose estimée par OpenPose (Cao et al., 2018), leur avatar généré superposé sur ceux-ci, ainsi que l’avatar isolé. Théoriquement, si l’utilisation de SMPLify-X et d’OpenPose avait été parfaite, alors l’image 2D de l’avatar recouvrant le sujet serait parfaite.

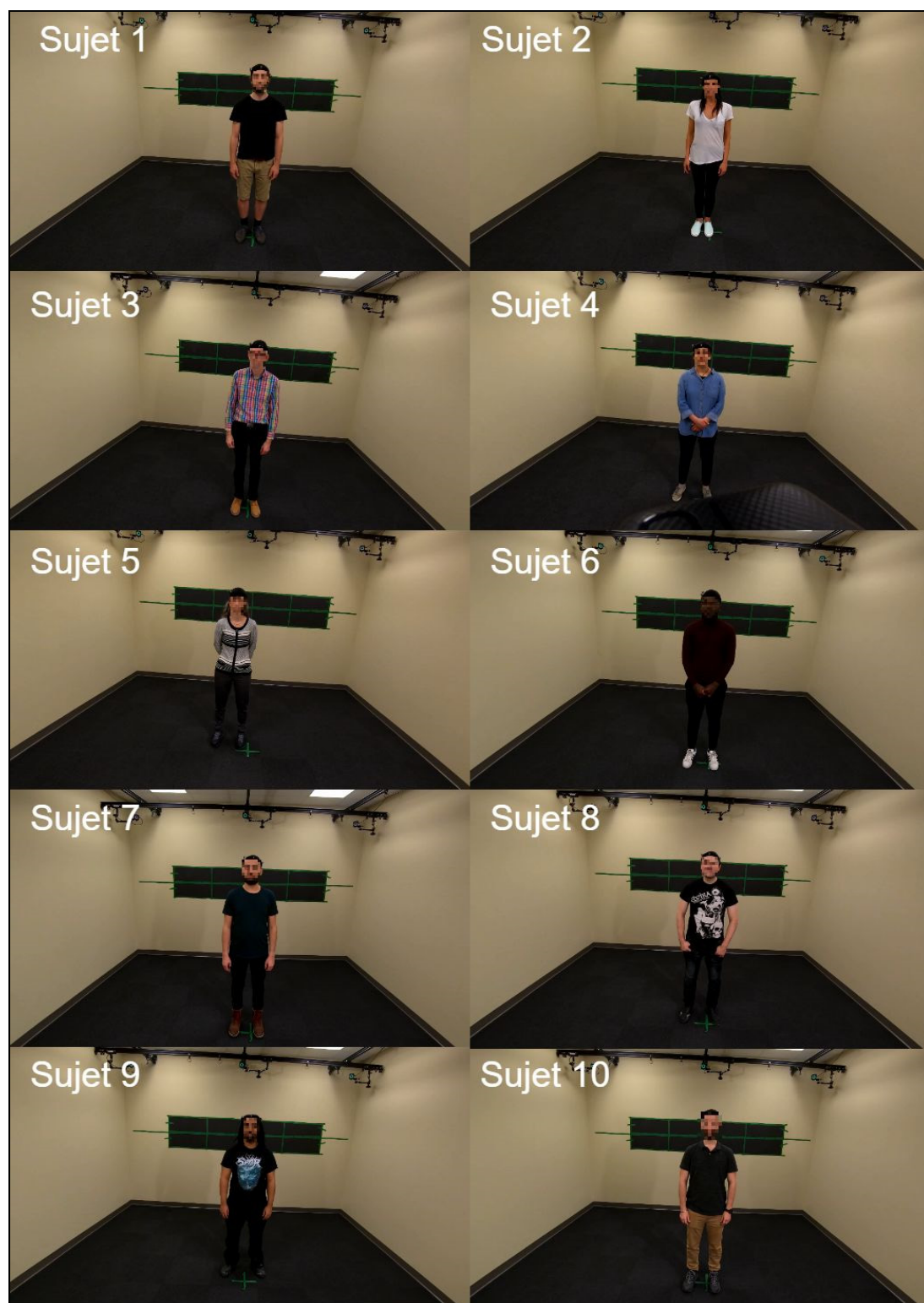


Figure 4.1 Sujet à la position initiale

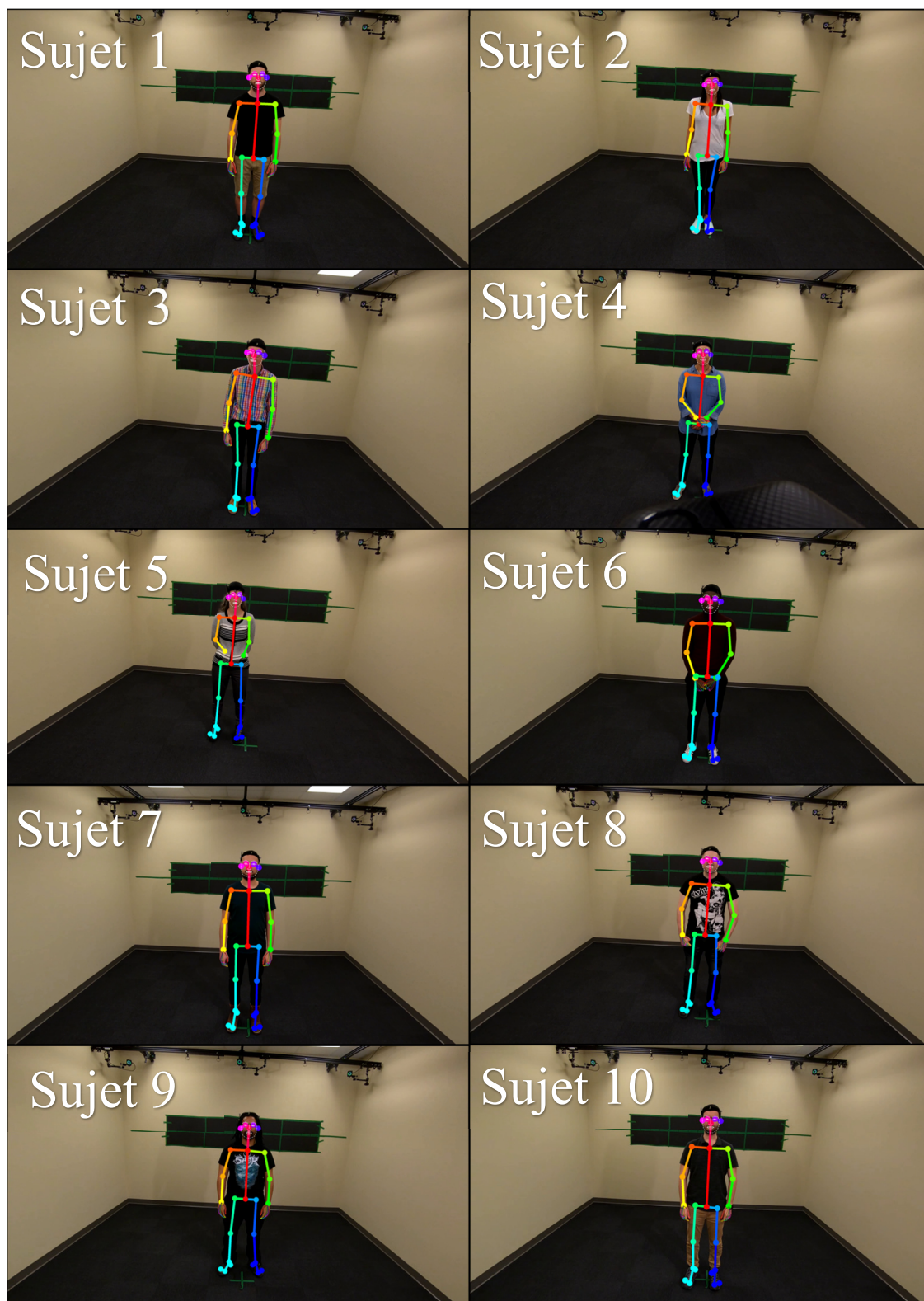


Figure 4.2 Détection de la pose sur sujet

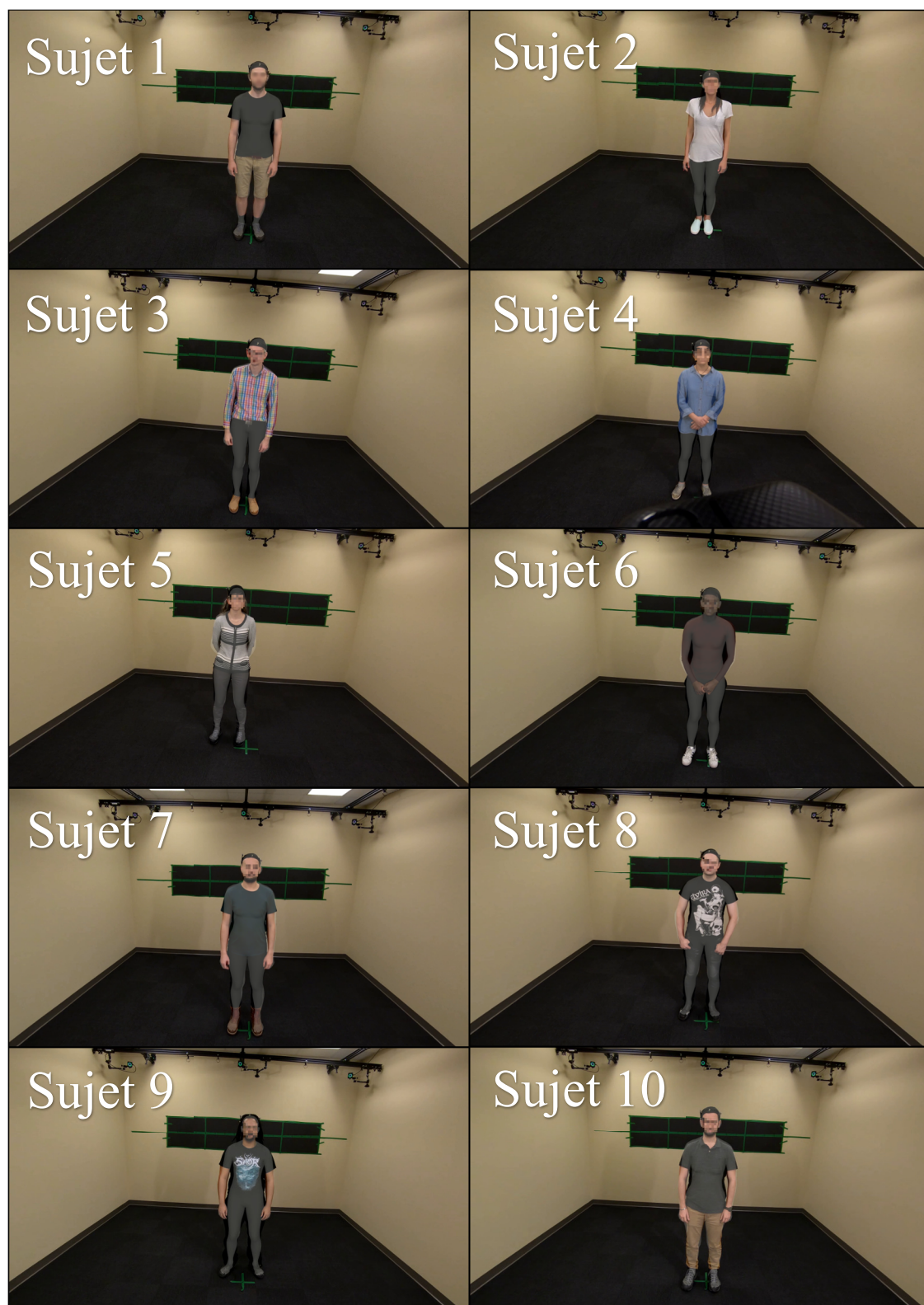


Figure 4.3 Superposition de l'avatar sur le sujet

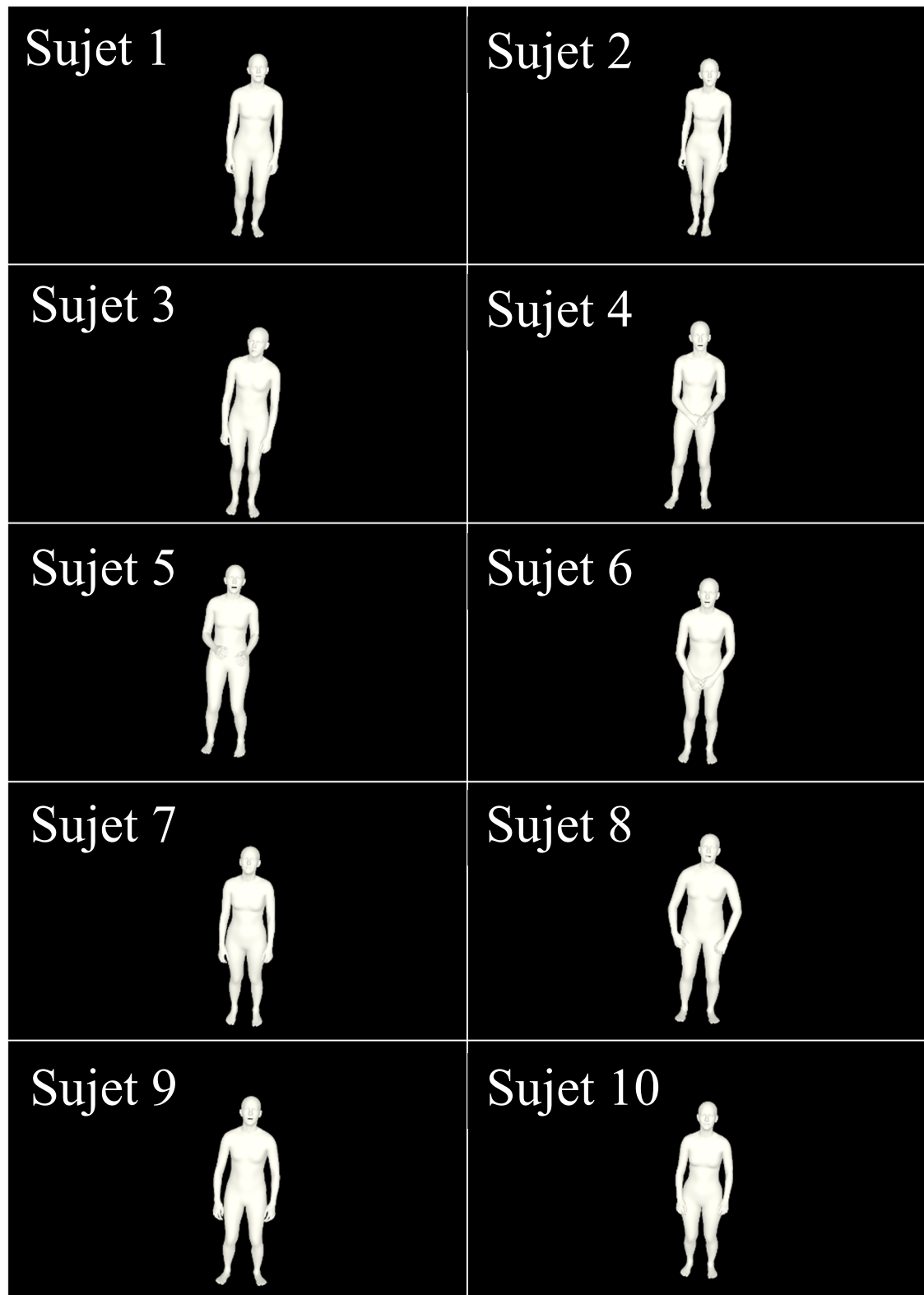


Figure 4.4 Avatars générés

Tel que démontré dans la Figure 4.2, la détection de la pose ainsi que les marqueurs du visage et des mains répondent aux attentes qui étaient d’avoir des marqueurs pour le visage, les mains, le corps, ainsi que les pieds afin de fournir ceux-ci comme données d’entrées à SMPLify-X. Aucun marqueur est à l’extérieur du sujet. Cependant, lorsque les mains sont une sur l’autre, les marqueurs des mains n’existent tout simplement pas ce qui empêchera donc le remplacement des mains du modèle SMPL-X. Aucune analyse plus approfondie sur la précision des marqueurs n’a été faite puisque ceci a été complété dans (Cao et al., 2018).

En comparant les avatars résultants au sujet initial tel que démontré dans la Figure 4.3, il est possible d’accepter cette solution comme avatar. Celui-ci recale très bien le sujet respectif. Le corps, les épaules, les cuisses ainsi que le visage sont particulièrement bien réussies. De plus, les pieds sont particulièrement bien recalés considérant que tous les sujets portaient des souliers et que ces souliers ne devaient pas être pris en compte lors de la génération de l’avatar dans cette recherche. Cependant, il est possible d’observer sur le sujet 2 que la tête de l’avatar est légèrement plus haute que le haut de la tête du sujet ce qui aura pour effet d’affecter la grandeur estimée. Il y a également le défaut que les mains de l’avatar du sujet 5 sont devant le corps lorsqu’en réalité, les mains devraient être derrière le corps et les doigts se croisent. Ceci est causé par le fait qu’il n’existe aucune position pour les mains et ses articulations dans l’estimation de la pose. Dans les sujets 4 et 6, les doigts ne se croisent pas comme ceux des sujets. Ceci est probablement dû au fait que la position du croisement des doigts comporte beaucoup d’occlusions qui ne sont pas évidentes à résoudre malgré les bons points de position 2D fournie par OpenPose (Cao et al., 2018). Une estimation de pose à trois dimensions aurait possiblement permis d’améliorer le résultat dans ce contexte. Bien que cela n’affecte pas directement le but de ce travail, il est bien de noter ce défaut pour de futures recherches. Il y a aussi une possibilité d’amélioration dans l’estimation de la position des bras. Presque tous les sujets ont un léger décalage entre l’emplacement des bras et les bras de l’avatar. Ceci ne devrait pas affecter grandement les résultats puisque la grosseur est respectée majoritairement par une estimation visuelle. Il existe aussi une erreur pour les sujets 2 et 3 au niveau des cuisses. Ceux-ci sont légèrement plus gros que la réalité. Une explication plausible à cela est que les deux sujets sont particulièrement minces. Le modèle de l’avatar générique est basé sur un modèle

générique et modifié par des paramètres en utilisant un réseau de neurones entraîné. Il est possible que le réseau soit entraîné avec peu de personnes ayant une taille beaucoup plus petite que la moyenne au niveau des cuisses ce qui a pour effet qu'il n'est pas capable d'estimer les paramètres adéquatement pour ce type de personnes au niveau des cuisses.

4.2 Ajustement de l'avatar à partir de la distance inter pupillaire

4.2.1 Angle de rotation

Dans cette expérience, la seule rotation importante était le lacet. Les résultats théoriques représentés par les données acquises avec le MOCAP peuvent être représentés par une courbe temps versus la rotation de la tête. Le sujet 10 est illustré dans la Figure 4.5 à titre d'exemple.

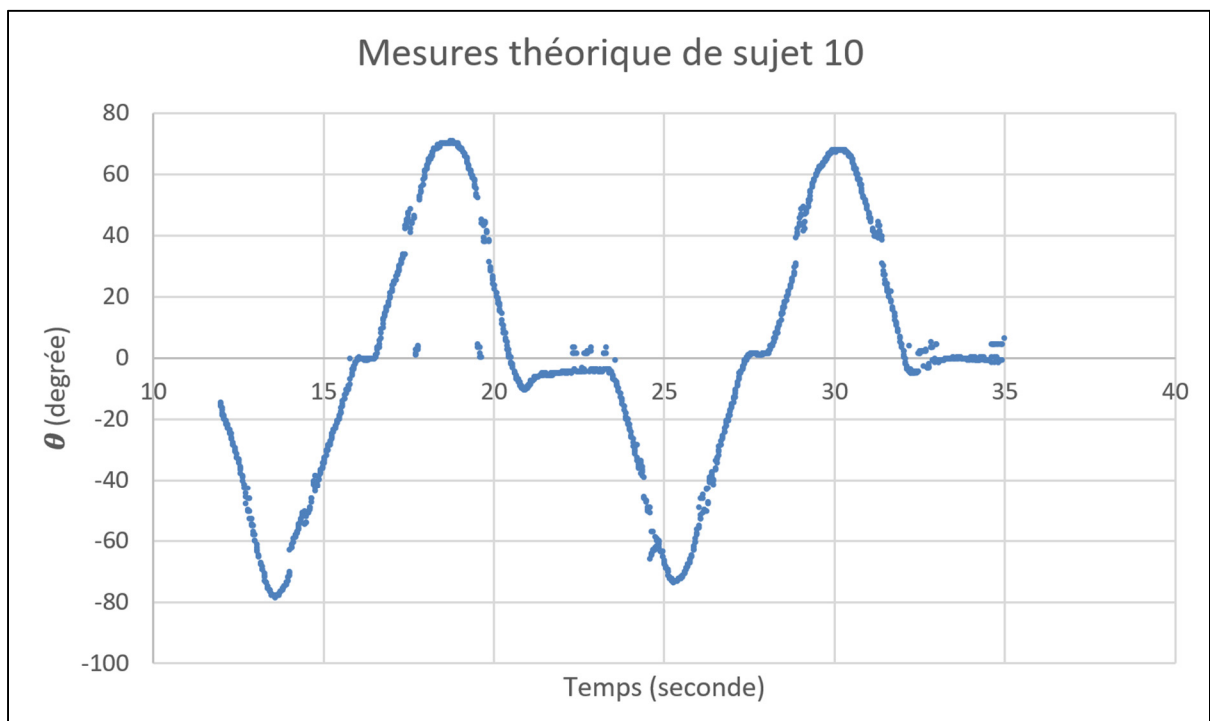


Figure 4.5 Courbe empirique du lacet du Sujet 10

La Figure 4.5 représente bien le lacet de la tête lors de la première période de l'expérience où le sujet exécute une rotation du lacet au maximum de gauche à droite à deux reprises. Au sein de toutes les courbes empiriques des sujets tels que la Figure 4.5, il y a des points qui peuvent

être considérés comme étant aberrants puisqu'ils ne suivent pas la courbe tel qu'attendu. Ceci peut être expliqué par le fait que la tête tournait à une vitesse incontrôlée (trop rapidement) ce qui avait pour effet au système Mocap de perdre suffisamment de points de références sur la tête du sujet qu'il n'était pas capable de reconstruire adéquatement l'emplacement de celui-ci. Par conséquent, ces points furent retirés de l'évaluation.

Selon la Figure 1.11, OpenFace 2.0 devrait être capable d'estimer la rotation de la tête à une erreur moyenne de $3,2^\circ$. Celle-ci est obtenu sur la Figure 4.6, nous pouvons observer la courbe théorique en bleu par rapport à la courbe estimée en rouge du sujet 10. Nous pouvons clairement identifier que la courbe estimée diverge de la courbe théorique autour de 40° .

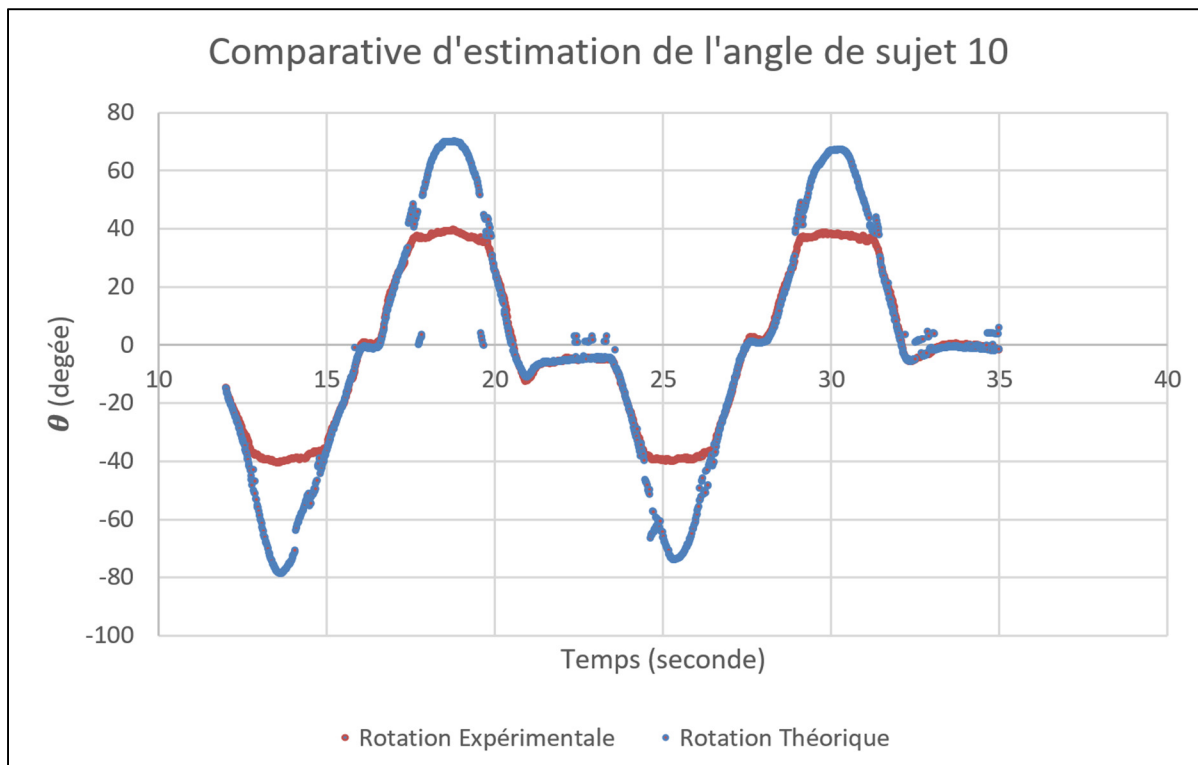
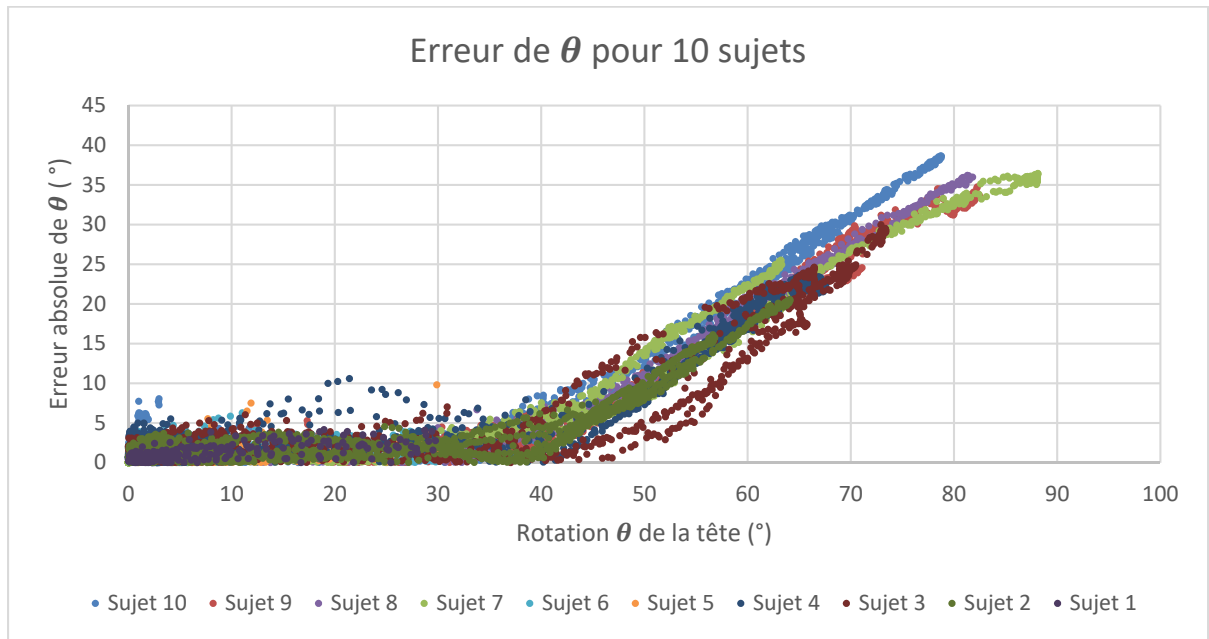


Figure 4.6 Estimation du lacet du sujet 10

Afin de mieux déterminer l'angle où l'algorithme de détection perd sa précision, la Figure 4.7 a été produite.

Figure 4.7 Erreur de $|\theta|$ de 10 SujetTableau 4.1 Résultat de $|\Delta\theta|$

	AVG	STD	MIN	MAX
$ \Delta\theta , \theta < 30^{\circ}$	1.08	0.96	0.00	10.57
$ \Delta\theta $	5.96	8.85	0.00	38.65

Celle-ci démontre que pour tous les sujets, les résultats de détection restent généralement acceptables (sous la barre des 10°) jusqu'à un angle de 30° . Tel que démontré dans le Tableau 4.1, la moyenne d'erreur de la rotation estimée peut être représentée par une fonction constante soit $y = 1.08^{\circ}$ avant 30° , par la suite, l'erreur augmente rapidement et il est impossible d'estimer précisément l'orientation de la tête. On peut donc définir que la limite d'estimation de l'angle de rotation de OpenFace (Baltrušaitis et al., 2016) est utilisable jusqu'à 30° . Celle-ci a été déterminée en observant la Figure 4.7.

4.2.2 Estimation de la distance pupillaire corrigée dans l'image

Une bonne estimation de \mathbf{PD}_p est cruciale à l'obtention de la distance d'un sujet puisqu'elle englobe les variables de l'équation (9). Comme décrit ci-dessus, la mesure de la \mathbf{PD}_I devrait rester identique à une distance constante entre le centre de la distance pupillaire et la caméra. La Figure 4.8 compare la mesure de \mathbf{PD}_p en utilisant notre méthodologie contre celle qui ne considère pas l'angle ($\mathbf{PD}_I = \mathbf{PD}_p$) (Deák et al., 2017) sur l'intervalle de -75° à 75° . Le

Tableau 4.2 établit les valeurs numériques représentées de cette figure. Ceux-ci démontrent que considérer l'angle de rotation du lacet de la tête aide grandement à avoir une estimation plus précise de \mathbf{PD}_p . Plus l'angle de rotation est petit, plus les deux méthodes se rapprochent.

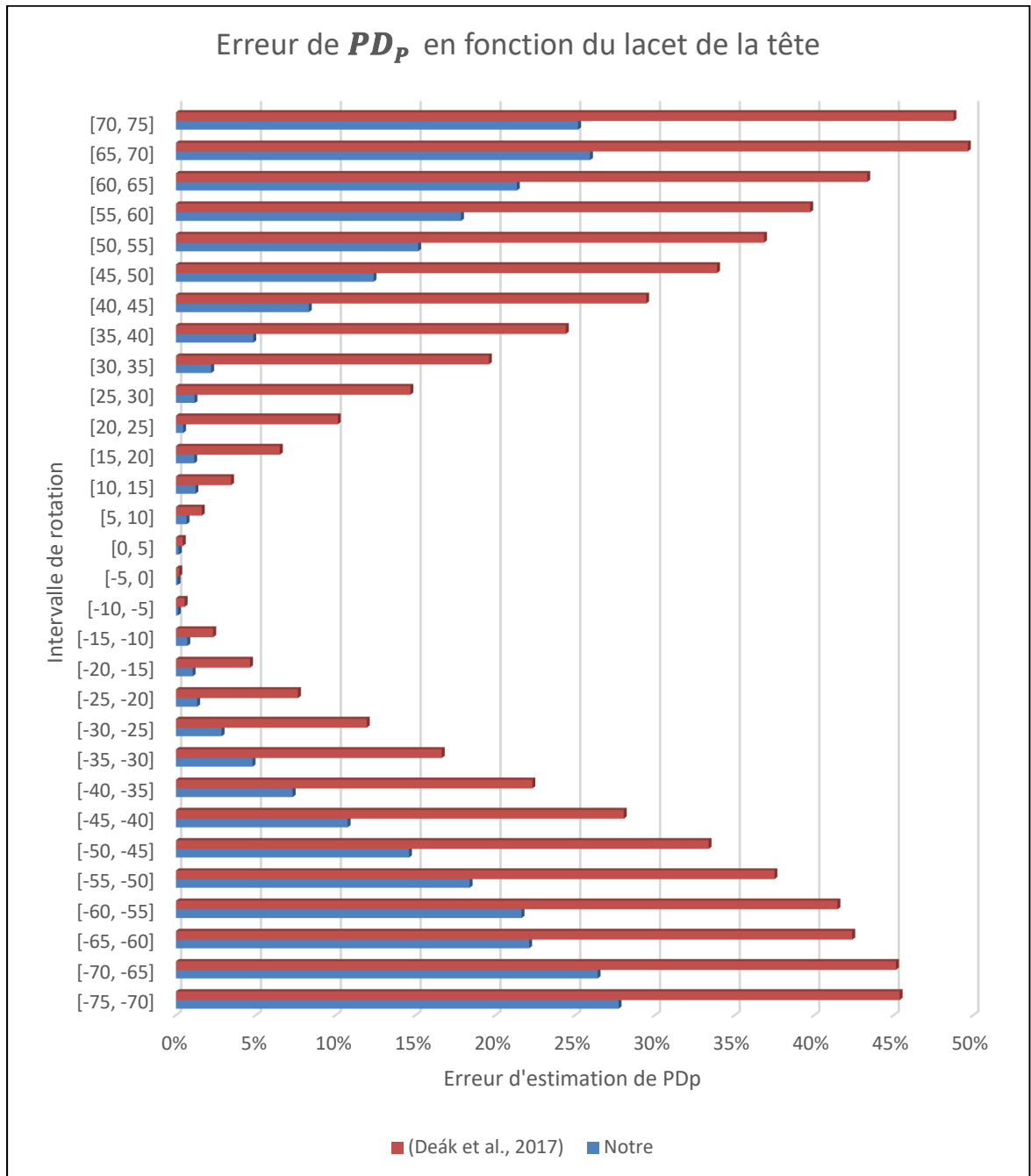
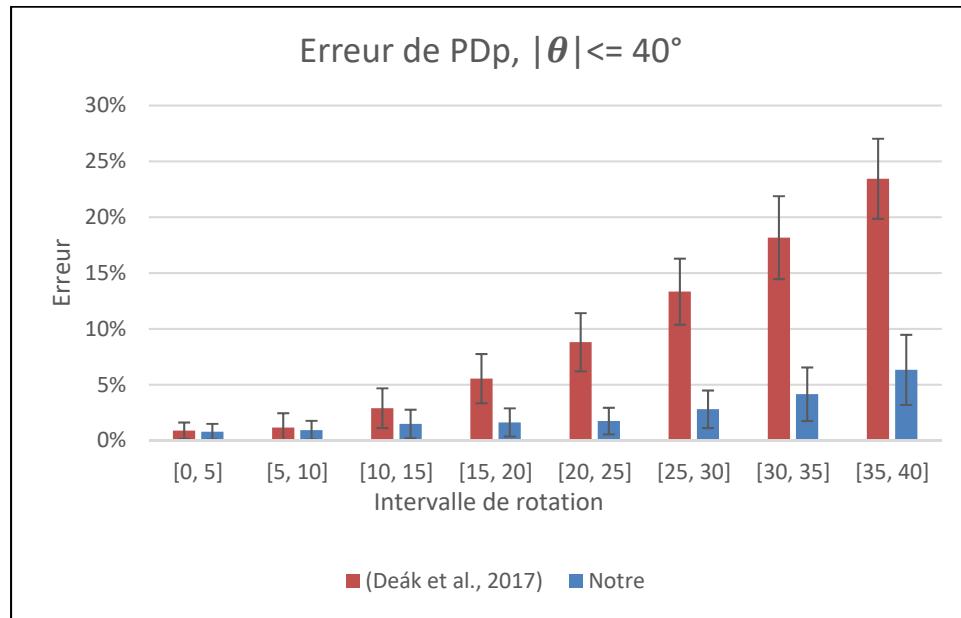


Figure 4.8 Erreur de PD_p en fonction du lacet de la tête

Tableau 4.2 Erreur de PD_p en fonction du lacet de la tête

Intervalle de rotation	Nb de points	Moyenne		σ	
		(Deák et al., 2017)	Notre	(Deák et al., 2017)	Notre
[-75, -70]	212	45.37%	27.72%	4.83%	4.53%
[-70, -65]	337	45.14%	26.41%	7.26%	7.09%
[-65, -60]	445	42.40%	22.12%	4.49%	4.75%
[-60, -55]	270	41.47%	21.65%	4.42%	5.21%
[-55, -50]	227	37.50%	18.40%	4.16%	4.85%
[-50, -45]	230	33.37%	14.59%	3.81%	3.87%
[-45, -40]	204	28.06%	10.75%	3.71%	3.54%
[-40, -35]	195	22.32%	7.31%	3.70%	3.46%
[-35, -30]	192	16.65%	4.80%	3.54%	2.63%
[-30, -25]	205	11.96%	2.86%	2.40%	1.94%
[-25, -20]	237	7.64%	1.33%	2.08%	1.45%
[-20, -15]	226	4.64%	1.05%	1.78%	1.59%
[-15, -10]	236	2.33%	0.71%	1.57%	1.49%
[-10, -5]	717	0.54%	0.12%	1.09%	1.11%
[-5, 0]	1314	0.21%	0.10%	1.09%	1.08%
[0, 5]	1008	0.43%	0.18%	1.12%	1.01%
[5, 10]	268	1.62%	0.67%	2.12%	1.42%
[10, 15]	220	3.46%	1.22%	1.89%	1.89%
[15, 20]	211	6.50%	1.14%	2.22%	1.89%
[20, 25]	207	10.14%	0.45%	2.51%	2.22%
[25, 30]	209	14.68%	1.16%	2.83%	2.85%
[30, 35]	204	19.60%	2.20%	3.29%	3.40%
[35, 40]	221	24.42%	4.85%	3.18%	3.58%
[40, 45]	291	29.45%	8.32%	3.07%	3.02%
[45, 50]	308	33.89%	12.38%	3.22%	3.70%
[50, 55]	271	36.84%	15.17%	2.52%	3.71%
[55, 60]	473	39.75%	17.86%	2.40%	3.98%
[60, 65]	381	43.34%	21.35%	4.32%	4.08%
[65, 70]	230	49.66%	25.94%	4.18%	3.83%
[70, 75]	55	48.77%	25.18%	1.79%	2.87%

Afin de mieux analyser les résultats, les résultats ont été mis en absolu tel que représenté dans le Tableau 4.3. La Figure 4.9 est une représentation visuelle des résultats sur l'intervalle absolu de 0° à 40° .

Figure 4.9 Erreur de PD_p , $|\theta| \leq 40^\circ$ Tableau 4.3 Erreur de PD_p , $|\theta|$

Intervalle de rotation	Nb de point	Moyenne		σ	
		(Guan, 2009)	Notre	(Guan, 2009)	Notre
[0, 5]	2322	0.88%	0.79%	0.74%	0.71%
[5, 10]	985	1.15%	0.93%	1.30%	0.84%
[10, 15]	456	2.90%	1.48%	1.78%	1.28%
[15, 20]	437	5.54%	1.62%	2.21%	1.26%
[20, 25]	444	8.81%	1.74%	2.61%	1.19%
[25, 30]	414	13.33%	2.80%	2.96%	1.68%
[30, 35]	396	18.17%	4.15%	3.71%	2.40%
[35, 40]	416	23.44%	6.33%	3.59%	3.14%
[40, 45]	495	28.88%	9.36%	3.42%	3.36%
[45, 50]	538	33.67%	13.33%	3.49%	3.92%
[50, 55]	498	37.14%	16.65%	3.38%	4.52%
[55, 60]	743	40.38%	19.25%	3.38%	4.74%
[60, 65]	826	42.83%	21.77%	4.44%	4.47%
[65, 70]	567	46.98%	26.22%	6.58%	5.99%
[70, 75]	267	46.07%	27.20%	4.59%	4.36%

D'après le Tableau 4.3, à angle de 0 à 5 degrés, notre solution se rapproche grandement d'une approche qui n'utiliserait pas la rotation de la tête telle que (Barron & Kakadiaris, 2000). C'est le comportement attendu puisque à cet intervalle, l'angle est quasiment nul et donc d'après (9), $\frac{PD_I}{\cos(\theta)} \cong PD_I$. La légère amélioration entre notre proposition et celle existante de 0.09% peut être expliqué par le fait que l'angle n'est pas seulement à 0° à l'intérieur de l'intervalle. Lorsque l'intervalle d'angle augmente, il est possible d'observer que l'erreur des deux solutions augmente. Néanmoins, la solution proposée améliore grandement les résultats en fonction de l'augmentation de l'intervalle. Plus précisément, elle présente une erreur de moins de 3% lorsqu'on ne prend en considération que les angles valides de la section 4.2. Il s'agit d'une amélioration considérable si on la compare à la solution précédente qui fournissait un résultat moyen allant jusqu'à 13,33% dans le même intervalle. En outre, en analysant les mêmes intervalles, il est possible d'identifier que la stabilité des résultats est également améliorée lorsque la rotation augmente. En analysant l'intervalle de [25, 30], la solution proposée améliore l'écart type de 2,96 % à 1,68 %.

Il est important de noter qu'il reste une erreur sur la solution proposée dans l'intervalle valide définie dans la section 4.2 et que cette erreur augmente à mesure que l'angle augmente. Ceci peut être expliqué par le fait que notre solution utilise les marqueurs détectés autour des yeux à partir de *OpenFace* (Baltrušaitis et al., 2016) pour estimer les centres des yeux qui seront utilisés pour calculer la distance pupillaire dans l'image (PD_I). Ces marqueurs ne prennent pas en compte la rotation de la tête. Ainsi, la rotation de la tête a un impact similaire sur les marqueurs que sur PD_I . Cet impact déplace légèrement le centre des yeux réduisant la distance pupillaire dans l'image. C'est pourquoi l'erreur augmente lorsque la rotation de la tête augmente. Dans une recherche future, il serait intéressant de proposer une solution qui tienne compte de la rotation de la tête sur les marqueurs, ce qui permettrait d'améliorer encore plus la précision du PD_P .

Une autre erreur qui n'a pas été prise en compte lors de la réalisation de l'expérience est que la distance du centre de l'écart pupillaire augmente légèrement lorsque le sujet tourne la tête puisque le centre de la tête s'éloigne de la caméra lorsque l'angle augmente. L'utilisation d'un

PD_I fixe introduit une erreur puisqu'il devrait augmenter et diminuer légèrement lorsque le sujet tourne la tête. La méthodologie calcule la distance entre le centre des deux yeux et la caméra. Lorsque le sujet tourne sa tête, ce centre se déplace légèrement en profondeur et donc PD_I varie légèrement. Cependant, étant donné que la véritable PD_I n'était pas calculable, une PD_I fixe a été utilisée (lorsque $\theta = 0$).

4.2.3 Distance du sujet

Les données théoriques utilisées pour la comparaison des résultats obtenus étaient la distance entre les marqueurs Mocap de la caméra et les marqueurs Mocap de la tête du sujet. La Figure 4.10 est une représentation de l'erreur d'estimation de la distance normalisée à distance de $2m \pm 0.2m$ à différentes rotations entre -75° à 75° en utilisant trois différentes méthodes : la méthodologie existante en rouge (Deák et al., 2017) , notre méthodologie utilisant la distance pupillaire moyenne en vert (notre moy. PD), notre méthodologie utilisant la vraie distance inter pupillaire en bleu (notre). Le Tableau 4.4 présente les valeurs numériques à la Figure 4.10 ainsi que l'écart type de ces résultats.

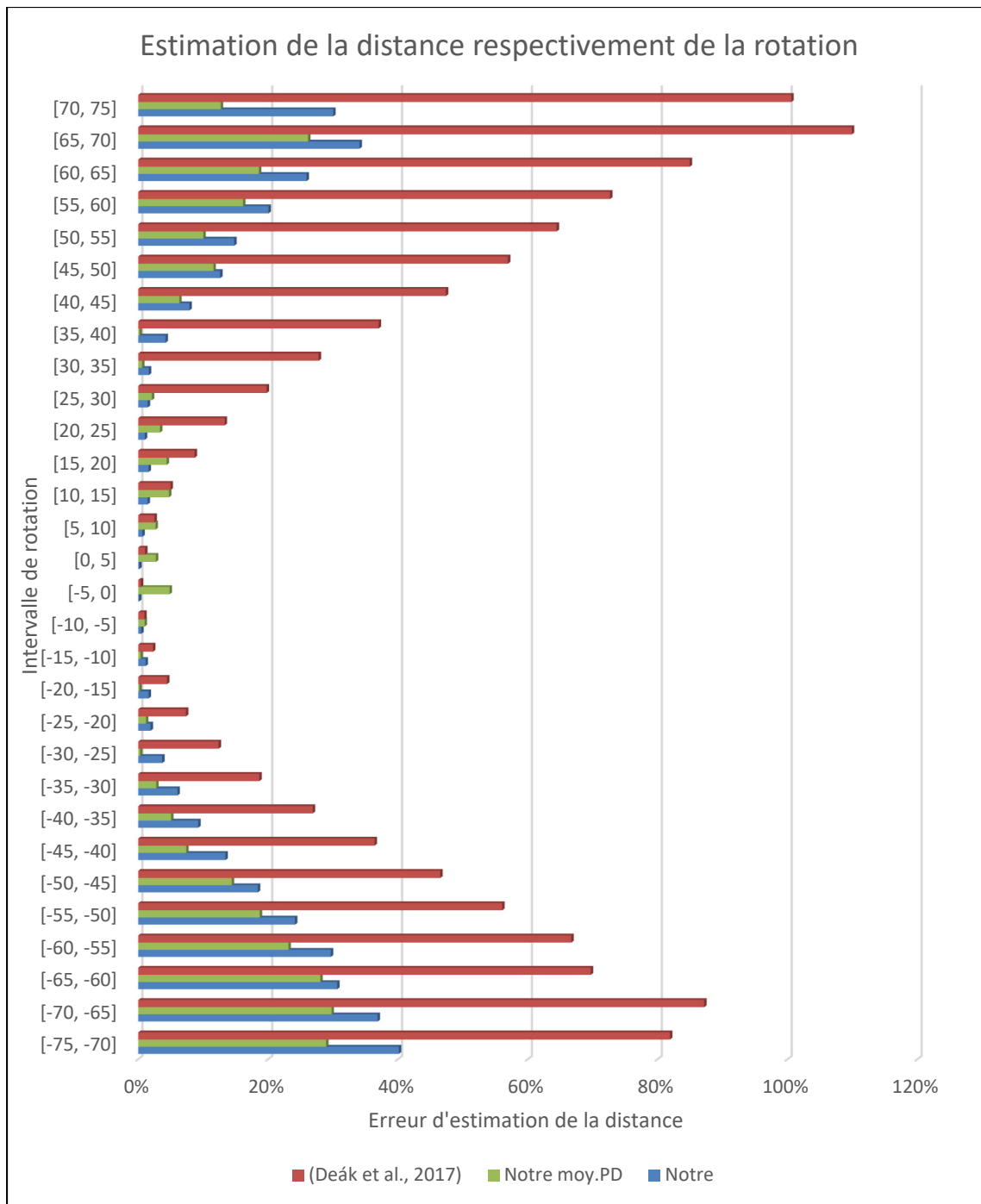


Figure 4.10 Estimation de la distance en fonction de la rotation

Tableau 4.4 Estimation de la distance en fonction de la rotation

Intervalle de rotation	Nb de points	Moyenne			σ		
		(Deák et al., 2017)	Notre moy.PD	Notre	(Deák et al., 2017)	Notre moy.PD	Notre
[-75, -70]	214	81.92%	28.96 %	40.15 %	27.40 %	9.47 %	11.63 %
[-70, -65]	346	87.19%	29.80 %	36.88 %	53.28 %	18.87 %	18.26 %
[-65, -60]	445	69.72%	28.06 %	30.65 %	14.16 %	10.55 %	8.08 %
[-60, -55]	270	66.72%	23.15 %	29.68 %	11.04 %	9.89 %	7.99 %
[-55, -50]	227	56.08%	18.73 %	24.15 %	8.68 %	9.93 %	6.62 %
[-50, -45]	230	46.53%	14.43 %	18.44 %	7.62 %	10.35 %	4.84 %
[-45, -40]	197	36.41%	7.40 %	13.45 %	6.81 %	9.52 %	4.12 %
[-40, -35]	186	26.89%	5.07 %	9.25 %	5.92 %	9.62 %	3.66 %
[-35, -30]	189	18.67%	2.78 %	6.06 %	5.54 %	10.43 %	2.99 %
[-30, -25]	204	12.40%	0.30 %	3.70 %	2.94 %	9.08 %	2.02 %
[-25, -20]	231	7.37%	1.20 %	1.92 %	2.53 %	8.88 %	1.69 %
[-20, -15]	233	4.46%	0.20 %	1.61 %	2.05 %	8.81 %	1.92 %
[-15, -10]	223	2.28%	0.37 %	1.15 %	1.63 %	7.33 %	1.69 %
[-10, -5]	617	0.95%	1.00 %	0.48 %	1.06 %	7.98 %	1.17 %
[-5, 0]	1289	0.42%	4.84 %	0.15 %	1.03 %	5.21 %	0.93 %
[0, 5]	1118	1.07%	2.73 %	0.17 %	1.23 %	6.11 %	0.84 %
[5, 10]	299	2.52%	2.68 %	0.65 %	2.54 %	7.43 %	1.43 %
[10, 15]	219	4.99%	4.76 %	1.45 %	2.16 %	8.12 %	1.76 %
[15, 20]	206	8.71%	4.38 %	1.60 %	2.97 %	8.29 %	1.83 %
[20, 25]	209	13.32%	3.36 %	1.03 %	3.64 %	9.17 %	2.12 %
[25, 30]	212	19.79%	2.10 %	1.48 %	4.57 %	10.08 %	2.84 %
[30, 35]	216	27.82%	0.61 %	1.64 %	5.91 %	11.09 %	3.45 %
[35, 40]	223	37.06%	0.23 %	4.22 %	5.86 %	10.02 %	4.20 %
[40, 45]	295	47.40%	6.33 %	7.90 %	7.22 %	10.75 %	5.15 %
[45, 50]	324	56.97%	11.60 %	12.64 %	9.97 %	10.88 %	6.63 %
[50, 55]	291	64.47%	10.04 %	14.80 %	6.95 %	11.65 %	8.48 %
[55, 60]	482	72.69%	16.17 %	20.07 %	7.21 %	9.91 %	6.45 %
[60, 65]	381	84.96%	18.61 %	25.94 %	16.25 %	7.38 %	6.84 %
[65, 70]	230	109.93%	26.17 %	34.10 %	21.10 %	13.69 %	8.82 %
[70, 75]	55	100.59%	12.72 %	30.04 %	9.49 %	7.48 %	8.61 %

Il est possible de voir l'amélioration apportée par notre solution proposée par rapport aux solutions existantes. Cette amélioration augmente au fur et à mesure que l'intervalle de rotation augmente. La Figure 4.11 est une représentation similaire, mais utilisant l'erreur absolue sur une rotation absolue de 0° à 40° afin de concentrer l'analyse des résultats.

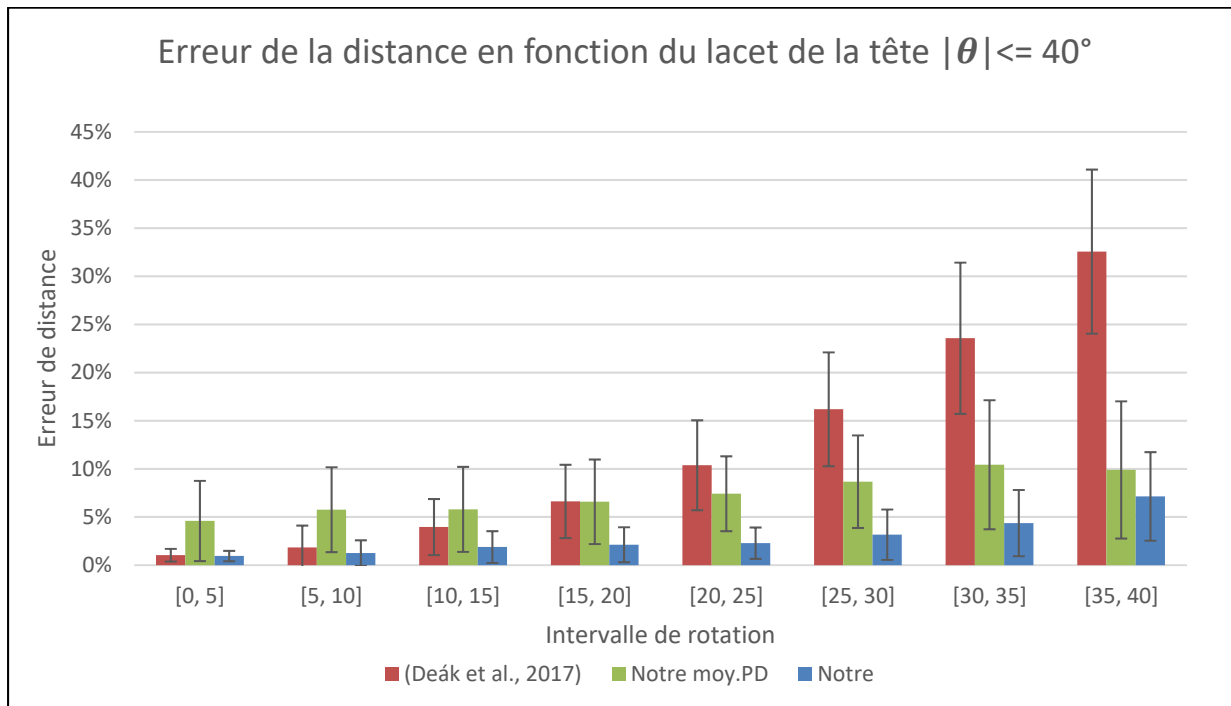


Figure 4.11 Erreur de la distance en fonction du lacet de la tête, $|\theta| \leq 40^\circ$

Tableau 4.5 Erreur de la distance en fonction du lacet de la tête, $|\theta| \leq 40^\circ$

Interval le de rotation	Nb de point	Moyenne			σ		
		(Deák et al., 2017)	Notre moy.PD	Notre	(Deák et al., 2017)	Notre moy. PD	Notre
[0, 5]	2125	1.03%	4.59%	0.95%	0.66%	4.17%	0.54%
[5, 10]	993	1.83%	5.76%	1.26%	2.28%	4.41%	1.32%
[10, 15]	452	3.96%	5.79%	1.88%	2.91%	4.41%	1.65%
[15, 20]	452	6.62%	6.58%	2.12%	3.80%	4.39%	1.82%
[20, 25]	421	10.37%	7.42%	2.28%	4.67%	3.89%	1.63%
[25, 30]	382	16.19%	8.66%	3.16%	5.91%	4.81%	2.62%
[30, 35]	377	23.56%	10.43%	4.37%	7.86%	6.70%	3.44%
[35, 40]	329	32.56%	9.89%	7.14%	8.52%	7.13%	4.60%

En analysant l'intervalle valide le plus élevé déterminé par la section 4.2 [25°, 30°], il est possible de voir que la solution proposée a une erreur de distance moyenne de 3,16% alors que l'erreur moyenne des solutions existantes s'élève à 16,19 %. Cela signifie que la solution proposée augmente la précision de la distance estimée de plus de cinq fois. À une distance de 2 mètres, cela représente une erreur de 63,2mm par rapport à 323,8mm. Il est intéressant de noter que l'utilisation de la distance pupillaire moyenne à cet intervalle d'angle fournit de meilleurs résultats que la proposition existante avec un résultat de 8,66 % d'erreur. Bien que tout angle estimé au-dessus de 30° soit imprécis selon la section 4.2, la méthodologie proposée améliore encore l'estimation de la distance par rapport aux solutions existantes. Il est également possible de voir les améliorations apportées par notre solution proposée dans la stabilité des résultats. En analysant l'intervalle valide le plus élevé [25°, 30°], l'écart-type de la méthodologie proposée est de 2,62 % alors que les solutions existantes avaient un écart-type de 5,91 %. Une fois encore, la solution proposée présente un écart-type deux fois moins important que les solutions existantes. En outre, lorsque l'on analyse l'intervalle dépassant l'intervalle valide, la solution proposée fournit de meilleurs résultats, comme le montre le tableau 5.

L'erreur résiduelle dans notre solution proposée peut être liée à l'erreur décrite à la section 4.2.2. Une erreur dans la variable PD_p entraîne une erreur proportionnelle à la distance. C'est pourquoi l'erreur de la section 4.2.2 (2,80 %) est si proche de l'erreur de distance (3,16 %) dans l'intervalle de [25°, 30°]. Une autre source d'erreur provient de la mesure théorique du point central de la distance pupillaire qui fut utilisé afin de calculer la distance théorique. Puisqu'il n'y avait pas de marqueur directement sur ce point, il fut calculé par l'intermédiaire de deux marqueurs à chaque côté de la tête. Cependant, le point central de ces deux marqueurs n'arrive pas nécessairement exactement au même point qu'au point central de la distance pupillaire ce qui introduit une erreur lors des calculs d'erreur de la distance pratique.

Il est également possible d'observer que sur l'intervalle de rotation totale possible [-75°, 75°], la solution proposée avec l'écart pupillaire moyen améliore les résultats. Tel que démontre la Figure 4.14 le seul impact d'utiliser la distance pupillaire moyenne est que la courbe est

déplacée. Ceci a pour effet d'augmenter l'erreur proportionnellement à la différence entre la distance pupillaire moyenne utilisée et la distance pupillaire réelle du sujet. Il est important de noter que dans notre échantillon de sujets pour l'expérience, deux des dix sujets avaient un écart pupillaire supérieur à 95 % de la population. Par conséquent, cela a eu un impact négatif important sur les résultats qui ont utilisé l'écart pupillaire moyen. Si nous avions un échantillon plus grand représentant plus adéquatement la population terrestre, nous aurions probablement observé un meilleur résultat qui démontre que même en utilisant la moyenne nous réduisons substantiellement l'erreur d'estimation.

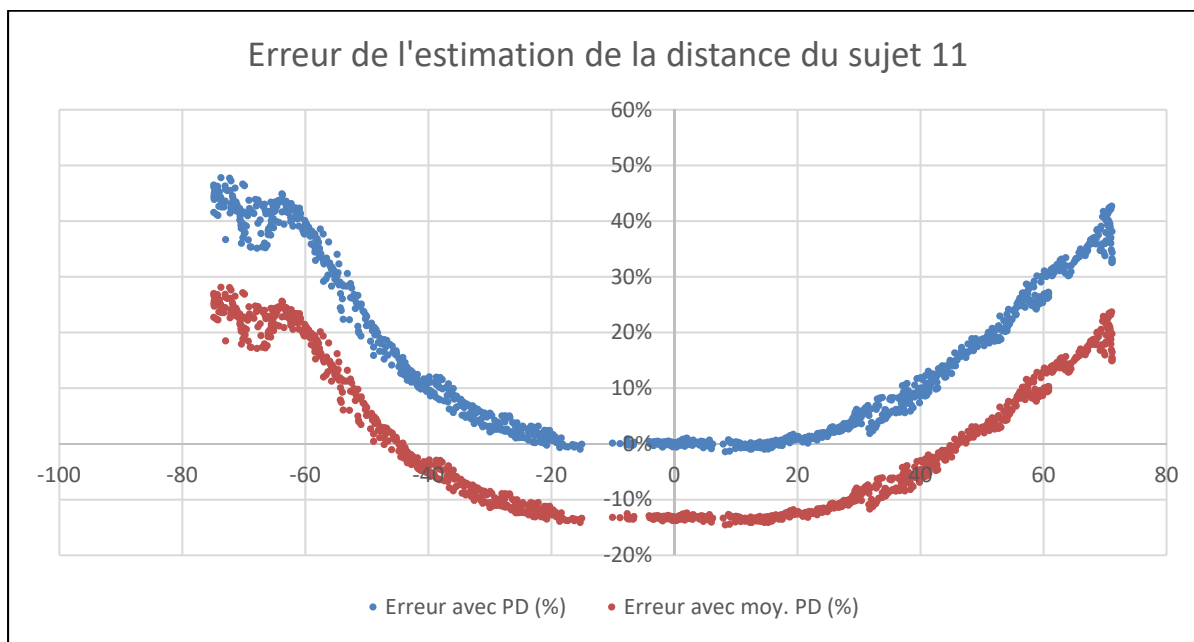


Figure 4.12 Comparatif de l'erreur d'estimation de distance entre **PD** et **PD** moyen

4.2.4 Dimension du sujet

Afin d'obtenir les différentes mesures anthropométriques expérimentales, Meshlab (Cignoni et al., 2011) a été utilisé en calculant la distance entre les différents sommets représentant la mesure désirée. Ceci a permis de s'assurer de toujours comparer les mêmes sommets et donc, les mêmes mesures anthropométriques entre chaque avatar. Ces mesures sont par la suite comparées aux mesures théoriques prises lors du protocole d'expérimentation.

Le Tableau 4.6 démontre l'erreur normalisée entre la mesure effectuée sur l'avatar à échelle estimée A_e et la mesure réelle. En analysant les différentes mesures, nous pouvons constater que l'idée de la méthodologie fonctionne. Il est possible d'obtenir des mesures et une représentation numérique tridimensionnelle d'un sujet à partir d'une seule image. Par exemple, l'erreur maximale sur les 10 sujets est de 7,69 % , soit 12.69 cm sur hauteur de 1.65 m. (Deák et al., 2017) réussit à avoir une erreur maximum de 10 %. Cependant, il est important de noter que notre solution ne fournit pas seulement la hauteur, mais un modèle à trois dimensions qui peut être déplacé et modifié. Il est aussi possible d'observer qu'il n'y a aucune mesure qui fut parfaite. Une erreur était toujours présente sur toute estimation. Intéressamment, il y a aussi des mesures telles que le poignet, l'avant-bras, et le biceps, qui ont un très bas écart type, mais une erreur moyenne de plus de 1 %. Ceci semble donc indiquer que dans la position dans dont les individus ont été capturés, l'algorithme SMPLify-X à une erreur sur l'ajustement de l'avatar qui est constant.

Tableau 4.6 Erreur d'estimation de mesure anthropométrique

	MOYENNE	ÉCART TYPE.	MIN	MAX
GRANDEUR	3.59 %	0.47 %	0.97 %	7.69 %
TÊTE	2.24 %	0.83 %	0.34 %	4.61 %
COUP	2.49 %	0.49 %	0.47 %	5.86 %
BASE DU COUP	1.71 %	1.17 %	0.79 %	6.19 %
ÉPAULE	1.97 %	1.03 %	0.64 %	7.23 %
POITRINE	1.46 %	1.81 %	1.21 %	5.83 %
TAILLE	3.20 %	0.89 %	0.12 %	6.57 %
HANCHES	2.54 %	1.23 %	0.10 %	5.64 %
BASSIN	1.72 %	0.19 %	0.03 %	4.19 %
BICEPS	1.96 %	0.04 %	0.02 %	5.85 %
AVANT-BRAS	2.56 %	0.05 %	0.20 %	6.13 %
POIGNET	3.20 %	0.03 %	0.33 %	5.23 %
CUISSE	3.01 %	0.59 %	0.74 %	5.43 %
TIBIA	2.27 %	0.75 %	0.43 %	6.19 %
CHEVILLE	3.35 %	0.89 %	0.99 %	6.02 %

L'erreur majeure de notre solution proposée est le repositionnement du modèle SMPLify-X sur l'image. Comme identifié dans la section 4.1, certaines caractéristiques morphologiques

n'ont pas été respectées par le modèle. Par exemple, sur le sujet 2, les cuisses, le mollet, la largeur de la tête, la largeur des bras, sur l'avatar était légèrement plus grande que sur le sujet réel. Ceci peut être observé via la Figure 4.13. Bien que ces altérations soient très mineures, elles entraînent une déviation du résultat. Plus la déviation est grande, plus l'erreur est grande.

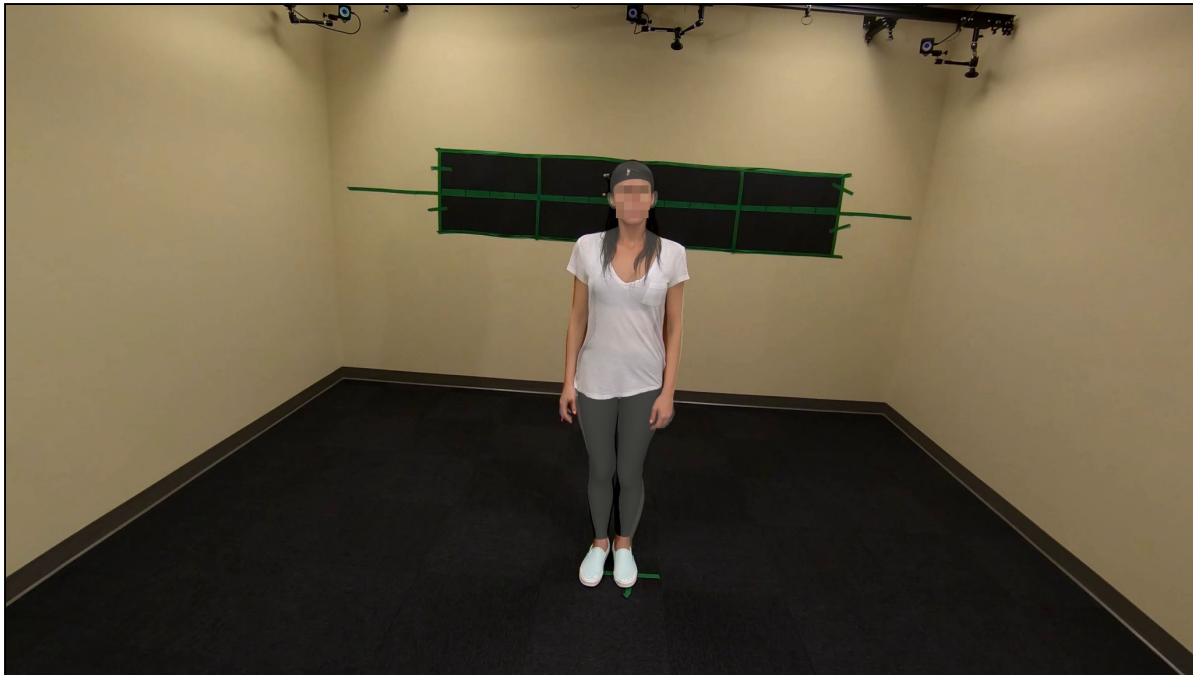


Figure 4.13 Avatar sur sujet 2

Il est important de noter que les mesures théoriques ont été effectuées sur les sujets en retirant ses vêtements lorsque cela était possible. Il y a une erreur humaine potentielle dans la mesure. Une comparaison plus efficace aurait été d'avoir une numérisation complète du corps humain. Cette numérisation aurait pu être utilisée comme référence et comparée au corps estimé pour obtenir une comparaison du corps entier au lieu d'une quantité limitée de mesures. Cela aurait également permis d'éliminer le facteur humain de l'équation. En revanche, dû aux limites d'équipement ainsi qu'à la difficulté de recrutement, cette option n'a pas été prise.

Il est aussi important de mentionner que les mesures ont été faites à la main via MeshLab. Il y a potentiellement une erreur humaine qui s'est infiltrée dans ce processus. Afin de remédier à

cette problématique et ayant une numérisation 3D théorique du sujet, il aurait été possible de développer un outil de comparaison de points précis sur le modèle.

4.3 Estimation de la distance inter pupillaire

Lors du traitement de données, il fut observé qu'une simple double intégration de l'accélération en plus de l'utilisation du gyroscope dans la caméra RealSense était problématique. Tel qu'expliqué dans (Muset & Emerich, 2012), puisque ceci est une double intégration, l'erreur sur l'accélération est aussi sous double intégration. La caméra RealSense a une trop grande erreur afin d'être utilisée directement sans filtre ou autre post-traitement. Puisque cela sort de sujet de cette recherche, la différence de la distance théorique (ΔD) a été obtenue en utilisant le Mocap afin de confirmer l'équation (12).

La Tableau 4.7 compare l'écart pupillaire mesuré à celui estimé. 3 mesures et 3 estimations ont été utilisées afin d'obtenir le résultat moyen. Cependant, la précision de notre outil de mesure (règle) n'était que de l'ordre du millimètre. Conséquemment, lors de la mesure de la distance pupillaire, la précision était limitée à un demi-millimètre tandis que l'estimation pouvait aller dans la précision décimale puisqu'il s'agit d'une équation algorithmique.

Tableau 4.7 PD mesuré comparé à l'estimé à ΔD de 2m

Sujet	PD Mesuré (mm)	Estimé (mm)	Erreur (mm)
1	62.5	62.8	0.3
2	60.5	60.3	-0.2
3	65.0	65.1	0.1
4	65.5	65.6	0.1
5	57.0	57.0	0.0
6	72.5	72.5	0.0
7	68.5	68.6	0.1
8	67.0	67.0	0.0
9	72.5	72.5	0.0
10	66.5	66.7	0.2

En comparant l'écart pupillaire mesuré à celui estimé, il est possible d'identifier que l'estimation est toujours dans la précision de l'outil, soit 0,5mm. Ceci montre qu'il est possible d'utiliser une différence de distance afin de déterminer l'écart pupillaire. Cependant, il serait intéressant de savoir quelle est la différence de distance minimale pour que cet algorithme fonctionne. La Figure 4.14 nous permet de repérer cette limite en illustrant l'erreur de l'écart pupillaire évalué en utilisant différentes distances ΔD . L'écart pupillaire de référence est celui estimé dans le Tableau 4.7. Le Tableau 4.8 et Tableau 4.9 nous permet d'analyser le graphique sur des intervalles plus détaillés. Le point de départ (0) sur l'axe horizontale est lorsque la caméra est à 2m du sujet. Lorsque la caméra se rapproche du sujet, les valeurs de ΔD sont donc négative.

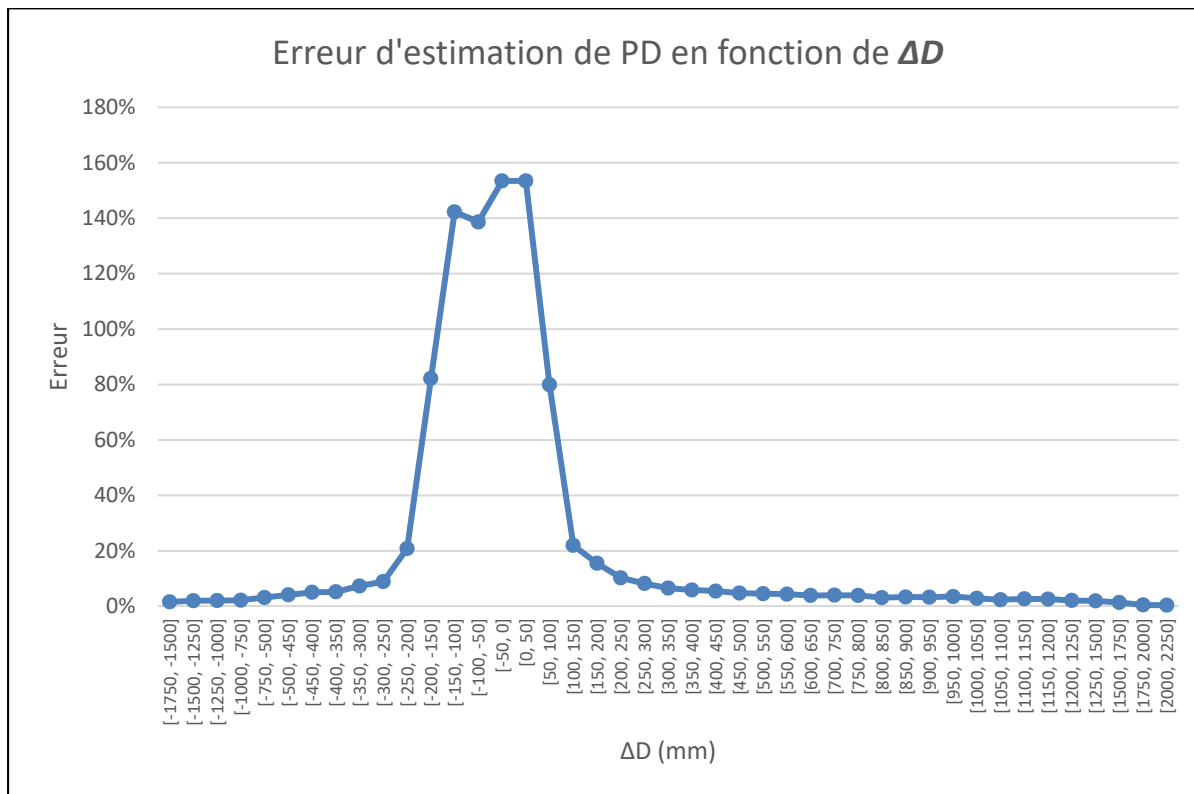


Figure 4.14 Erreur d'estimation de PD en fonction de ΔD

Tableau 4.8 Erreur d'estimation de ***PD*** en fonction de ***ΔD***

Intervalle de ΔD (mm)	Nb de points	Moyenne	σ	Min	Max
[-1750, -1500]	2151	1.41 %	0.85 %	0.00 %	4.54 %
[-1500, -1250]	1712	1.85 %	0.83 %	0.02 %	5.49 %
[-1250, -1000]	1468	1.86 %	1.05 %	0.01 %	6.90 %
[-1000, -750]	1292	2.01 %	1.48 %	0.01 %	8.09 %
[-750, -500]	1331	2.99 %	2.26 %	0.00 %	17.62 %
[-500, -450]	263	3.94 %	3.26 %	0.01 %	20.91 %
[-450, -400]	262	4.85 %	3.82 %	0.05 %	22.34 %
[-400, -350]	267	5.04 %	4.00 %	0.03 %	20.74 %
[-350, -300]	284	7.14 %	4.95 %	0.02 %	28.95 %
[-300, -250]	296	8.74 %	7.25 %	0.01 %	49.36 %
[-250, -200]	358	20.65 %	124.68 %	0.01 %	2043.47 %
[-200, -150]	365	81.96 %	828.90 %	0.05 %	15792.68 %
[-150, -100]	1014	142.06 %	885.07 %	0.12 %	17421.35 %
[-100, -50]	464	138.44 %	237.07 %	0.09 %	4211.63 %
[-50, 0]	790	153.22 %	404.41 %	0.67 %	7383.25 %
[0, 50]	1097	153.26 %	1103.69 %	0.54 %	33624.06 %
[50, 100]	249	79.71 %	268.21 %	0.10 %	2166.78 %
[100, 150]	272	21.78 %	33.82 %	0.36 %	397.07 %
[150, 200]	227	15.32 %	13.86 %	0.17 %	100.50 %
[200, 250]	211	10.10 %	8.66 %	0.08 %	39.71 %
[250, 300]	194	8.05 %	7.18 %	0.08 %	29.81 %
[300, 350]	199	6.37 %	6.31 %	0.01 %	26.39 %
[350, 400]	187	5.71 %	5.04 %	0.07 %	17.93 %
[400, 450]	184	5.29 %	4.48 %	0.02 %	16.19 %
[450, 500]	183	4.57 %	3.90 %	0.05 %	13.80 %
[500, 550]	176	4.36 %	3.74 %	0.12 %	14.59 %
[550, 600]	170	4.21 %	3.20 %	0.15 %	12.25 %
[600, 650]	189	3.70 %	2.91 %	0.06 %	10.98 %
[700, 750]	178	3.83 %	2.38 %	0.21 %	10.49 %
[750, 800]	185	3.75 %	2.15 %	0.45 %	9.55 %
[800, 850]	196	2.92 %	1.73 %	0.22 %	9.47 %
[850, 900]	179	3.17 %	1.78 %	0.32 %	8.83 %
[900, 950]	154	3.11 %	1.95 %	0.02 %	9.29 %
[950, 1000]	168	3.34 %	1.77 %	0.04 %	9.59 %
[1000, 1050]	181	2.69 %	1.57 %	0.06 %	8.10 %
[1050, 1100]	195	2.19 %	1.53 %	0.09 %	8.43 %

Tableau 4.9 Erreur d'estimation de PD en fonction de ΔD - suite

[1100, 1150]	204	2.50 %	1.66 %	0.23 %	9.46 %
[1150, 1200]	205	2.42 %	1.44 %	0.12 %	7.98 %
[1200, 1250]	185	1.91 %	1.24 %	0.02 %	6.34 %
[1250, 1500]	807	1.76 %	1.61 %	0.00 %	11.50 %
[1500, 1750]	1293	1.19 %	0.95 %	0.00 %	6.53 %
[1750, 2000]	805	0.31 %	0.90 %	0.00 %	5.47 %
[2000, 2250]	166	0.21 %	1.05 %	0.01 %	4.51 %

Il est possible de noter qu'une distance minimale de 300 mm est nécessaire pour obtenir un résultat qui n'est pas sur la courbe de changement abrupte. Plus spécifiquement, elle donne un résultat inférieur à 7%. Appliqué à l'écart pupillaire moyen (62,85 mm), cela représente une erreur de 4,4 mm. Bien que minime, une telle erreur aurait un impact important sur le calcul de la distance si elle devait être exploitée dans l'équation (9). Par exemple, à 2 mètres, une erreur de 7% représente 140 mm. En continuant d'augmenter l'intervalle de ΔD , l'erreur réduit graduellement. La distance totale couverte par l'expérience ne s'est pas stabilisée. En autres mots, il serait possible d'augmenter ΔD et peut-être atteindre de meilleurs résultats. Les meilleurs résultats ont été obtenus dans l'intervalle de distance la plus éloignée [2000mm, 2250mm] avec une erreur moyenne de $0,41 \pm 1,05\%$. Cela correspond à une erreur de 0,26 mm sur l'écart pupillaire ce qui équivaut à une erreur de 8,2 mm sur une distance de 2 mètres en utilisant l'équation (9).

Le défi de cette expérience est que nous ne disposions pas d'un outil de très haute précision pour mesurer l'écart pupillaire au micromètre et que nous avons donc dû utiliser l'estimation initiale comme référence aux résultats. Il est fort probable que cette référence n'était pas parfaitement alignée avec la mesure de la vérité terrain aux micromètres et a donc introduit une erreur. Ceci pourrait expliquer pourquoi nous avons encore une erreur. En outre, comme expliqué dans la section 4.2.3 ainsi que 4.2.2, les marqueurs PD_p ne sont pas ajustés avec la rotation de la tête. Bien que la rotation soit très minime pendant la deuxième partie de l'expérience, il y en avait encore une. Par conséquent, PD_p comporte une erreur qui, à son tour, injecte une erreur dans l'estimation.

CHAPITRE 5

CONSIDÉRATIONS ÉTHIQUES

La première préoccupation éthique autour de cette étude était de s'assurer que l'algorithme proposé n'était pas biaisé. La solution proposée est basée sur une équation mathématique et une projection géométrique. Ainsi, elle ne fait pas la différence entre les différentes couleurs de peau, ethnies ou tailles. Cependant, les outils utilisés pour estimer les marqueurs faciaux et pour repositionner le modèle sur le sujet dans une image sont tous deux des réseaux neuronaux entraînés sur des bases de données. Ces bases de données peuvent poser un problème éthique si elles ne contiennent pas un ensemble de données représentatif de la diversité humaine mondiale. Cela créerait statistiquement une intelligence artificielle qui serait encline à mieux travailler sur le type spécifique d'humains présents dans l'ensemble de données d'entraînement.

Une autre question éthique soulevée dans cette étude concerne la création de l'ensemble de données utilisé pour valider la théorie. La procédure ainsi que les données devaient respecter des règles éthiques. Celles-ci ont été garanties et approuvées par le comité d'éthique de l'ÉTS démontré dans l'ANNEXE III.

Une autre préoccupation éthique que ce projet soulève est l'obtention de mesures anthropométrique d'un individu sans son autorisation. Bien qu'en espace public il est autorisé de filmer dans la majorité des pays, permettre d'obtenir de telles informations sur des personnes sans leur consentement soulève la question jusqu'où permettrons-nous d'utiliser une caméra en public. De plus, comment peut-on contrôler les informations extrapolées à la suite du traitement d'image. Par exemple, ces informations pourraient être utilisées à des fins commerciales ou d'études scientifiques. Combiné avec un réseau de neurones, ils pourraient potentiellement aider à l'identification d'individus. Il est important de comprendre que plus le traitement d'image avance, plus nous ouvrons des possibilités d'avancement pour le bien et le mal. Tout dépend de l'utilisateur de cet algorithme.

CONCLUSION

Ce mémoire propose un processus qui recrée une représentation numérique tridimensionnelle de l'être humain à partir d'une seule image et de son écart pupillaire. Un tel avatar du sujet est une étape importante vers l'estimation précise et stable des mesures anthropométriques à partir d'une seule image puisque les mesures peuvent être effectuées directement sur l'avatar. Le mémoire caractérise premièrement les limitations de l'estimation de la rotation de la tête de OpenFace (Baltrusaitis et al., 2018) et qu'après 30° de rotation, l'estimation perd son efficacité. Par la suite, une amélioration est proposée sur l'estimation de la distance d'un humain à une caméra qui utilise la calibration de la caméra et la distance inter pupillaire en prenant compte de la rotation de la tête. Ceci a permis de générer une erreur moyenne normalisée de 3.16% lors d'une rotation de 30° contrairement à la méthode antérieure (Deák et al., 2017) qui obtenait 16.19%. Cette recherche exploite ensuite la distance estimée en vue de calculer un facteur d'échelle qui est ensuite appliqué à un modèle SMPL-X (Pavlakos et al., 2019) normalisé pour obtenir un modèle à l'échelle ce qui n'a pas été fait dans les recherches précédentes. Cette technique testée sur 10 sujets permet d'améliorer l'erreur maximale obtenue dans (BenAbdelkader & Yacoob, 2008; Guan, 2009; Wang, Wang, Lin, & Yuille, 2018) à seulement 7.69%. Elle propose également un algorithme qui utilise la géométrie de projection en perspective pour déduire la distance inter pupillaire d'un sujet à partir de 2 images afin d'augmenter la précision de la méthodologie proposée. Cette méthodologie obtient une erreur moyenne normalisée de 0.21%. Une partie des avancées de ce mémoire ont été présentées et publiées lors de la conférence IEEE MMSP 2021 (Yuen & Vazquez, 2021). Ce mémoire est la première étape d'une recherche plus poussée sur l'utilisation de l'écart pupillaire pour estimer les dimensions 3D réelles.

Il y a beaucoup d'améliorations potentielles dans cette recherche. Tout d'abord, on pourrait étudier la raison pour laquelle la distance pupillaire dans l'image ajustée (PD_p) contient toujours une erreur croissante lorsque la rotation de la tête augmente. Cela pourrait être simplement dû au fait que la rotation de la tête n'est pas prise en compte dans les centres pupillaires estimés avec les marqueurs. Cela permettrait d'obtenir une solution encore plus

stabilisée sur l'angle croissant de la rotation de la tête ce qui augmenterait la précision du facteur d'échelle de l'avatar. L'estimation de la distance pupillaire à partir de deux images serait aussi améliorée.

Trouver un moyen d'utiliser des images non calibrées serait également un excellent complément à cette méthodologie. Ceci permettrait de prendre n'importe quelle image et de déterminer automatiquement les mesures anthropométriques d'un individu présent, peu importe la source de celle-ci. Ceci permettrait aussi d'ajouter toutes les bases de données dont la calibration de la caméra n'était pas fournie afin de vérifier et valider la méthodologie.

Une autre voie d'amélioration de cette mémoire est d'améliorer SMPLify-X qui ajuste les paramètres du SMPL-X. Avoir un meilleur modèle d'avatar, qui ne soit ni plus grand ni plus petit que le sujet, augmenterait la précision de l'estimation.

Il y a aussi l'automatisation de l'estimation de la distance pupillaire utilisant le gyroscope et l'accéléromètre d'un appareil mobile qui mérite d'être investiguée. Lors du travail de ce mémoire une tentative fut faite et à échouée. N'étant pas la problématique que tente de résoudre ce document, cette voie n'a pas plus été investiguée malgré son potentielle et son utilité.

Pour analyser plus en profondeur l'efficacité des méthodologies proposées, une base de données plus importante serait nécessaire. Le fait de disposer d'une numérisation du corps entier dans une telle base de données permettrait également d'obtenir une comparaison plus représentative de l'avatar estimé par rapport à l'avatar théorique. Il serait également intéressant d'observer à quelle distance maximale l'algorithme fonctionne. Cela n'a pas été possible à déterminer puisqu'à la distance maximale de notre jeu de données, les algorithmes se comportent aussi bien sur une erreur normalisée.

ANNEXE I

DÉMARCHE AFIN D'OBTENIR UNE MESURE À PARTIR DE DEUX IMAGES ET DEUX DISTANCES CONNUES

$$D_{s_1} = \frac{f \cdot \widehat{PD} \cdot I_w}{PD_{p_1} \cdot w_r}$$

$$D_{s_2} = \frac{f \cdot \widehat{PD} \cdot I_w}{PD_{p_2} \cdot w_r}$$

$$D_{s_2} - D_{s_1} = \frac{f \cdot \widehat{PD} \cdot I_w}{PD_{p_2} \cdot w_r} - \frac{f \cdot \widehat{PD} \cdot I_w}{PD_{p_1} \cdot w_r}$$

$$\Delta D_s = \frac{f \cdot \widehat{PD} \cdot I_w \cdot PD_{p_1}}{PD_{p_2} \cdot w_r \cdot PD_{p_1}} - \frac{f \cdot \widehat{PD} \cdot I_w \cdot PD_{p_2}}{PD_{p_1} \cdot w_r \cdot PD_{p_2}}$$

$$\Delta D_s = \left(\frac{f \cdot I_w}{w_r} \right) \cdot \left(\frac{\widehat{PD} \cdot PD_{p_1}}{PD_{p_1} \cdot PD_{p_2}} - \frac{\widehat{PD} \cdot PD_{p_2}}{PD_{p_1} \cdot PD_{p_2}} \right)$$

$$\frac{\Delta D_s \cdot w_r}{f \cdot I_w} = \left(\frac{\widehat{PD} \cdot PD_{p_1}}{PD_{p_1} \cdot PD_{p_2}} - \frac{\widehat{PD} \cdot PD_{p_2}}{PD_{p_1} \cdot PD_{p_2}} \right)$$

$$\frac{\Delta D_s \cdot w_r}{f \cdot I_w} = \widehat{PD} \cdot \left(\frac{PD_{p_1}}{PD_{p_1} \cdot PD_{p_2}} - \frac{PD_{p_2}}{PD_{p_1} \cdot PD_{p_2}} \right)$$

$$\widehat{PD} = \frac{\Delta D_s \cdot w_r}{f \cdot I_w} \cdot \left(\frac{PD_{p_1} \cdot PD_{p_2}}{PD_{p_1} - PD_{p_2}} \right)$$

ANNEXE II


GRILLE DE MESURE D'UN SUJET

Tableau-A III-1 Grille de mesure d'un sujet

#	Sujet	Mesure (mm)
	Date	
	Sujet	
	Distance inter pupillaire	
	Grandeur	
Corp		
1	Tête	
2	Coup	
3	Base du coup	
4	Épaule	
5	Poitrine	
6	Taille	
7	Hanches	
8	Bassin	
Bras droite		
9	Biceps	
10	Avant-Bras	
11	Poignet	
Bras gauche		
12	Biceps	
13	Avant-Bras	
14	Poignet	
Jambe droite		
15	Cuisse	
16	Tibia	
17	Cheville	
Jambe gauche		
18	Cuisse	
19	Tibia	
20	Cheville	

ANNEXE III

PROTOCOL D'ACQUISITION DE DONNÉES



Projet H20190101

FORMULAIRE D'INFORMATION ET DE CONSENTEMENT

TITRE DU PROJET DE RECHERCHE

Estimation de mesures anthropométriques de l'être humain avec une caméra vidéo

CHERCHEUR PRINCIPAL

Carlos Vazquez, professeur au département de LOG-TI - École de technologie supérieure (ÉTS)

ÉTUDIANT

Michael Buron Yuen, Étudiant en maîtrise, Département LOG-TI, ÉTS

INTRODUCTION

Nous vous invitons à participer à un projet de recherche. Cependant, avant d'accepter de participer à ce projet et de signer ce formulaire d'information et de consentement, veuillez prendre le temps de lire, de comprendre et de considérer attentivement les renseignements qui suivent.

Ce formulaire peut contenir des mots que vous ne comprenez pas. Nous vous invitons à poser toutes les questions que vous jugerez utiles au chercheur responsable de ce projet ou à un membre de l'équipe de recherche, et à leur demander de vous expliquer tout mot ou renseignement qui n'est pas clair.

NATURE ET OBJECTIFS DU PROJET DE RECHERCHE

L'estimation de la position d'un être humain dans une image est un sujet présentement en grande évolution. Ce projet tente de pousser le sujet encore plus loin en estimant différentes mesures d'un être humain à l'aide d'une seule caméra de couleur. Pour ce faire, l'algorithme utilise la distance entre le centre des yeux (distance inter-pupillaire) ainsi que la rotation de la tête afin d'estimer la distance à laquelle le sujet se retrouve par rapport à la caméra. Ainsi, cette expérience tente d'évaluer l'efficacité de l'algorithme en comparant les données théoriques aux données pratiques.

Formulaire d'information et de consentement – Version 15 avril 2019

p.1/5

Figure A IV-1 Protocol d'acquisition de données

PARTICIPATION DEMANDÉE

Ce projet de recherche comporte 3 étapes réparties sur une séance d'environ 30 minutes.

1. Installation du matériel et prise de mesures (Durée : environ 15 min)

Si vous répondez aux critères d'éligibilité et si vous consentez à participer au projet, nous vous demanderons dans un premier temps de mettre un casque sur lequel nous avons installé un laser pour mesurer l'orientation de la tête et des capteurs inertiels pour mesurer l'accélération et l'orientation du déplacement de la tête.

Une fois le casque installé, nous mesurerons manuellement, à l'aide d'une règle à mesurer, votre distance inter-pupillaire. Finalement, nous mesurerons différentes parties de votre corps (longueur du bras, hauteur, etc.). Puis, nous prendrons deux photos de votre visage. Ces photos nous permettront de mesurer de façon automatique la distance inter-pupillaire.

2. Mesure de la rotation de la tête et de la distance par rapport à la caméra (Durée : environ 15 min)

Dans cette étape, nous vous demanderons de faire une rotation lente de la tête de gauche à droite et de droite à gauche de façon à parcourir la totalité du champ de vision possible tout en gardant votre corps immobile. Pendant ce temps, une caméra vidéo vous filmera. Ceci sera répété trois fois.

Finalement, vous vous déplacerez lentement sur une ligne droite d'une distance de 0.5m à 5m de la caméra. Comme pour l'étape précédente, vous serez filmé(e) et vous devrez répéter l'exercice trois fois.

3. Retrait du matériel et fin de la séance (Durée : environ 5 minutes)

Une fois l'ensemble des étapes précédentes terminées, il vous sera demandé de retirer le casque et la séance prendra fin.

RISQUES ET INCONVÉNIENTS ASSOCIÉS À LA PARTICIPATION

Outre le temps consacré à la participation, il existe un risque relié à l'utilisation du laser. Toutes les précautions seront prises pour que le laser ne pointe jamais dans votre direction durant les acquisitions.

AVANTAGES ET BÉNÉFICES

Il se peut que vous retiriez un bénéfice personnel de votre participation à ce projet de recherche, mais nous ne pouvons vous l'assurer. Par ailleurs, les résultats obtenus contribueront à l'avancement des connaissances scientifiques dans ce domaine de recherche.

CONFIDENTIALITÉ

Durant votre participation à ce projet de recherche, le chercheur responsable de ce projet ainsi que les membres de l'équipe de recherche recueilleront, dans un dossier de recherche, les renseignements vous concernant et nécessaires pour répondre aux objectifs scientifiques de ce projet de recherche.

Figure A IV-1 Protocol d'acquisition de données (Suite)

Ces renseignements peuvent comprendre votre nom, votre adresse courriel ainsi que les résultats de toutes les activités et procédures réalisées dans le cadre du projet.

Tous les renseignements recueillis demeureront confidentiels, dans les limites prévues par la loi. Afin de préserver votre identité et la confidentialité de vos renseignements, un numéro de code vous sera attribué. La clé du code reliant votre nom à votre dossier de recherche sera conservée par le chercheur responsable de ce projet de recherche.

Ces données de recherche seront conservées pendant au moins 10 ans par le chercheur responsable de ce projet de recherche.

Les données de recherche pourront être publiées ou faire l'objet de discussions scientifiques, mais il ne sera pas possible de vous identifier.

À des fins de surveillance, de contrôle, de protection, de sécurité, votre dossier de recherche pourra être consulté par une personne mandatée par des organismes réglementaires ainsi que par des représentants de l'organisme subventionnaire, de l'École de technologie supérieure ou du Comité d'éthique de la recherche. Ces personnes et ces organismes adhèrent à une politique de confidentialité.

Vous avez le droit de consulter votre dossier de recherche pour vérifier les renseignements recueillis et les faire rectifier au besoin.

UTILISATION SECONDAIRE DES DONNÉES DE RECHERCHE

Le chercheur responsable prévoit utiliser vos données de recherche pour réaliser d'autres projets de recherche dans le domaine de la vision par ordinateur.

Ces projets de recherche seront évalués et approuvés par le Comité d'éthique de la recherche de l'ÉTS avant leur réalisation. De plus, le Comité en assurera le suivi. Vos données de recherche seront conservées de façon sécuritaire sur des serveurs de l'ÉTS. Afin de préserver votre identité et la confidentialité de vos données de recherche, vous ne serez identifié(e) que par un numéro de code.

Vos données de recherche seront conservées aussi longtemps qu'elles peuvent avoir une utilité pour l'avancement des connaissances scientifiques. Lorsqu'elles n'auront plus d'utilité, vos données de recherche seront détruites. Par ailleurs, notez qu'en tout temps, vous pouvez demander la non-utilisation de vos données de recherche en vous adressant au chercheur responsable de ce projet de recherche.

Acceptez-vous que vos données de recherche soient utilisées à ces conditions? ☐ Oui ☐ Non

PARTICIPATION VOLONTAIRE ET DROIT DE RETRAIT

Votre participation à ce projet de recherche est volontaire. Vous êtes donc libre de refuser d'y participer. Vous pouvez également vous retirer de ce projet à n'importe quel moment, sans avoir à donner de raison, en informant l'équipe de recherche.

Si vous décidez de vous retirer du projet, les renseignements et données recueillis seront détruits, sauf si le chercheur ne possède plus la clé de correspondance permettant de faire le lien entre votre identité et votre code de recherche.

Le chercheur principal, le Comité d'éthique de la recherche ou l'établissement peuvent mettre fin à votre participation, sans votre consentement. Cela peut se produire si de nouvelles découvertes ou informations

Figure A IV-1 Protocol d'acquisition de données (Suite)

indiquent que votre participation au projet n'est plus dans votre intérêt, si vous ne respectez pas les consignes, ou encore s'il existe des raisons administratives d'abandonner le projet.

Toute nouvelle connaissance acquise durant le déroulement du projet qui pourrait avoir un impact sur votre décision de continuer à participer à ce projet vous sera communiquée rapidement.

COMPENSATION

Vous ne recevrez pas de compensation financière pour votre participation à ce projet de recherche.

POSSIBILITÉ DE COMMERCIALISATION

Votre participation au projet de recherche pourrait mener à la création de produits commerciaux qui pourraient être éventuellement protégés par voie de brevet ou autres droits de propriété intellectuelle. Cependant, dans un tel cas, vous ne pourrez en retirer aucun avantage financier.

VOS DROITS EN TANT QUE PARTICIPANT

En acceptant de participer à ce projet de recherche, vous ne renoncez à aucun de vos droits ni ne libérez le chercheur principal et l'établissement de leur responsabilité civile et professionnelle.

SUIVI ÉTHIQUE

Le comité d'éthique de la recherche de l'ÉTS a approuvé ce projet de recherche et en assure le suivi. De plus, il approuvera au préalable toute révision et toute modification apportée au formulaire d'information et de consentement et au protocole de recherche.

PERSONNES-RESSOURCES

Pour toute question en lien avec le projet de recherche, vous pouvez contacter le chercheur principal au carlos.vazquez@etsmtl.ca. Vous pouvez également contacter michael.buron.1@ens.etsmtl.ca.

Pour toute question en lien avec vos droits en tant que participant à la recherche, vous pouvez contacter la coordonnatrice du Comité d'éthique de la recherche de l'École de technologie supérieure en téléphonant au (514) 396-8800 poste 7807.

Figure A IV-1 Protocol d'acquisition de données (Suite)

CONSENTEMENT

Participant(e)

Je reconnais avoir lu le présent formulaire de consentement et avoir disposé de suffisamment de renseignements et du temps nécessaire pour prendre ma décision. Après réflexion, je consens volontairement à participer à ce projet de recherche, aux conditions énoncées.

Nom du(de la) participant(e)

Signature

Date

Personne qui obtient le consentement

J'ai expliqué au(à la) participant(e) tous les aspects pertinents de la recherche et j'ai répondu aux questions qu'il(elle) m'a posées.

☐ Je m'engage à remettre une copie du présent formulaire signé au(à la) participant(e).

Nom de la personne qui obtient le
consentement

Signature

Date

Signature et engagement du chercheur responsable de ce projet de recherche

Je certifie qu'on a expliqué au(à la) participant(e) le présent formulaire d'information et de consentement, que l'on a répondu aux questions qu'il(elle) avait.

Je m'engage, avec l'équipe de recherche, à respecter ce qui a été convenu au formulaire d'information et de consentement.

Chercheur responsable

Signature

Date



Projet H20190101
Formulaire d'information et de consentement – version 15 avril 2019

p.5/5

Figure A IV-1 Protocol d'acquisition de données (Suite)

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Al-Amri, S. S., Kalyankar, N., & Khamitkar, S. (2010). Image segmentation by using edge detection. *International journal on computer science and engineering*, 2(3), 804-807.
- Albiol, A., Monzo, D., Martin, A., Sastre, J., & Albiol, A. (2008). Face recognition using HOG-EBGM. *Pattern Recognition Letters*, 29(10), 1537-1543.
- Alldieck, T., Magnor, M., Xu, W., Theobalt, C., & Pons-Moll, G. (2018). Detailed Human Avatars from Monocular Video. Dans *2018 International Conference on 3D Vision (3DV)* (pp. 98-109). doi: 10.1109/3DV.2018.00022
- Alp Güler, R., Neverova, N., & Kokkinos, I. (2018). Densepose: Dense human pose estimation in the wild. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7297-7306).
- Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. Dans *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition* (pp. 3686-3693).
- Andriluka, M., Roth, S., & Schiele, B. (2010). Monocular 3d pose estimation and tracking by detection. Dans *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 623-630). IEEE.
- Arcoverde, E., Duarte, R., Barreto, R., Magalhaes, J., Bastos, C., Ing Ren, T., & Cavalcanti, G. (2014). Enhanced real-time head pose estimation system for mobile device. *Integrated Computer Aided Engineering*, 21, 281-293. doi: 10.3233/ICA-140462
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2012). 3D constrained local model for rigid and non-rigid facial tracking. Dans *2012 IEEE conference on computer vision and pattern recognition* (pp. 2610-2617). IEEE.
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. Dans *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.
- Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. Dans *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 59-66). IEEE.
- Banks, M. S., Read, J. C. A., Allison, R. S., & Watt, S. J. (2012). Stereoscopy and the Human Visual System. *SMPTE motion imaging journal*, 121(4), 24-43. doi: 10.5594/j18173.
Repéré à <https://pubmed.ncbi.nlm.nih.gov/23144596>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490636/>

- Barron, C., & Kakadiaris, I. A. (2000). Estimating anthropometry and pose from a single image. Dans *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)* (Vol. 1, pp. 669-676). IEEE.
- Belagiannis, V., & Zisserman, A. (2017). Recurrent human pose estimation. Dans *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 468-475). IEEE.
- BenAbdelkader, C., & Yacoob, Y. (2008). Statistical body height estimation from a single image. Dans *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 1-7). doi: 10.1109/AFGR.2008.4813453
- Bertoni, L., Kreiss, S., & Alahi, A. (2019). Monoloco: Monocular 3d pedestrian localization and uncertainty estimation. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6861-6871).
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008*.
- Caprile, B., & Torre, V. (1990). Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2), 127-139.
- Chen, W., & Jiang, B. C. (1991). 3-D camera calibration using vanishing point concept. *Pattern Recognition*, 24(1), 57-67.
- Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., & Urtasun, R. (2016). Monocular 3d object detection for autonomous driving. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2147-2156).
- Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2020). Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5386-5395).
- Chou, J., Solli, D. R., & Jalali, B. (2008). Real-time Spectroscopy with Sub-GHz Resolution using Amplified Dispersive Fourier Transformation. *arXiv preprint arXiv:0803.1654*.
- Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A. L., & Wang, X. (2017). Multi-context attention for human pose estimation. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1831-1840).

- Cignoni, P., Ranzuglia, G., Callieri, M., Corsini, M., Ganovelli, F., Pietroni, N., & Tarini, M. (2011). MeshLab.
- Coon, J. (2016). Systems and methods for obtaining a pupillary distance measurement using a mobile computing device: Google Patents.
- Corke, P. I. (2005). The Machine Vision Toolbox: a MATLAB toolbox for vision and vision-based control. *IEEE robotics & automation magazine*, 12(4), 16-25.
- De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. (2015). *IntraFace*. doi: 10.1109/FG.2015.7163082
- Deák, A., Kainz, O., Michalko, M., & Jakab, F. (2017). Estimation of human body height from uncalibrated image. Dans *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 1-4). IEEE.
- Desai, P. R., Desai, P. N., Ajmera, K. D., & Mehta, K. (2014). A review paper on oculus rift-a virtual reality headset. *arXiv preprint arXiv:1408.1173*.
- Dodgson, N. (2004). *Variation and extrema of human interpupillary distance* (Vol. 5291). doi: 10.1117/12.529999
- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38(2), 270-277. doi: 10.1037/0022-3514.38.2.270
- Faugeras, O. (1993). *Three-dimensional computer vision: a geometric viewpoint*. MIT press.
- Furtado, J. S., Liu, H. H., Lai, G., Lacheray, H., & Desouza-Coelho, J. (2019). Comparative analysis of optitrack motion capture systems. Dans *Advances in Motion Sensing and Control for Robotic Applications* (pp. 15-31). Springer.
- Govil, A., You, S., & Neumann, U. (2000). A video-based augmented reality golf simulator. Dans *Proceedings of the eighth ACM international conference on Multimedia* (pp. 489-490).
- Gremban, K. D., Thorpe, C. E., & Kanade, T. (1988). Geometric camera calibration using systems of linear equations. Dans *Proceedings. 1988 IEEE International Conference on Robotics and Automation* (pp. 562-567). IEEE.
- Guan, Y.-P. (2009). *Unsupervised human height estimation from a single image* (Vol. 2). doi: 10.4236/jbise.2009.26061
- Gudet, D. (2015). L'intégration du chemin chez les enfants de 5 à 12 ans. Dans.

- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.
- Haseeb, M. A., Guan, J., Ristic-Durrant, D., & Gräser, A. (2018). DisNet: a novel method for distance estimation from monocular camera. *10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18), IROS*.
- Hidalgo, G., Raaj, Y., Idrees, H., Xiang, D., Joo, H., Simon, T., & Sheikh, Y. (2019). Single-network whole-body pose estimation. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 6982-6991).
- Hirshberg, D. A., Loper, M., Rachlin, E., & Black, M. J. (2012). Coregistration: Simultaneous alignment and modeling of articulated 3D shape. Dans *European conference on computer vision* (pp. 242-255). Springer.
- Ji, X., Fang, Q., Dong, J., Shuai, Q., Jiang, W., & Zhou, X. (2020). A survey on monocular 3D human pose estimation. *Virtual Reality & Intelligent Hardware*, 2(6), 471-500.
- Jin, S., Xu, L., Xu, J., Wang, C., Liu, W., Qian, C., . . . Luo, P. (2020). Whole-Body Human Pose Estimation in the Wild. *arXiv preprint arXiv:2007.11858*.
- Johnson, S., & Everingham, M. (2010). Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. Dans *bmvc* (Vol. 2, pp. 5). Citeseer.
- Karlinsky, L., & Ullman, S. (2012). Using linking features in learning non-parametric part models. Dans *European Conference on Computer Vision* (pp. 326-339). Springer.
- Koh, C. J., & Kamiyama, K. (2019). Systems and methods for full body measurements extraction using multiple deep learning networks for body feature measurements: Google Patents.
- Kowalczyk, Z., & Merta, T. (2015). Evaluation of position estimation based on accelerometer data. Dans *2015 10th International Workshop on Robot Motion and Control (RoMoCo)* (pp. 246-251). IEEE.
- Kytö, M., Nuutinen, M., & Oittinen, P. (2011). *Method for measuring stereo camera depth accuracy based on stereoscopic vision* (Vol. 7864). doi: 10.1117/12.872015
- Lazova, V., Insafutdinov, E., & Pons-Moll, G. (2019). 360-degree textures of people in clothing from a single image. Dans *2019 International Conference on 3D Vision (3DV)* (pp. 643-653). IEEE.
- Li, M., Zhou, Z., Li, J., & Liu, X. (2018). Bottom-up pose estimation of multiple person with bounding box constraint. Dans *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 115-120). IEEE.

- Li, R., Xiu, Y., Saito, S., Huang, Z., Olszewski, K., & Li, H. (2020). Monocular Real-Time Volumetric Performance Capture. *arXiv preprint arXiv:2007.13988*.
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), Article 248. doi: 10.1145/2816795.2818013. Repéré à <https://doi.org/10.1145/2816795.2818013>
- Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics*, 21(4), 163-169.
- Ma, Q., Yang, J., Ranjan, A., Pujades, S., Pons-Moll, G., Tang, S., & Black, M. J. (2020). Learning to dress 3d people in generative clothing. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6469-6478).
- Martinez, B., & Valstar, M. (2016). Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition. Dans (pp. 63-100). doi: 10.1007/978-3-319-25958-1_4
- Meagher, D. (1982). Geometric modeling using octree encoding. *Computer graphics and image processing*, 19(2), 129-147.
- Muset, B., & Emerich, S. (2012). Distance measuring using accelerometer and gyroscope sensors. *Carpathian Journal of Electronic and Computer Engineering*, 5, 83.
- Nayar, S. K., & Gupta, M. (2012). Diffuse structured light. Dans *2012 IEEE International Conference on Computational Photography (ICCP)* (pp. 1-11). doi: 10.1109/ICCPHOT.2012.6215216
- Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. Dans *European conference on computer vision* (pp. 483-499). Springer.
- Norton, K., & Olds, T. (1996). *Anthropometrica: a textbook of body measurement for sports and health courses*. UNSW press.
- Pang, G., & Liu, H. (2001). Evaluation of a low-cost MEMS accelerometer for distance measurement. *Journal of Intelligent and Robotic Systems*, 30(3), 249-265.
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 10975-10985).
- Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., & Schiele, B. (2016). Deepcut: Joint subset partition and labeling for multi person pose estimation.

Dans *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4929-4937).

Pishchulin, L., Wuhler, S., Helten, T., Theobalt, C., & Schiele, B. (2017). Building statistical shape spaces for 3d human modeling. *Pattern Recognition*, 67, 276-286.

Pons-Moll, G., Romero, J., Mahmood, N., & Black, M. J. (2015). Dyna: a model of dynamic human shape in motion. *ACM Trans. Graph.*, 34(4), Article 120. doi: 10.1145/2766993. Repéré à <https://doi.org/10.1145/2766993>

Qi, W., Li, F., & Zhenzhong, L. (2010). Review on camera calibration. Dans *2010 Chinese Control and Decision Conference* (pp. 3354-3358). IEEE.

Robinette, K., Blackwell, S., Daanen, H., Boehmer, M., Fleming, S., Brill, T., . . . Burnside, D. (2002). Civilian american and european surface anthropometry resource final report. *AFRL-HE-WP-TR*.

Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2304-2314).

Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6), 1113-1133. doi: 10.1109/TPAMI.2014.2366127

Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006). A toolbox for easily calibrating omnidirectional cameras. Dans *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5695-5701). IEEE.

Seifert, K., & Camacho, O. (2007). Implementing positioning algorithms using accelerometers. *Freescale Semiconductor*, 1, 13.

Simek, K. (2013). Dissecting the camera matrix, part 3: The intrinsic matrix. URL <http://ksimek.github.io/2013/08/13/intrinsic>.

Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. Dans *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1145-1153).

Sturm, P. F., & Maybank, S. J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. Dans *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (Vol. 1, pp. 432-437). IEEE.

- Tang, J., & Li, J. (2020). End-to-End Monocular Range Estimation for Forward Collision Warning. *Sensors*, 20(20), 5941.
- Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 648-656).
- Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4), 323-344.
- Wang, C., Wang, Y., Lin, Z., & Yuille, A. (2018). Robust 3D Human Pose Estimation from Single Images or Video Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. doi: 10.1109/TPAMI.2018.2828427
- Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. Dans *Advances in Neural Information Processing Systems* (pp. 492-502).
- Xuanwu, Y., Guijin, W., Chenbo, S., & Qingmin, L. (2014). Efficient active depth sensing by laser speckle projection system. *Optical Engineering*, 53(1), 1-10. doi: 10.1117/1.OE.53.1.013105. Repéré à <https://doi.org/10.1117/1.OE.53.1.013105>
- Yang, S. X., Larsen, P. K., Alkjær, T., Juul-Kristensen, B., Simonsen, E. B., & Lynnerup, N. (2014). Height estimations based on eye measurements throughout a gait cycle. *Forensic science international*, 236, 170-174.
- Yin, X., Wang, G., Shi, C., & Liao, Q. (2014). Efficient active depth sensing by laser speckle projection system. *Optical Engineering*, 53(1), 013105. Repéré à <https://doi.org/10.1117/1.OE.53.1.013105>
- Yu, H., & Oh, J. (2021). Anchor Distance for 3D Multi-Object Distance Estimation from 2D Single Shot. *IEEE Robotics and Automation Letters*, 6(2), 3405-3412.
- Yuen, M. B., & Vazquez, C. (2021). Human Subject Distance Estimation Using the Pupillary Distance and Head Orientation. Dans *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.
- Zhang, Y., Li, Y., Zhao, M., & Yu, X. (2020). A regional regression network for monocular object distance estimation. Dans *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.

- Zhang, Z. (2002). Method and system for calibrating digital cameras: Google Patents.
- Zheng, Z., Yu, T., Wei, Y., Dai, Q., & Liu, Y. (2019). Deephuman: 3d human reconstruction from a single image. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 7739-7749).
- Zhu, J., & Fang, Y. (2019). Learning object-specific distance from a monocular image. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3839-3848).
- Al-Amri, S. S., Kalyankar, N., & Khamitkar, S. (2010). Image segmentation by using edge detection. *International journal on computer science and engineering*, 2(3), 804-807.
- Albiol, A., Monzo, D., Martin, A., Sastre, J., & Albiol, A. (2008). Face recognition using HOG-EBGM. *Pattern Recognition Letters*, 29(10), 1537-1543.
- Alldieck, T., Magnor, M., Xu, W., Theobalt, C., & Pons-Moll, G. (2018). Detailed Human Avatars from Monocular Video. Dans *2018 International Conference on 3D Vision (3DV)* (pp. 98-109). doi: 10.1109/3DV.2018.00022
- Alp Güler, R., Neverova, N., & Kokkinos, I. (2018). Densepose: Dense human pose estimation in the wild. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7297-7306).
- Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. Dans *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition* (pp. 3686-3693).
- Andriluka, M., Roth, S., & Schiele, B. (2010). Monocular 3d pose estimation and tracking by detection. Dans *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 623-630). IEEE.
- Arcoverde, E., Duarte, R., Barreto, R., Magalhaes, J., Bastos, C., Ing Ren, T., & Cavalcanti, G. (2014). Enhanced real-time head pose estimation system for mobile device. *Integrated Computer Aided Engineering*, 21, 281-293. doi: 10.3233/ICA-140462
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2012). 3D constrained local model for rigid and non-rigid facial tracking. Dans *2012 IEEE conference on computer vision and pattern recognition* (pp. 2610-2617). IEEE.
- Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. Dans *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.

- Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. Dans *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 59-66). IEEE.
- Banks, M. S., Read, J. C. A., Allison, R. S., & Watt, S. J. (2012). Stereoscopy and the Human Visual System. *SMPTE motion imaging journal*, 121(4), 24-43. doi: 10.5594/j18173. Repéré à <https://pubmed.ncbi.nlm.nih.gov/23144596>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3490636/>
- Barron, C., & Kakadiaris, I. A. (2000). Estimating anthropometry and pose from a single image. Dans *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)* (Vol. 1, pp. 669-676). IEEE.
- Belagiannis, V., & Zisserman, A. (2017). Recurrent human pose estimation. Dans *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 468-475). IEEE.
- BenAbdelkader, C., & Yacoob, Y. (2008). Statistical body height estimation from a single image. Dans *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition* (pp. 1-7). doi: 10.1109/AFGR.2008.4813453
- Bertoni, L., Kreiss, S., & Alahi, A. (2019). Monoloco: Monocular 3d pedestrian localization and uncertainty estimation. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6861-6871).
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *arXiv preprint arXiv:1812.08008*.
- Caprile, B., & Torre, V. (1990). Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2), 127-139.
- Chen, W., & Jiang, B. C. (1991). 3-D camera calibration using vanishing point concept. *Pattern Recognition*, 24(1), 57-67.
- Chen, X., Kundu, K., Zhang, Z., Ma, H., Fidler, S., & Urtasun, R. (2016). Monocular 3d object detection for autonomous driving. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2147-2156).
- Cheng, B., Xiao, B., Wang, J., Shi, H., Huang, T. S., & Zhang, L. (2020). Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5386-5395).

- Chou, J., Solli, D. R., & Jalali, B. (2008). Real-time Spectroscopy with Sub-GHz Resolution using Amplified Dispersive Fourier Transformation. *arXiv preprint arXiv:0803.1654*.
- Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A. L., & Wang, X. (2017). Multi-context attention for human pose estimation. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1831-1840).
- Cignoni, P., Ranzuglia, G., Callieri, M., Corsini, M., Ganovelli, F., Pietroni, N., & Tarini, M. (2011). MeshLab.
- Coon, J. (2016). Systems and methods for obtaining a pupillary distance measurement using a mobile computing device: Google Patents.
- Corke, P. I. (2005). The Machine Vision Toolbox: a MATLAB toolbox for vision and vision-based control. *IEEE robotics & automation magazine*, 12(4), 16-25.
- De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. (2015). *IntraFace*. doi: 10.1109/FG.2015.7163082
- Deák, A., Kainz, O., Michalko, M., & Jakab, F. (2017). Estimation of human body height from uncalibrated image. Dans *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 1-4). IEEE.
- Desai, P. R., Desai, P. N., Ajmera, K. D., & Mehta, K. (2014). A review paper on oculus rift-a virtual reality headset. *arXiv preprint arXiv:1408.1173*.
- Dodgson, N. (2004). *Variation and extrema of human interpupillary distance* (Vol. 5291). doi: 10.1117/12.529999
- Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology*, 38(2), 270-277. doi: 10.1037/0022-3514.38.2.270
- Faugeras, O. (1993). *Three-dimensional computer vision: a geometric viewpoint*. MIT press.
- Furtado, J. S., Liu, H. H., Lai, G., Lacheray, H., & Desouza-Coelho, J. (2019). Comparative analysis of optitrack motion capture systems. Dans *Advances in Motion Sensing and Control for Robotic Applications* (pp. 15-31). Springer.
- Govil, A., You, S., & Neumann, U. (2000). A video-based augmented reality golf simulator. Dans *Proceedings of the eighth ACM international conference on Multimedia* (pp. 489-490).

- Gremban, K. D., Thorpe, C. E., & Kanade, T. (1988). Geometric camera calibration using systems of linear equations. Dans *Proceedings. 1988 IEEE International Conference on Robotics and Automation* (pp. 562-567). IEEE.
- Guan, Y.-P. (2009). *Unsupervised human height estimation from a single image* (Vol. 2). doi: 10.4236/jbise.2009.26061
- Gudet, D. (2015). L'intégration du chemin chez les enfants de 5 à 12 ans. Dans.
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.
- Haseeb, M. A., Guan, J., Ristic-Durrant, D., & Gräser, A. (2018). DisNet: a novel method for distance estimation from monocular camera. *10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18), IROS*.
- Hidalgo, G., Raaj, Y., Idrees, H., Xiang, D., Joo, H., Simon, T., & Sheikh, Y. (2019). Single-network whole-body pose estimation. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 6982-6991).
- Hirshberg, D. A., Loper, M., Rachlin, E., & Black, M. J. (2012). Coregistration: Simultaneous alignment and modeling of articulated 3D shape. Dans *European conference on computer vision* (pp. 242-255). Springer.
- Ji, X., Fang, Q., Dong, J., Shuai, Q., Jiang, W., & Zhou, X. (2020). A survey on monocular 3D human pose estimation. *Virtual Reality & Intelligent Hardware*, 2(6), 471-500.
- Jin, S., Xu, L., Xu, J., Wang, C., Liu, W., Qian, C., . . . Luo, P. (2020). Whole-Body Human Pose Estimation in the Wild. *arXiv preprint arXiv:2007.11858*.
- Johnson, S., & Everingham, M. (2010). Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation. Dans *bmvc* (Vol. 2, pp. 5). Citeseer.
- Karlinsky, L., & Ullman, S. (2012). Using linking features in learning non-parametric part models. Dans *European Conference on Computer Vision* (pp. 326-339). Springer.
- Koh, C. J., & Kamiyama, K. (2019). Systems and methods for full body measurements extraction using multiple deep learning networks for body feature measurements: Google Patents.
- Kowalczyk, Z., & Merta, T. (2015). Evaluation of position estimation based on accelerometer data. Dans *2015 10th International Workshop on Robot Motion and Control (RoMoCo)* (pp. 246-251). IEEE.

- Kytö, M., Nuutinen, M., & Oittinen, P. (2011). *Method for measuring stereo camera depth accuracy based on stereoscopic vision* (Vol. 7864). doi: 10.1117/12.872015
- Lazova, V., Insaftudinov, E., & Pons-Moll, G. (2019). 360-degree textures of people in clothing from a single image. Dans *2019 International Conference on 3D Vision (3DV)* (pp. 643-653). IEEE.
- Li, M., Zhou, Z., Li, J., & Liu, X. (2018). Bottom-up pose estimation of multiple person with bounding box constraint. Dans *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 115-120). IEEE.
- Li, R., Xiu, Y., Saito, S., Huang, Z., Olszewski, K., & Li, H. (2020). Monocular Real-Time Volumetric Performance Capture. *arXiv preprint arXiv:2007.13988*.
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: a skinned multi-person linear model. *ACM Trans. Graph.*, 34(6), Article 248. doi: 10.1145/2816795.2818013. Repéré à <https://doi.org/10.1145/2816795.2818013>
- Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics*, 21(4), 163-169.
- Ma, Q., Yang, J., Ranjan, A., Pujades, S., Pons-Moll, G., Tang, S., & Black, M. J. (2020). Learning to dress 3d people in generative clothing. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6469-6478).
- Martinez, B., & Valstar, M. (2016). Advances, Challenges, and Opportunities in Automatic Facial Expression Recognition. Dans (pp. 63-100). doi: 10.1007/978-3-319-25958-1_4
- Meagher, D. (1982). Geometric modeling using octree encoding. *Computer graphics and image processing*, 19(2), 129-147.
- Muset, B., & Emerich, S. (2012). Distance measuring using accelerometer and gyroscope sensors. *Carpathian Journal of Electronic and Computer Engineering*, 5, 83.
- Nayar, S. K., & Gupta, M. (2012). Diffuse structured light. Dans *2012 IEEE International Conference on Computational Photography (ICCP)* (pp. 1-11). doi: 10.1109/ICCPHOT.2012.6215216
- Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. Dans *European conference on computer vision* (pp. 483-499). Springer.
- Norton, K., & Olds, T. (1996). *Anthropometrica: a textbook of body measurement for sports and health courses*. UNSW press.

- Pang, G., & Liu, H. (2001). Evaluation of a low-cost MEMS accelerometer for distance measurement. *Journal of Intelligent and Robotic Systems*, 30(3), 249-265.
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 10975-10985).
- Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., & Schiele, B. (2016). Deepcut: Joint subset partition and labeling for multi person pose estimation. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4929-4937).
- Pishchulin, L., Wuhler, S., Helten, T., Theobalt, C., & Schiele, B. (2017). Building statistical shape spaces for 3d human modeling. *Pattern Recognition*, 67, 276-286.
- Pons-Moll, G., Romero, J., Mahmood, N., & Black, M. J. (2015). Dyna: a model of dynamic human shape in motion. *ACM Trans. Graph.*, 34(4), Article 120. doi: 10.1145/2766993. Repéré à <https://doi.org/10.1145/2766993>
- Qi, W., Li, F., & Zhenzhong, L. (2010). Review on camera calibration. Dans *2010 Chinese Control and Decision Conference* (pp. 3354-3358). IEEE.
- Raza, S. N., Rehman, H. R. u., Lee, S. G., & Choi, G. S. (2019). Artificial Intelligence based Camera Calibration. Dans *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 1564-1569). doi: 10.1109/IWCMC.2019.8766666
- Robinette, K., Blackwell, S., Daanen, H., Boehmer, M., Fleming, S., Brill, T., . . . Burnside, D. (2002). Civilian american and european surface anthropometry resource final report. *AFRL-HE-WP-TR*.
- Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., & Li, H. (2019). Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2304-2314).
- Sariyanidi, E., Gunes, H., & Cavallaro, A. (2015). Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6), 1113-1133. doi: 10.1109/TPAMI.2014.2366127
- Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006). A toolbox for easily calibrating omnidirectional cameras. Dans *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5695-5701). IEEE.

- Seifert, K., & Camacho, O. (2007). Implementing positioning algorithms using accelerometers. *Freescale Semiconductor*, 1, 13.
- Simek, K. (2013). Dissecting the camera matrix, part 3: The intrinsic matrix. URL <http://ksimek.github.io/2013/08/13/intrinsic>.
- Simon, T., Joo, H., Matthews, I., & Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. Dans *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 1145-1153).
- Sturm, P. F., & Maybank, S. J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. Dans *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)* (Vol. 1, pp. 432-437). IEEE.
- Tang, J., & Li, J. (2020). End-to-End Monocular Range Estimation for Forward Collision Warning. *Sensors*, 20(20), 5941.
- Tompson, J., Goroshin, R., Jain, A., LeCun, Y., & Bregler, C. (2015). Efficient object localization using convolutional networks. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 648-656).
- Tsai, R. (1987). A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4), 323-344.
- Wang, C., Wang, Y., Lin, Z., & Yuille, A. (2018). Robust 3D Human Pose Estimation from Single Images or Video Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. doi: 10.1109/TPAMI.2018.2828427
- Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. Dans *Advances in Neural Information Processing Systems* (pp. 492-502).
- Xuanwu, Y., Guijin, W., Chenbo, S., & Qingmin, L. (2014). Efficient active depth sensing by laser speckle projection system. *Optical Engineering*, 53(1), 1-10. doi: 10.1117/1.OE.53.1.013105. Repéré à <https://doi.org/10.1117/1.OE.53.1.013105>
- Yang, S. X., Larsen, P. K., Alkjær, T., Juul-Kristensen, B., Simonsen, E. B., & Lynnerup, N. (2014). Height estimations based on eye measurements throughout a gait cycle. *Forensic science international*, 236, 170-174.

- Yin, X., Wang, G., Shi, C., & Liao, Q. (2014). Efficient active depth sensing by laser speckle projection system. *Optical Engineering*, 53(1), 013105. Repéré à <https://doi.org/10.1117/1.OE.53.1.013105>
- Yu, H., & Oh, J. (2021). Anchor Distance for 3D Multi-Object Distance Estimation from 2D Single Shot. *IEEE Robotics and Automation Letters*, 6(2), 3405-3412.
- Yuen, M. B., & Vazquez, C. (2021). Human Subject Distance Estimation Using the Pupillary Distance and Head Orientation. Dans *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)* (pp. 1-6). IEEE.
- Zhang, Y., Li, Y., Zhao, M., & Yu, X. (2020). A regional regression network for monocular object distance estimation. Dans *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)* (pp. 1-6). IEEE.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.
- Zhang, Z. (2002). Method and system for calibrating digital cameras: Google Patents.
- Zheng, Z., Yu, T., Wei, Y., Dai, Q., & Liu, Y. (2019). Deephuman: 3d human reconstruction from a single image. Dans *Proceedings of the IEEE International Conference on Computer Vision* (pp. 7739-7749).
- Zhu, J., & Fang, Y. (2019). Learning object-specific distance from a monocular image. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3839-3848).
- Ajanohoun, J., Paquette, E., & Vázquez, C. (2021). Multi-view human model fitting using bone orientation constraint and joints triangulation. Dans *2021 IEEE International Conference on Image Processing (ICIP)* (pp. 1094-1098). IEEE.
- Li, R., Xiu, Y., Saito, S., Huang, Z., Olszewski, K., & Li, H. (2020). Monocular real-time volumetric performance capture. Dans *European Conference on Computer Vision* (pp. 49-67). Springer.
- Saito, S., Simon, T., Saragih, J., & Joo, H. (2020). Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 84-93).
- Xiu, Y., Yang, J., Tzionas, D., & Black, M. J. (2022). Icon: Implicit clothed humans obtained from normals. Dans *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13296-13306).

He, T., Xu, Y., Saito, S., Soatto, S., & Tung, T. (2021). ARCH++: Animation-ready clothed human reconstruction revisited. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 11046-11056).