

Transformée en ondelettes et IA pour la reconstruction d'un signal PPG en contact à partir de sa version sans contact

Frédéric BOUSEFSAF, Djamaledine DJELDJI, Yassine OUZAR, Choubeila MAAOUI, Alain PRUSKI

LCOMS, Université de Lorraine
7 rue Marconi, 57070 Metz, France
frederic.bousefsaf@univ-lorraine.fr

Résumé – La mesure de signaux photopléthysmographiques (PPG) sans contact est une technique de mesure non invasive permettant d'estimer un ensemble de fonctions vitales par analyse vidéo délivrée par caméra. Nous proposons, dans cet article, une méthode permettant de convertir un signal PPG mesuré par caméra en un signal PPG mesuré en contact (cPPG). L'objectif à plus long terme consistera à transformer le signal cPPG en signal de tension artérielle afin de proposer une chaîne de traitement permettant d'estimer la tension à partir d'une vidéo. La méthode que nous proposons dans cet article repose sur la transformée en ondelettes et sur des modèles d'IA modernes. Les résultats reflètent la pertinence de l'approche et montrent qu'une estimation de la pression artérielle à partir d'un signal PPG caméra converti en signal en contact est envisageable.

Abstract – Imaging photoplethysmography (iPPG) is an optical technique dedicated to the assessment of several vital functions using a simple camera. We here propose a method for converting iPPG to contact PPG (cPPG) signals for, in future works, translating this cPPG signal to blood pressure. This would allow remote measurement of blood pressure from video. The continuous wavelet transform of cPPG and iPPG signals and deep neural networks are employed in this study. The results exhibit good agreements towards several metrics, showing that the neural architectures properly estimated cPPG from iPPG signals through their CWT representations.

1 Introduction

Les recherches portant sur la mesure de signaux physiologiques par des technologies sans contact ont connu des avancées significatives ces dernières années [1]. La photopléthysmographie (PPG) est mesurable à distance en observant les fines fluctuations de la couleur de la peau d'une personne. Le domaine est en plein essor et est soutenu par un ensemble d'études [2]. Des méthodes issues de la vision par ordinateur, du traitement d'images et de l'intelligence artificielle (IA) ont été utilisées ou développées spécifiquement pour transformer avec fiabilité la vidéo d'entrée en paramètres biomédicaux. Ces méthodes reposent principalement sur des modèles neuronaux [3].

Les recherches dans ce domaine s'orientent désormais vers la mesure de nouveaux paramètres physiologiques tels que la tension artérielle [4]. La mesure de la tension par analyse vidéo est complexe et peu de travaux montrent sa faisabilité. Deux directions sont à l'étude : (i) la mesure du *pulse transit time*, paramètre admis comme étant corrélé avec la pression artérielle [5] ainsi que (ii) l'étude directe de l'onde PPG [4]. Les résultats de ces études sont mitigés.

Des tentatives d'utilisation de modèles d'IA ont récemment été proposées [6]. L'apprentissage des modèles est cependant contraint par les faibles quantités de données actuellement disponibles. L'apprentissage d'un modèle neuronal profond permettant d'estimer avec précision la pression artérielle par analyse vidéo est donc difficilement envisageable pour le moment. Nous avons récemment montré que les caractéristiques tempo-

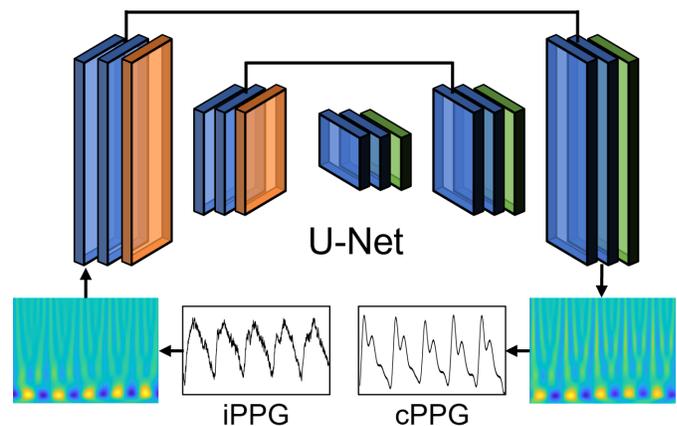


FIGURE 1 – Vue d'ensemble de la méthode proposée. Le signal iPPG est calculé à partir d'une analyse vidéo du visage de la personne. Sa représentation en ondelettes traverse le réseau U-Net. La transformée inverse de la représentation en ondelettes prédite permet de former le signal PPG en contact (cPPG) mesuré traditionnellement via un capteur placé sur le doigt.

relles, de courbure et de surface des signaux PPG évoluent de manière comparable entre les mesures caméra et les mesures en contact pris au doigt ou à l'oreille [7]. Ce point est important car il motive la présente étude. Nous partons de l'hypothèse qu'un signal PPG mesuré par caméra (imaging PPG, iPPG) peut être converti en un signal PPG en contact (cPPG) par le

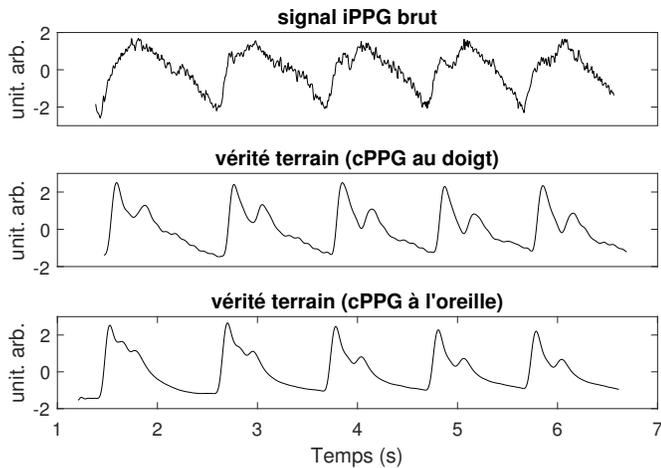


FIGURE 2 – Illustration de signaux extraits du participant # 1 durant une phase de maintien de respiration.

biais d'un modèle d'IA utilisant pour entrée la représentation en ondelettes continue des signaux. L'objectif final et à plus long terme consistera à transformer ce signal cPPG en signal de tension artérielle car ce champ est soutenu par une littérature plus mature, avec des études présentant des méthodes dont les performances respectent les standards internationaux [8].

2 Méthodes

2.1 Base de données et protocole expérimental

Les données utilisées pour apprendre les modèles neuronaux présentés en section 2.3 ont été présentées dans un article publié précédemment. 12 volontaires ont participé à l'étude. L'âge des participants est compris entre 20 et 35 ans. Ils ont été placés à environ 1 mètre d'une caméra rapide (125 fps). Les références ont été acquises grâce à des capteurs PPG en contact placés au doigt et à l'oreille. Deux essais de 60 secondes ont été proposés aux participants de l'étude. Premier test : nous demandons aux participants de rester au calme et de respirer normalement. Second test : il était demandé aux participants de retenir leur respiration autant que faire se peut, l'objectif étant de provoquer des variations physiologiques qui modifient la pression artérielle et impactent les signaux PPG enregistrés. Nous renvoyons le lecteur vers la publication originale pour plus de détails concernant la procédure et le matériel utilisé [7].

La base de données contient 724 signaux échantillonnés sur 256 points. Chaque signal contient 5 ondes PPG. La base de données est aléatoirement séparée en deux jeux : 75 % est dédié à l'entraînement des réseaux (soit 543 signaux) et 25 % à la validation (181 signaux).

2.2 Traitement des images et des signaux

Le front correspond à une région d'intérêt pertinente en matière de rapport signal sur bruit [9]. La région est détectée à partir d'un modèle composé de 68 points épousant les formes prin-

cipales du visage. Ces différents points sont suivis le long de la vidéo et certains d'entre eux permettent de calculer automatiquement la position du front. En pratique, les algorithmes de détection du visage et des caractéristiques faciales respectivement inclus dans les bibliothèques OpenCV et Dlib ont été utilisés.

Le signal PPG caméra est construit à partir d'une moyenne spatiale sur le canal vert des pixels du front. Cette technique a été utilisée dès les toutes premières publications relatives à la mesure de signaux PPG sans contact par caméra [9]. Les tendances basses fréquences du signal brut sont supprimées par un filtre passe-bas spécifique [10]. Une détection robuste des vallées est ensuite calculée pour extraire chaque onde. In fine, chaque signal de la base de données est échantillonné sur 256 points et contient 5 ondes PPG successives. Un signal calculé à partir d'une des vidéos est présenté en figure 2. Les signaux en contact de référence mesurés au doigt et à l'oreille sont aussi présentés sur cette figure. Tous les signaux ont été centrés (moyenne nulle) et réduits (écart type égal à un).

Il est proposé, dans cet article, d'exploiter la représentation en ondelettes pour entraîner les différentes architectures neuronales présentées en section 2.3. L'utilisation directe du signal iPPG en entrée d'un modèle d'IA est soutenue par une littérature très faible [6] et des essais préliminaires mais non concluants ont été menés par notre équipe de recherche (résultats non publiés).

La transformée en ondelettes continue d'un signal correspond à une représentation temps-fréquence calculée à partir d'une fonction prototype communément appelée ondelette mère. Contrairement à la transformée de Fourier, la transformée en ondelettes permet de détecter des variations abruptes de fréquence dans les signaux. Différentes ondelettes mères ont été développées et le choix dépend principalement de l'application et des propriétés du signal analysé. L'ondelette mère de Morlet, déjà utilisée dans de précédents travaux relatifs à l'analyse de la PPG par caméra [11], a été retenue dans cette étude.

La transformée en ondelettes continue a été calculée sur chaque signal PPG dans la plage de fréquences physiologiques des battements du coeur humain, soit $[0.6, 4.5]$ Hz [2]. La représentation en ondelettes qui servira à entraîner les architectures neuronales est de dimension 256×256 . Un signal caméra et en contact au doigt avec leur représentation en ondelettes respective (partie réelle) sont présentés en figure 1. Notons la différence de forme entre les signaux et de phase entre les représentations en ondelettes : la partie réelle du signal caméra démarre sur une série de coefficients de faible intensité (pseudo-ellipse bleue) tandis que la partie réelle du signal en contact démarre sur des coefficients de forte intensité (pseudo-ellipse jaune). Il s'agit d'une particularité que le réseau de neurones apprendra pendant la phase d'entraînement.

2.3 Développement des architectures neuronales

Nous proposons d'exploiter l'architecture U-Net initialement utilisé dans le cadre de la segmentation d'images médicales. Cette architecture est constituée d'une branche descendante (en-

codeur) complétée par une branche ascendante (décodeur), donnant une forme de U au réseau. La branche descendante contient un enchevêtrement de couches de convolution et de *pooling*. La branche ascendante intègre des couches de déconvolution connectées aux convolutions de la branche descendante. Les connexions permettent de restaurer l’information spatiale. Une représentation schématique du réseau est proposée en figure 1.

Des squelettes (*backbones*) peuvent être intégrés dans la partie encodeur du réseau U-Net. Les paramètres internes du squelette sont bloqués pendant l’entraînement (les poids du réseau restent fixes). Il s’agit en pratique de modèles pré-entraînés sur la base de données ImageNet pour des tâches de reconnaissance d’objet dans les images [12]. L’apprentissage d’un réseau U-Net soutenu par un squelette consiste à optimiser les paramètres internes de la partie décodeur. Cette stratégie est similaire à un apprentissage par transfert. Différents squelettes populaires ont été testés : la version 16 couches (VGG-16), la version 101 couches de ResNet [13], la version 201 couches du réseau DenseNet [14] ainsi que les réseaux Inception [15] InceptionV3 et InceptionResNetV2. Les techniques conventionnelles de régularisation n’ont pas été introduites tandis qu’un schéma de normalisation (i.e. *batch normalization*) est utilisé dans les réseaux possédant un squelette. La tâche ne correspond pas à une classification de données mais à une régression sous la forme d’une reconstruction pixel à pixel d’une représentation en ondelettes sur deux canaux. Le nombre de variables à entraîner (poids et biais) est compris entre 2 et 9 millions.

3 Résultats et discussion

3.1 Performances des apprentissages

La fonction de coût (*loss*) des moindres carrés (*mean squared error*) a été utilisée :

$$MSE = \frac{1}{n} \sum_{i,j} (CWT_{i,j} - \widehat{CWT}_{i,j})^2 \quad (1)$$

CWT correspond à la transformée en ondelettes (voir section 2.2) du signal PPG en contact et \widehat{CWT} à celle prédite par le réseau de neurones à partir de la transformée calculée sur le signal caméra.

Les valeurs minimales des courbes d’évolution de la fonc-

Réseau	MSE_{doigt}	$MSE_{oreille}$
U-Net1	0.327	0.231
VGG-16	0.282	0.228
ResNeXt101	0.316	0.227
InceptionResNetV2	0.323	0.238
InceptionV3	0.318	0.233
DenseNet201	0.308	0.229

TABLE 1 – Minimum de la fonction de coût pour chaque modèle. U-Net1 correspond au réseau initial n’intégrant pas de squelette. Les autres réseaux correspondent à des architectures U-Net soutenues par un squelette.

tion de coût pour chaque réseau sont répertoriées dans le tableau 1. Indépendamment du site de mesure, le réseau utilisant VGG-16 pour squelette présente la plus faible *MSE*, traduisant ainsi les meilleurs performances en terme de reconstruction de la représentation en ondelettes. Notons tout de même que les valeurs minimales sont proches, en particulier celles calculées à partir des signaux PPG mesurés sur l’oreille.

Nous pouvons aussi observer, toujours dans le tableau 1, une meilleure performance générale (plus faible *MSE*) sur les reconstructions en ondelettes des signaux en contact mesurés à l’oreille par rapport aux signaux en contact mesurés au doigt. Nous supposons que cet écart reflète les différences de forme d’onde mesurée entre les sites, la forme d’une onde PPG caméra étant en général plus proche d’une onde mesurée à l’oreille que d’une onde mesurée au doigt [7].

3.2 Validation point à point des signaux reconstruits

Les modèles neuronaux entraînés délivrent une représentation en ondelettes sur deux plans (une partie réelle et une partie imaginaire). Le signal PPG temporel est reconstruit à partir de la transformée inverse. Un exemple est présenté en figure 3, où il est possible d’apprécier la qualité de la prédiction. L’écart de phase est correctement rectifié par le réseau. Nous pouvons observer que le rebond caractéristique de l’onde PPG est convenablement reproduit alors qu’il est presque toujours absent sur le signal caméra. Nous voyons que le signal a été lissé et que la largeur des ondes est plus faible, montrant que le réseau corrige les coefficients hautes fréquences qui transcrivent les bruits ainsi que les coefficients des fréquences centrales qui déterminent la partie pulsée du signal.

La RMSE (équation 2) a été calculée entre les différentes paires de signaux. Les amplitudes de ces derniers étant arbitraires et normalisées, nous proposons d’observer l’erreur absolue moyenne en pourcentage (*mean absolute percentage error*, voir équation 3).

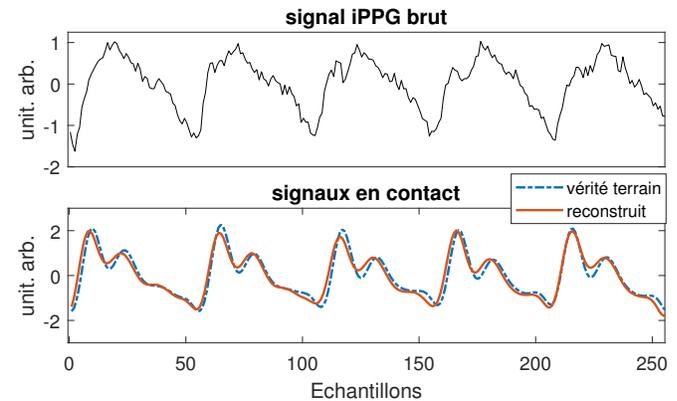


FIGURE 3 – Exemple de reconstruction d’un signal en contact au doigt (figure du bas) à partir du signal caméra (figure du haut). Notons la bonne qualité de la reconstruction même si quelques erreurs sont visuellement perceptibles.

$$RMSE = \sqrt{MSE} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_i \left| \frac{p_i - q_i}{p_i} \right| \quad (3)$$

La *MAPE* est ici calculée entre deux signaux (p et q dans l'éq. 3). Les résultats sont présentés dans le tableau 2 où nous pouvons observer la performance des prédictions délivrées par les modèles neuronaux. L'erreur sur le réseau utilisant le squelette VGG-16 est légèrement plus faible, ce qui cohérent avec les résultats présentés en section 3.1 et dans le tableau 1.

4 Synthèse des contributions et travaux futurs

Nous avons proposé, dans cet article, une architecture neuronale permettant de reconstruire avec précision une onde PPG en contact à partir d'une onde PPG sans contact estimée par analyse vidéo. La reconstruction est effectuée par le biais de la représentation temps-fréquence du signal via sa transformée en ondelettes continue. Les réseaux de neurones proposés correspondent à des architectures U-Net avec et sans squelette. Le signal reconstruit est proche de la vérité terrain en contact.

La motivation principale de ce travail correspond à la possibilité de proposer une estimation de la pression artérielle par l'analyse d'ondes PPG mesurées par caméra. La prochaine étape consistera donc à intégrer les signaux reconstruits dans des modèles d'IA dédiés à l'estimation de la pression artérielle par signaux en contact, ces derniers pouvant être collectés sur de larges bases de données publiques (e.g. MIMIC).

Des pistes d'amélioration de ce travail sont envisagées. Nous proposons dans un premier temps d'étoffer la base de données qui est actuellement limitée en volume et en nombre de participants. Les vidéos exploitées dans cette recherche ont été acquises par une caméra rapide (125 fps). Nous envisageons

Réseau	cPPG _{doigt} vs \widehat{cPPG}_{doigt}	cPPG _{oreille} vs $\widehat{cPPG}_{oreille}$
U-Net1	0.25 (0.06)	0.20 (0.04)
U-NetVGG16	0.23 (0.05)	0.20 (0.04)
U-Net _{ResNeXt101}	0.24 (0.06)	0.20 (0.04)
U-Net _{InceptionResNetV2}	0.25 (0.06)	0.20 (0.04)
U-Net _{InceptionV3}	0.25 (0.06)	0.20 (0.04)
U-Net _{DenseNet201}	0.24 (0.06)	0.20 (0.04)

TABLE 2 – *RMSE (MAPE)* (voir les équations 2 et 3) calculées entre les prédictions délivrées par les différentes architectures neuronales et les vérités terrain. $cPPG_{doigt}$ et $cPPG_{oreille}$ correspondent aux signaux de vérité terrain mesurés au doigt et à l'oreille respectivement (voir courbe bleue sur la figure 3 pour un exemple typique). \widehat{cPPG}_{doigt} et $\widehat{cPPG}_{oreille}$ correspondent aux prédictions calculées par la transformée inverse des représentations en ondelettes délivrées par les modèles neuronaux (voir courbe orange sur la figure 3).

d'étudier dans des travaux futurs les signaux formés à partir de caméras classiques (30 fps). Les ondes PPG acquises par de tels capteurs sont moins détaillées et donc plus complexes à analyser. Il sera en contrepartie possible d'entraîner les modèles avec un volume plus conséquent de données, de nombreuses bases dédiées à l'étude de signaux PPG mesurés par des caméras classiques étant désormais publiquement disponibles. Une intégration directe de la vidéo plutôt que des représentations temps-fréquence dans l'architecture U-Net fera l'objet de travaux de recherche sur le plus long terme.

Références

- [1] D. McDuff, "Camera measurement of physiological vital signs," *arXiv preprint arXiv :2111.11547*, 2021.
- [2] S. Zauneder et al., "Cardiovascular assessment by imaging photoplethysmography—a review," *Bio. Eng.*, 2018.
- [3] A. Ni et al., "A Review of Deep Learning-Based Contactless Heart Rate Measurement Methods," *Sensors*, 2021.
- [4] H. Luo et al., "Smartphone-based blood pressure measurement using transdermal optical imaging technology," *Circulation : Card. Imag.*, vol. 12, p. e008857, 2019.
- [5] N. Sugita et al., "Contactless Technique for Measuring Blood-Pressure Variability from One Region in Video Plethysmography," *Journal of Med. and Bio. Eng.*, 2018.
- [6] F. Schrumpp et al., "Assessment of Non-Invasive Blood Pressure Prediction from PPG and rPPG Signals Using Deep Learning," *Sensors*, vol. 21, no. 18, p. 6022, 2021.
- [7] D. Djeldjli et al., "Remote estimation of pulse wave features related to arterial stiffness and blood pressure using a camera," *Biomed. Sig. Proc. and Control*, vol. 64, 2021.
- [8] J. Cheng et al., "Prediction of arterial blood pressure waveforms from photoplethysmogram signals via fully convolutional neural nets," *Comp in Bio and Med*, 2021.
- [9] W. Verkruyse et al., "Remote plethysmographic imaging using ambient light," *Optics express*, vol. 16, 2008.
- [10] M. Tarvainen et al., "An advanced detrending method with application to HRV analysis," *IEEE Trans. on Bio-med. Eng.*, vol. 49, no. 2, pp. 172–175, Feb. 2002.
- [11] F. Bousefsaf et al., "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Bio. Sig. Proc. and Control*, vol. 8, pp. 568–574, 2013.
- [12] E. C. Too et al., "A comparative study of fine-tuning deep learning models for plant disease identification," *Comp. and Elec. in Agr.*, vol. 161, pp. 272–279, 2019.
- [13] K. He et al., "Deep residual learning for image recognition," in *IEEE CVPR*, 2016, pp. 770–778.
- [14] G. Huang et al., "Densely connected convolutional networks," in *IEEE CVPR*, 2017, pp. 4700–4708.
- [15] C. Szegedy et al., "Inception-v4, inception-resnet and the impact of res. connections on learning," in *AAAI*, 2017.