HAL
open science

Vers le développement d'un capteur photopléthysmographique sans contact

Serge Bobbia

To cite this version:

Serge Bobbia. Vers le développement d'un capteur photopléthysmographique sans contact. Traitement des images [eess.IV]. Université de Bourgogne - Franche Comté, 2019. Français. NNT : . tel-02068347

HAL Id: tel-02068347

https://hal.science/tel-02068347v1

Submitted on 14 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

PRÉPARÉE À L'UNIVERSITÉ DE BOURGOGNE

École doctorale n°37

Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Informatique et Instrumentation de l'Image

par

SERGE BOBBIA

Vers le développement d'un capteur photopléthysmographique sans contact

Thèse présentée et soutenue à Dijon, le 15/02/19

Composition du Jury :

PINNA ANDREA	Maître de conférences HDR à Sorbonne Université	Rapporteur
ROSENBERGER CHRISTOPHE	Professeur à l'ENSICAEN	Rapporteur
VRAY DIDIER	Professeur à l'INSA Lyon	Président du jury
YANG FAN	Professeur à l'Université de Bourgogne Franche-Comté	Examinatrice
BENEZETH YANNICK	Maître de conférences à l'Université de Bourgogne Franche-Comté	Encadrant de thèse
DUBOIS JULIEN	Maître de conférences HDR à l'Université de Bourgogne Franche-Comté	Directeur de thèse

N° | X | X | X |

SPIM
école doctorale **sciences pour l'ingénieur et microtechniques**

Titre : Vers le développement d'un capteur photopléthysmographique sans contact

Mots-clés : Traitement d'images, Ingénierie biomédicale, Photopléthysmographie sans contact, Segmentation en superpixels

Résumé :

La mesure cardiaque sans contact réalisée grâce aux méthodes de photopléthysmographie sans contact est un domaine de recherche très actif. Depuis l'introduction en 2010 d'une nouvelle méthode de mesure avec des capteurs optiques d'entrée de gamme (*webcam PC*), les travaux de recherche se sont multipliés. Ainsi, on observe une plus grande diversité des méthodes proposées afin de réaliser la mesure. Également, la précision de la mesure a grandement progressé et les scénarios et possibilités d'usage de la technologie sont aujourd'hui très nombreux. Au cœur de ce processus de mesure, la segmentation dans l'image de la ou les zones d'intérêt est une étape clé. Nous proposons dans cette thèse une méthode innovante afin de réaliser la mesure photoplethysmographique sans contact en identifiant implicitement les zones de peau vivante dans la vidéo. Nous avons montré que notre approche permet d'améliorer la qualité de la mesure en favorisant les zones dans l'image où le signal est de plus grande qualité. Afin de rendre possible l'intégration de notre solution, nous avons proposé une nouvelle méthode de segmentation en *superpixels*, nommée IBIS, qui permet de réduire la complexité algorithmique de cette étape du traitement. Ce faisant, nous avons démontré la faisabilité de l'intégration de notre solution au sein d'une plateforme embarquée. Les différentes méthodes ont été évaluées au travers de plusieurs expérimentations afin de valider leurs performances. Notre méthode de segmentation en *superpixels* est comparée aux méthodes de l'état de l'art tandis que nous avons implémenté plusieurs des méthodes de mesure du signal photoplethysmographique afin de discuter de l'impact de notre approche sur la qualité de la mesure photoplethysmographique. Que ce soit pour la segmentation en *superpixels* ou pour l'estimation du rythme cardiaque sans contact, nous avons montré une importante plus-value de nos méthodes comparées à celles disponibles dans la littérature. Les différents travaux présentés dans ce document ont été valorisés au travers de publications en conférences et revue.

Title: Vers le développement d'un capteur photopléthysmographique sans contact

Keywords: Image processing, Biomedical engineering, Remote photoplethysmography, Superpixels segmentation

Abstract:

Heart-rate estimation performed with remote photoplethysmography is a very active research field. Since pioneer works in 2010, which demonstrated the feasibility of the measure with low-grade consumers' camera (webcam), the number of scientific publications have increased significantly in the domain. Hence, we observe a multiplication of the methods in order to retrieve the photoplethysmographic signal which has led to an increased precision and quality of the heart-rate estimation. Region of interest segmentation is a key step of the processing pipeline in order to maximize the quality of the measured signal. We propose a new method to perform remote photoplethysmographic measurement using an implicit living skin identification method. Hence, we have shown that our approach lead to an improvement in both quality of the signal measured and precision of the heart-rate estimation by favoring more contributive area. As we are working with hardware integration constraint, we propose a new *superpixels* segmentation method which requires significantly less computation power than state of the art methods by reducing the algorithmic complexity of this step. Moreover, we have demonstrated the integration and real time capabilities by implementing our solution to an embedded device. All of our proposed method have been evaluated through different experimentations. Our new segmentation method, called IBIS, have been compared to state of the art methods to quantify the quality of the produced segmentation. To quantify the impact of our approach on the quality of the photoplethysmographic measure, we have implemented and compared state of the art methods with our proposed method. For both the superpixels segmentation and remote heart-rate estimation, our methods have shown great results and advantages compared to state of the art ones. Our works have been reviewed by the scientific community through several conference presentations and journal publications.

Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon, France

A mes parents.

Ma mère, pour être la meilleure des professeures.
Mon père, pour m'avoir transmis sa curiosité et à qui j'aurais adoré présenter ces travaux.

Remerciements

Ces travaux de thèse ont été supporté par le Conseil Régional de Bourgogne Franche Comté et l'Union Européenne au travers du programme FEDER.

Mes remerciements vont à monsieur Christophe Rosenberger, professeur à l'ENSICAEN, ainsi qu'à monsieur Pinna Andrea, maître de conférences HDR à Sorbonne Université, pour avoir accepté de rapporter mon travail. À madame Yang Fan, professeur à l'Université de Bourgogne Franche-Comté, ainsi qu'à monsieur Vray Didier, professeur à l'INSA Lyon, pour avoir accepté d'examiner mon travail. À mes encadrants de thèse, le maître de conférences Benezeth Yannick et le maître de conférences HDR Dubois Julien pour leur soutien et leur patience.

Mes remerciements vont au laboratoire d'accueil Le2i qui aura mis à disposition les ressources nécessaires à la bonne conduite de cette thèse. À l'équipe de recherche qui a facilité mon intégration au monde de la recherche. Aux doctorants et docteurs : Bonazza Pierre (qui est un excellent covoiturage), Boscaro Anthony, Bouderbane Mustapha, Khokhlova Margarita, Macwan Richard, Marin Yoan, Moinet Axel et tous les autres pour leur aide et leur soutien durant ces trois années de thèse communes.

Mes remerciements vont à ma famille et mes amis qui m'ont permis de m'évader d'un superpixels ou d'une page de code trop complexe le temps d'un instant, pour toujours repartir de plus belle.

Enfin, mes remerciements vont à ma compagne Roy Ludivine, qui a su me motiver et me supporter le temps de cette thèse. Qui a toujours été mon principal soutien durant ces trois années. Et qui, dans les moments difficiles, m'a toujours encouragé à avancer dans ce projet.

UNION EUROPEENNE

RÉGION BOURGOGNE FRANCHE COMTÉ

avec le Fonds européen de développement régional (FEDER)

Sommaire

1 Introduction **1**

1.1 Contexte : mesure de constantes vitales 1

1.2 Les capteurs cardiaques 3

1.3 Applications de la mesure cardiaque optique sans contact 4

1.3.1 Médical 4

1.3.2 Paramédical et loisir 6

1.3.3 Sécurité des personnes 6

1.4 Plan de la thèse 7

I État de l'art et problématique **9**

2 Les systèmes photopléthysmographiques **11**

2.1 Principes de fonctionnement 12

2.2 Limites de la photopléthysmographie au contact 15

2.3 Du point aux pixels : la photopléthysmographie sans contact 16

2.3.1 Impact du capteur 18

2.3.2 Capteur matriciel : mesures multi-points pour la pression artérielle 19

2.3.3 Longueurs d'onde pour la rPPG 20

3 Concepts et méthodes rPPG existantes **23**

3.1 Modèles d'interactions lumière-tissu pour la rPPG 24

3.2 Algorithmes rPPG existants 26

3.2.1 Algorithmes par séparation de sources 27

3.2.2 Algorithmes par projection 29

3.3 Problématiques et objectifs de la thèse 34

II Mesure rPPG : algorithme et validation **39**

4 Détection non-supervisée de la région d'intérêt pour la mesure rPPG **41**

4.1 Motivations 42
4.1.1 Travaux connexes sur la détection non supervisée de la région d'intérêt . 43
4.1.2 Segmentation vidéo par superpixels temporels : présentation de TSP . . . 45
4.2 Segmentation en superpixels temporels pour la rPPG 47
4.2.1 Métrique de qualité d'un signal rPPG avec un modèle en fréquence . . . 48
4.2.2 Fusion des signaux rPPG 51
4.2.3 Modélisation d'un signal rPPG avec des modèles de Markov cachés . . . 52
4.2.4 Métrique de qualité d'un signal rPPG avec les modèles de Markov cachés 55

5 Validation expérimentale de la mesure rPPG avec une détection non supervisée de la région d'intérêt 57
5.1 Protocole expérimental 58
5.1.1 Données de test : UBFC-rPPG 58
5.1.2 Méthodes implémentées 60
5.1.3 Métriques 61
5.2 Résultats 62
5.2.1 Validation préliminaire de la segmentation non supervisée : cas d'usage simple 62
5.2.2 Validation de la segmentation non supervisée : cas d'usage réaliste 64
5.2.3 Validation de la métrique de qualité du signal basée sur les modèles de Markov cachés 70
5.3 Discussions 72

III Optimisations et implémentation pour une mesure rPPG temps réel 75

6 Segmentation en superpixels à complexité réduite pour la rPPG 77
6.1 Segmentation par superpixels 78
6.2 Présentation de IBIS 80
6.2.1 Identification implicite des frontières 81
6.2.2 Analyse de la mise à jour des seeds en fonction des itérations 83
6.2.3 Optimisation 84
6.2.4 Évaluation de la complexité algorithmique 87
6.3 Propagation temporelle de la segmentation et post-traitement pour la mesure rPPG 92
6.4 Validation 93
6.4.1 Plateforme PC à processeur multi-cœurs 94
6.4.2 Evaluation de la qualité de segmentation de IBIS 96

6.4.3 Mesure rPPG 99
6.5 Discussions 101

7 Implémentation sur plateforme embarquée GPU 103
7.1 Intégration et optimisation : IBIScuda 104
7.2 Comparaison des méthodes sur GPU 108
7.2.1 gSLICr et IBIScuda 108
7.2.2 Evolution des temps de calcul sur Jetson TX1 et Geforce GTX 1080 ti . . 109
7.3 Gain par limitation du nombre d'itérations 112
7.3.1 Impact du nombre d'itérations 112
7.3.2 Performances maximales d'IBIScuda sur Jetson TX1 et Geforce GTX 1080 ti 113

8 Conclusions et perspectives 119
8.1 Conclusions 120
8.2 Valorisation 123
8.3 Ouvertures 124

1

Introduction

Sommaire

1.1 Contexte : mesure de constantes vitales 1
1.2 Les capteurs cardiaques 3
1.3 Applications de la mesure cardiaque optique sans contact 4
1.3.1 Médical 4
1.3.2 Paramédical et loisir 6
1.3.3 Sécurité des personnes 6
1.4 Plan de la thèse 7

Cette thèse de doctorat se situe dans le cadre d'un programme Jeune Chercheur Entrepreneur (JCE) financé par la région Bourgogne Franche-Comté et le Fond Européen de Développement Régional (FEDER). Elle est effectuée au sein du laboratoire Le2i de l'Université de Bourgogne et de l'école doctorale des Sciences Pour l'Ingénieur et Microtechniques (SPIM).

1.1/ Contexte : mesure de constantes vitales

La mesure de paramètres biologiques vitaux est actuellement un champ d'étude très actif. Dans le milieu médical, avec la démocratisation des appareils de mesures automatisées, il est possible d'effectuer un monitoring quasi constant des signes vitaux d'un patient. Ceux-ci sont habituellement définis au nombre de 4 :

— La température,
— Le rythme cardiaque,
— Le rythme respiratoire,
— La pression artérielle.

La capacité respiratoire et la saturation en oxygène dans le sang, notée SpO_2 sont quelques fois ajoutées à cette liste. Leurs mesures, unités et notations doivent respecter les standards du système LOINC (Logical Observation Identifiers Names and Codes) établis en 1994 par le Regenstrief Institute, une organisation de recherche médicale américaine à but non lucratif. Ce standard comprend aujourd'hui plus de 71 000 termes pour tous types d'observations.

L'automatisation au cœur des systèmes modernes permet la surveillance et le déclenchement d'alerte en cas de mesure continue révélant une anomalie. Il est à noter qu'en fonction de l'âge, les plages d'acceptation des différents signes vitaux peuvent varier. Ainsi, au repos, la température d'une personne doit être comprise entre 35°C et 38°C pour ne pas être en hypothermie ou hyperthermie. Le rythme cardiaque doit être compris entre 100 et 160 battements par minute (bpm)

pour un nouveau né contre 60 à 100 bpm pour un adulte. Le rythme respiratoire doit être compris entre 30 et 50 cycles par minute pour un nouveau né contre 12 et 20 cycles par minute pour un adulte. Enfin la pression artérielle, qui se décompose en deux mesures : Systolique pendant la contraction cardiaque et Diastolique pendant le remplissage du cœur entre deux battements, notée Systolique/Diastolique doit être comprise entre 90-115/60-75 et 140-159/90-99 mmHg pour ne pas être en hypotension ni hypertension. C'est la connaissance de ces plages de valeurs, et l'utilisation de capteurs précis, qui permet l'automatisation de la surveillance de ces constantes. Ainsi nombre de pathologies sont détectées en fonction de la variation de ces paramètres.

Aujourd'hui, toutes ces mesures sont effectuées au contact dans un contexte médical ou paramédical, et quelques fois de manière invasive suivant la précision souhaitée. La pression artérielle par exemple ne peut être surveillée en continue qu'à l'aide d'un cathéter. La mesure externe se faisant par *contrepulsion* (brassard avec stéthoscope), elle n'est pas adaptée à une mesure continue. La température se mesure au contact, bien que des systèmes optiques permettent l'estimation de la température dans le domaine infrarouge, les dispositifs médicaux ne sont pas équipés de ce genre de systèmes. Ces deux mesures sont bien souvent faites de façon ponctuelle ou périodique pour de longues périodes d'observation, de l'ordre de plusieurs jours par exemple. Le rythme respiratoire, la saturation en oxygène et le rythme cardiaque sont des paramètres qui sont plus adaptés à une surveillance continue. Ainsi, les capteurs qui permettent cette mesure doivent répondre à plusieurs critères. Une grande fiabilité de la mesure est nécessaire compte tenu de l'importance des données et diagnostiques associés. En ce sens la réglementation exige que les appareils qui fournissent une mesure de paramètres vitaux dans le cadre médical, où les mesures impacteront les choix des praticiens, soient certifiés par un organisme agréé ISO 13485.

FIGURE 1.1 – Exemple de système pour la mesure cardiaque. Dans un cadre médical (gauche) : oxymètre de pouls, et dans un cadre récréatif (droite) : montre connectée.

Dans un contexte sociétal plus général, la démocratisation et la généralisation de l'utilisation des appareils électroniques (smartphone, montre connectée, ...) entraîne une utilisation de la mesure de paramètres vitaux auprès des particuliers. Le sport est un des grands consommateurs de ce type de mesure, et il est possible de trouver sur le marché de nombreuses montres connectées qui permettent de mesurer le rythme cardiaque (*e.g.* Figure 1.1 droite). La mesure permet alors des entrainements personnalisés pour mettre l'accent sur le renforcement cardiaque, l'élimination des graisses et autres. Ces dispositifs permettent également l'analyse des phases du sommeil par exemple, ou encore la surveillance du cycle menstruel chez les femmes si l'on ajoute les données de température.

La lutte contre le stress en entreprise est aussi un enjeu de plus en plus important et considérée par les entreprises qui sont désireuses d'optimiser leurs ressources humaines. La variabilité du rythme cardiaque est ainsi un paramètre très important dans l'objectif de la création de produits à

destination des entreprises dans le marché du bien être. Il a été montré une importante corrélation entre la cohérence cardiaque, établi à partir de cette mesure, et le niveau de stress voire l'état émotionnel. Plusieurs entreprises proposent déjà des exercices personnalisés qui visent à permettre une régulation du stress. La mesure du rythme cardiaque est faite en général au contact, avec un oxymètre de pouls (*e.g.* Figure 1.1 gauche).

Dans le cadre de cette thèse nous aborderons la problématique de la mesure du rythme cardiaque sans contact. Il n'existe pas aujourd'hui de système reconnu suffisamment fiable, bénéficiant de la certification médicale par exemple, qui fournisse une mesure cardiaque sans contact. Bien qu'il soit d'ores et déjà possible de trouver des exemples d'utilisation de cette technologie liée à l'estimation du rythme cardiaque ou de la cohérence cardiaque dans le commerce et notamment dans le cas d'applications mobiles, toutes ces mesures sont considérées dans un cadre récréatif.

C'est dans ce contexte que cette thèse propose l'étude et l'amélioration des méthodes de mesure cardiaque sans contact, et plus précisément des méthodes optiques de mesure sans contact que l'on nommera *photopléthysmographie sans contact* en accord avec la littérature.

1.2/ Les capteurs cardiaques

Le rythme cardiaque est défini comme le nombre moyen de contractions du muscle cardiaque en 1 minute, on le note battements par minute (bpm). Du fait de son évolution lente, la mesure du rythme cardiaque peut nécessiter des temps d'intégration minimum de 10 à 20 secondes pour des mesures non-invasives. Bien évidemment, la qualité et précision de la mesure sont directement liées à ces temps d'intégration. Ainsi, une période d'observation plus courte fournira une mesure instantanée temporellement mieux résolue mais sera moins précise et plus difficile à obtenir, notamment avec des systèmes optiques dont la précision n'est pas suffisante pour travailler avec des temps d'intégration très courts. Suivant les pathologies que l'on cherche à détecter, il faut surveiller l'évolution du rythme cardiaque sur de longues périodes de temps, ce qui peut demander plusieurs jours ou semaines d'observation. Ainsi, en fonction des besoins et du niveau de criticité de la mesure en matière de précision ou de temps de surveillance, différents systèmes existent afin de mesurer le rythme cardiaque. Dans le cadre médical, cette mesure est effectuée de 3 façons :

- La méthode hémodynamique consiste à introduire un fil cathéter dans une artère reliée à un capteur de pression. Les variations de pression ainsi mesurées permettent d'établir l'exact moment du cycle cardiaque et son rythme instantané.
- L'électrocardiographie permet d'enregistrer les modulations de polarisation des membranes cellulaires du cœur et ainsi d'avoir une mesure très précise de l'activité du muscle cardiaque. Elle s'applique en disposant jusqu'à 12 électrodes sur le corps du patient, suivant un schéma bien défini.
- La photopléthysmographie est une méthode optique non invasive qui permet d'évaluer les variations du volumes sanguins dans les tissus superficiels par la variation de l'absorption de lumière dans ces tissus. Cette méthode permet également de déterminer le niveau d'oxygénation. Les oxymètres de pouls sont généralement fixés au doigt ou à l'oreille.

Même si la mesure optique présente de nombreux avantages (facilité d'utilisation, faible cout, ...), la mesure fournie est moins précise qu'avec un ECG. Ce système permet par exemple la mesure du pic de l'onde R, qui est utilisé dans le diagnostique de certaines pathologies [23]. L'ECG reste la méthode de mesure de référence dans le cadre médical. C'est par son aspect pratique et son faible cout que la mesure optique s'est imposée comme outils de mesure privilégiée pour le monitoring non critique.

La nécessité de contact pour les capteurs existants entraîne une multitude de contraintes. Le contact prolongé avec la peau peut causer des lésions cutanées ou des irritations. De plus, les mouvements perturbent grandement la mesure. L'information électrique mesurée par l'électrocardiogramme est perturbée par l'activité des muscles actifs et elle est, de plus, très contraignante pour l'utilisateur en fonction du nombre d'éléctrodes employées. Au regard des limitations des outils actuels de mesure du rythme cardiaque, la photopléthymographie sans contact (rPPG) apporte une forte plus-value. En effet, cette méthode de mesure permet d'extraire à distance, typiquement de plus de 50 cm à plusieurs mètres voire dizaine de mètres [48], un signal physiologique directement corrélé à l'activité cardiaque. De plus, la lumière ambiante est bien souvent suffisante comme source lumineuse. Également, la mesure peut être conduite en utilisant un capteur optique d'entrée de gamme puisqu'ils offrent une sensibilité suffisante pour capter les subtiles modifications de couleur de la peau [63]. Il est alors théoriquement possible d'estimer les mêmes informations qu'avec la méthode photopléthysmographique au contact.

L'activité de recherche dans le cadre de la photopléthysmographie sans contact étant assez récente, avec une augmentation significative de l'activité de recherche dans ce domaine depuis 2007 [36, 42]. La dénomination de la technologie n'est pas encore arrêtée dans la communauté et on peut trouver de nombreuses références sous diverses appellations. La technologie a été premièrement nommée PPG imaging (PPGI) en 1996 [13], les dénominations suivantes font également référence à cette technologie : camera-based PPG (cbPPG) [107], non-contact PPG (ncPPG) [45], imaging PPG (iPPG/IPPG) [57], remote imaging PPG (RIPPG) [96], distance PPG (DistancePPG) [100], pulse camera (PulseCam) [116], cardiovascular camera (CardioCam) [75], video PPG (vPPG) [114], videoplethysmography (VPG) [121], videobased heart-rate (VHR) [92]. Dans ce document, cette technologie sera toujours référencée comme photopléthysmographie sans contact ou abrégée rPPG (pour remote PPG).

1.3/ Applications de la mesure cardiaque optique sans contact

Les applications de la mesure cardiaque optique et sans contact sont multiples. Elles touchent à de nombreux aspects de notre vie et sont ici regroupées en 3 catégories : le domaine médical, le paramédical ou l'utilisation à destination des particuliers (loisirs) et la sécurité. Bien évidemment, cette liste n'est pas exhaustive et il existe très certainement d'autres applications potentielles. Cependant, cela permet de se rendre compte du potentiel de la technologie.

1.3.1/ Médical

Dans le cadre médical, l'utilisation de la rPPG peut servir au maintien à domicile des personnes en réhabilitation suite à une opération ou pour les personnes âgées en raison des capacités de mesure passive qui ne requiert aucune manipulation de la part des usagés. La technologie peut également être employée pour le bien-être quotidien en fournissant un retour sur la qualité du sommeil, le conseil sur l'exécution d'exercice etc [50], voire d'un domicile entièrement monitoré [15]. Ainsi, en 2011, Poh et al. [53] proposent l'intégration d'une caméra RGB à un miroir afin de fournir une interface intelligente avec un retour sur l'information cardiaque. Cet exemple d'application crée la possibilité pour la surveillance à long terme des patients, mesure du rythme cardiaque au repos de façon quotidienne pour la détection d'arythmie par exemple sans aucune instrumentation nécessaire. La surveillance nocturne, pour estimer la qualité du sommeil, a d'ores et déjà été testée avec un capteur rPPG installé au-dessus du lit afin de maintenir la surveillance durant du sommeil [29]. Il faut noter que la mesure nocturne est possible grâce à l'utilisation d'une source de lumière

proche infra-rouge.

La technologie a pu être adaptée avec succès à des scénarios particuliers, et notamment dans le cas de l'instrumentation de couveuses [134]. En 2013, Aarts et al. [67] proposent une étude pilote afin de réaliser la mesure du rythme cardiaque sans contact en instrumentant une couveuse dans une unité de soin intensif néonatal. La technologie est testée sur 19 prématurés dont le rythme cardiaque est mesuré avec succès. La passivité du système permet d'éviter des risques d'infection ou de lésions cutanées chez les très jeunes bébés causées par les électrodes ou la pression au contact des oxymètres de pouls. Qui plus est, les surfaces de peau disponibles ne sont pas bien adaptés à l'ECG ou l'oxymétrie au contact. Le signal polychromatique fourni par les caméras RGB utilisées pour réaliser la mesure rPPG rend possible l'estimation d'autres constantes vitales. Le rythme respiratoire et la concentration en oxygène dans le sang (SpO_2) sont deux mesures qu'il est également possible de mesurer sans contact [87]. Pour cette même application, en 2014, Villarroel et al. [89] étendent la mesure au rythme respiratoire et à la saturation en oxygène en utilisant comme seul source le signal rPPG. La suppression du contact avec la peau revêt une importance toute particulière suivant les applications envisagées, en cas de brûlures ou de plaies qui empêchent le contact. La technologie a également été testée dans le cadre de la chirurgie cardiaque et notamment pour la surveillance post-opératoire [77, 120].

La mesure du rythme cardiaque permet de déceler des pathologies critiques, par exemple pour prévenir des complications dans le cadre sportif [16]. L'une des pathologies communes directement décelable à partir de l'observation continue du rythme cardiaque est la tachycardie. On considère qu'elle survient chez un adulte moyen pour un rythme cardiaque mesuré au repos, assis et sans fournir d'effort particulier, supérieur à 100 bpm. A l'inverse, la bradycardie correspond à un rythme cardiaque mesuré au repos inférieur à 50 bpm. Ces valeurs ne sont pas absolues et, suivant l'état de forme de l'individu, peuvent être revues à la hausse ou à la baisse par le praticien. Dans la pratique un dispositif ECG mobile, nommée *HOLTER*, permet la surveillance de longue durée des patients. A titre d'exemple la Figure 1.2 montre l'utilisation de l'un de ces dispositifs et permet de mettre en perspective tout l'intérêt d'une mesure sans contact pour assurer le confort du patient.

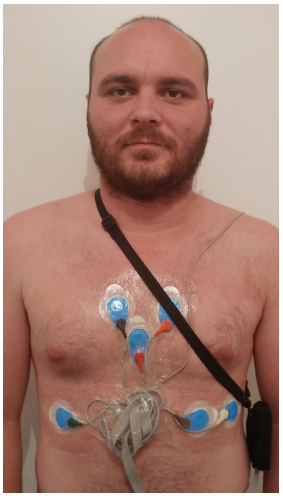


FIGURE 1.2 – Exemple d'installation d'un dispositif ECG mobile *HOLTER*.

1.3.2/ Paramédical et loisir

Le sport est une des principales applications pouvant bénéficier des avantages d'une technologie sans contact. En effet, le rythme cardiaque est très utile dans le cadre d'entrainements pour évaluer la santé du sportif [5]. Cependant, la nécessité de contact des capteurs est ici très limitante puisque les systèmes employés ne permettent pas une grande liberté de mouvement. En effet les systèmes plus compacts, telles que les montres connectées, ne sont pas toujours suffisamment précis pour répondre aux besoins des sportifs. Les méthodes de mesure des paramètres physiologiques comme le rythme cardiaque ou le rythme respiratoire requièrent l'utilisation d'un système de mesure ECG en ceinture autour de la poitrine et/ou d'un capteur optique au poignet [86, 90]. Les avancées technologiques permettent une mesure bien plus robuste au mouvement en cas d'exercice [135], cependant la mesure optique est toujours très perturbée par les mouvements intenses. Il existe plusieurs travaux qui mettent en avant l'utilisation de la mesure rPPG dans le cadre sportif [94] et pointent les contraintes spécifiques qui y sont liées comme l'apparition d'information périodique ou pseudo-périodique qui intègre le spectre de fréquences du rythme cardiaque et compliquent ainsi la mesure [80]. Les pathologies comme l'arythmie peuvent être détectées juste après le sport pendant le ralentissement du rythme cardiaque. En 2011 Sun et al. [55] montrent la bonne qualité de la mesure et son efficacité pour ce scénario.

Comme nous l'avons décrit précédemment, la variabilité du rythme cardiaque (HRV) est un paramètre fondamental dans les applications liées à l'évaluation du stress. En 2013, Bousefsaf et al. [68] exploitent la mesure rPPG afin de valider la bonne qualité de la mesure de HRV. La corrélation entre un stress induit par une activité cognitive et la mesure de rPPG valide la faisabilité de la mesure et son exploitation. Depuis ces travaux, le HRV mesuré à partir du signal rPPG a été utilisé à plusieurs reprises comme outils de mesure objective du système nerveux autonome [88, 138].

1.3.3/ Sécurité des personnes

Les systèmes de sécurité modernes sont focalisés autour de la surveillance de l'individu et la récupération d'informations objectives le concernant ou bien sur la détection de menaces ou risques potentiels. Dans le premier cas, la biométrie est un champ de recherche très actif qui propose depuis longtemps maintenant d'identifier et reconnaître les utilisateurs par leurs caractéristiques biologiques uniques. A ce titre, les empreintes digitales représentent un exemple très connu qui répond à ces critères. Depuis plusieurs années maintenant, la reconnaissance faciale est utilisée afin de compléter [10, 27] voire de se passer de l'usage des empreintes, et surtout de permettre une identification sans être au contact du capteur. La principale faiblesse de ce genre de système est qu'ils sont facilement trompés par des images, masques ou vidéos [81, 117]. Les méthodes les plus avancées de résistance à ce genre d'attaque, appelées *anti-spoofing*, passent par l'analyse en détail de l'image afin de détecter les textures en présence [79, 110]. Malheureusement, ces méthodes montrent vite leurs limites puisqu'elles ne considèrent pas d'informations supplémentaires à celles fournies par l'image, ce qui impose des contraintes uniquement sur la qualité de l'imitation. En intégrant un aspect temporel et la détection d'un rythme cardiaque cohérent, il est possible de proposer des systèmes beaucoup plus robustes à ce genre d'attaque. En 2016, Liu et al. [118] utilisent la détection d'un signal rPPG afin de renforcer la robustesse d'un système biométrique aux risques d'usurpation d'identité. Dans le cas de la surveillance passive d'individus, groupes ou foules, l'identification à distance permet la détection et possiblement la localisation d'une personne et peut bénéficier de l'estimation du rythme cardiaque pour éliminer les faux positifs [48, 108]. En 2014, Apvrille et al. [78] ont équipé des drônes autonomes afin de permettre

la fouille et la détection d'individus dans des zones sinistrées, dans le cas de catastrophes naturelles par exemple, la photopléthysmographie sans contact permet alors la détection de personnes vivantes et fournit une information utile aux équipes de soins.

La sécurité automobile est un secteur qui va également bénéficier du déploiement de la technologie rPPG. L'installation de caméras à bord des véhicules, afin de prévenir les accidents, est de plus en plus courante [37]. Les véhicules détectent plusieurs paramètres et activités propres au conducteur, que ce soit par la détection et surveillance du rythme de clignotement des yeux [60] ou du suivi du regard [105]. La rPPG se place en complément de ces mesures et permet de mesurer le rythme cardiaque afin de fournir une indication sur l'état de santé du conducteur [101, 127, 137], ou le HRV afin de détecter l'état de fatigue [124]. En 2006, Byeon et al. [31] proposent l'usage de cette mesure afin de détecter l'état de somnolence du conducteur. De plus, l'usage d'un capteur optique suffit à obtenir ce paramètre ce qui n'oblige pas à une instrumentation lourde du conducteur.

Cette thèse a pour objectif, au regard des travaux déjà existants, de proposer une méthode innovante de mesure rPPG et d'intégrer cette méthode sur une plateforme embarquée pour avoir au final un capteur photopléthysmographique sans contact. La méthode de mesure doit être la moins contraignante possible, que ce soit par le contrôle de l'environnement, le contrôle des sources lumineuses par exemple, ou par la nécessité de détection d'une zone spécifique du corps. Le capteur doit pouvoir effectuer une mesure en temps réel, dont la contrainte essentielle se limite à la capacité de fonctionner avec un rythme minimal d'échantillonnage du capteur vidéo de 25 images par seconde, ce qui est la fréquence d'échantillonnage minimale pour une caméra (*e.g.* webcam pour PC) standard.

1.4/ Plan de la thèse

Ce document se décompose en 3 parties. La première partie est focalisée sur l'étude des méthodes et modèles employées jusqu'alors dans la littérature pour la mesure du rythme cardiaque. Cette partie est constituée de deux chapitres. Le chapitre 2 retrace l'évolution de la technologie photopléthysmographique au travers des mesures par transmitance et reflectance mis au point dès 1937 [1]. Les limitations de la méthode au contact sont alors abordées ainsi qu la nécessité grandissante de méthodes moins invasives. La transition vers la technologie sans contact et les contraintes associées au capteur sont ensuite détaillés. Le chapitre 3 est dédié à l'étude des méthodes de l'état de l'art. Premièrement, les différents modèles d'interactions lumière-tissu utilisés dans la littérature sont présentés ainsi que les méthodes de rPPG qui les exploitent. Puis, dans un second temps, les verrous scientifiques concernant la mesure rPPG sont identifiés par l'analyse des contraintes liées aux méthodes employées. Ceux-ci permettent finalement la définition de la problématique de cette thèse qui sera le guide des travaux présentés en parties 2 et 3.

La seconde partie présente les travaux qui ont été menés afin d'une part d'effectuer la mesure rPPG par segmentation implicite de la région d'intéret et d'autre part d'améliorer la qualité de l'information rPPG. Dans le chapitre 4, la méthode de segmentation par superpixels temporels [70] employée est présentée ainsi que la méthode d'estimation de la qualité du signal rPPG à partir d'un modèle fréquentiel et d'un modèle temporel, et enfin la méthode de fusion de données qui est employée afin de maximiser l'information rPPG. Le chapitre 5 est articulé autour des modalités de l'expérimentation en présentant les méthodes de l'état de l'art implémentées pour comparaison ainsi que les données de test employées. Les conditions de tests, implémentations des méthodes et choix des métriques, sont alors abordées et justifiées. Enfin les résultats et conclusions de l'expérimentation sont présentés afin de valider la méthode.

La troisième et dernière partie présente des travaux menés afin de réaliser l'implémentation de la méthode de mesure rPPG proposée. Pour cela, une nouvelle méthode de segmentation par superpixels mis en oeuvre afin de réduire le temps de calcul nécessaire à la segmentation du flux vidéo a été proposée et est détaillée dans le chapitre 6. Cette nouvelle méthode est évaluée sur les aspects de la qualité de la segmentation en comparaison avec les méthodes de l'état de l'art ainsi que du point de vue de la qualité de l'estimation du rythme cardiaque [139]. Le chapitre 7 permet de détailler l'implémentation sur une cible matérielle de type processeur graphique (GPU) de la partie la plus critique de notre algorithme, du point de vue ressource de calcul, à savoir la segmentation de la zone utile. Les performances d'implémentation sont comparées à partir de deux plateformes de puissances de calcul très différentes afin de déterminer les limitations propres à la méthode mais également celles imposées par les plateformes matérielles. Les optimisations ont permis de répondre aux contraintes temps réel nécessaires à la réalisation d'un capteur rPPG.

I

État de l'art et problématique

2

Les systèmes photopléthysmographiques

Sommaire

2.1 Principes de fonctionnement 12
2.2 Limites de la photopléthysmographie au contact 15
2.3 Du point aux pixels : la photopléthysmographie sans contact 16
2.3.1 Impact du capteur 18
2.3.2 Capteur matriciel : mesures multi-points pour la pression artérielle . . . 19
2.3.3 Longueurs d'onde pour la rPPG 20

La photopléthysmographie au contact (PPG) [7] ou sans contact (rPPG) [47] permet l'estimation de paramètres physiologiques en mesurant respectivement les variations de l'information lumineuse transmise ou retrodiffusée par la peau, à l'aide d'un capteur photoélectrique monochromatique ou polychromatique. Ces variations sont induites par la variation temporelle du volume sanguin dans les tissus microvascularisés présents sous la peau [1,2,6,33]. La mesure s'appuie sur le fait que le sang présent dans les artères a des niveaux d'absorption différents suivants certaines longueurs d'onde. De plus, ces variations sont principalement induites par l'évolution du volume sanguin au regard des autres éléments biologiques [11].

Dans ce chapitre, nous détaillerons la méthode optique de mesure du rythme cardiaque, appelée photopléthysmographie. Dans un premier temps, le fonctionnement des capteurs au contact sera détaillé ainsi que leurs limitations. Puis nous décrirons le fonctionnement de la technologie sans contact. Les différents types de capteurs matriciels et leur impact sur la mesure fut également un sujet d'étude dans le développement de la méthode rPPG, nous en présenterons les conclusions dans un second temps. Enfin, l'étude de la réponse en quantité d'information rPPG en fonction de la longueur d'onde représente l'une des études pionnières de la technologie et est donc détaillée à ce niveau. Ces éléments permettent de mieux identifier les phénomènes physiques en cause et ainsi de permettre une meilleure compréhension des enjeux et limitations de la technologie.

2.1/ Principes de fonctionnement

En se basant sur le modèle dichromatique, il est couramment considéré dans le cas de la (r)PPG [6, 123] que la peau réfléchit la lumière grâce à deux phénomènes. La partie spéculaire de la lumière réfléchie est constituée de la lumière directement réfléchie par la surface de la peau. La partie diffuse de la lumière réfléchie est constituée de la lumière non-absorbée par les tissus puis ré-émise jusqu'en surface (cf. Figure 2.1). En fonction de la longueur d'onde, les rayons lumineux peuvent interagir plus ou moins profondément dans les tissus.

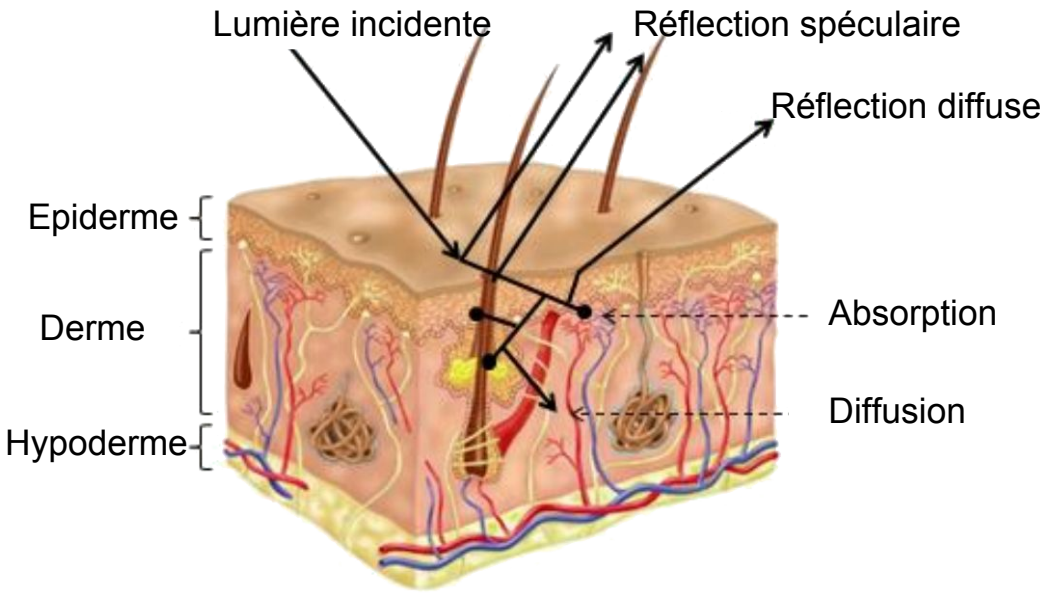


Figure 2.1 – Représentation de l'interaction lumière - peau.

Les oxymètres de pouls permettent la mesure de l'information photopléthysmographique au contact en transmittance ou réflectance. La mesure en transmittance se fait au doigt ou à l'oreille par l'utilisation d'une source lumineuse, LED, et d'un photorecepteur placé de part et d'autre du volume de tissus utilisé pour la mesure. La lumière transmise par la source lumineuse est modulée par le volume d'hémoglobine dans les capillaires sous la peau puis captée par le photorécepteur. En reflectance, la source lumineuse et le photorécepteur sont placés côte à côte. La lumière directement réfléchie par la peau, qui ne contient pas d'informations de pulsation, est également mesurée. Elle devient une source de perturbation [35].

Initialement observées en 1937 par Hertzman et Spealman [1], les variations du volume sanguin mesurées par transmittance au travers du doigt à l'aide d'un élément photoélectrique ont permis la mise en évidence du phénomène photopléthysmographique. Un an plus tard, la mesure de ce phénomène a été réalisée sur plusieurs surfaces de peau [2]. Les applications d'une telle technologie ont permis l'arrivée de capteurs efficaces tout d'abord en transmittance [4] puis en réflectance [6, 7]. Dans le cas de capteurs PPG au contact en transmittance, la sonde est généralement placée au doigt ou à l'oreille. Le volume de tissus illuminés est alors faible. Cependant ces derniers étant très irrigués, la lumière transmise au travers des tissus est également modulée, et par conséquent la mesure peut être réalisée.

Les contraintes techniques d'un tel dispositif sont actuellement bien identifiées, pour autant, elles sont non négligeables. La modulation est plus importante pour une certaine gamme de longueurs d'onde, par conséquent les propriétés spectrales de la source lumineuse sont primordiales. De plus, l'intensité de la source doit également être prise en compte pour obtenir suffisamment d'information au vu des pertes dues à la réflexion spéculaire et à l'absorption de la lumière dans les tissus [3].

Le signal mesuré a une apparence caractéristique qui met en évidence les phénomènes de systole et de diastole, Figure 2.2.

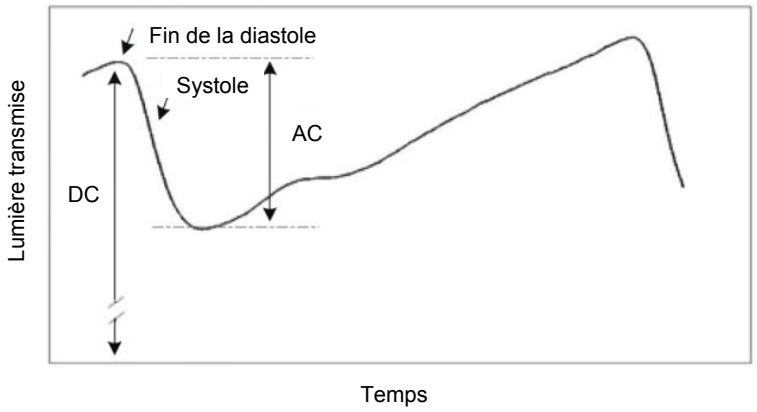


FIGURE 2.2 – Forme caractéristique du signal PPG avec les phénomènes de diastole et de systole [84] pour une mesure PPG en transmitance.

La systole est une phase qui correspond à l'expulsion du sang dans le système vasculaire. L'intensité de la lumière transmise décroit durant la systole. En effet, le volume sanguin augmente alors dans les capillaires, et par conséquent la quantité de lumière absorbée. Avec le reflux cardiaque durant la phase de diastole, le niveau d'absorption diminue et il est possible d'observer le signal PPG caractéristique ci-dessus (cf. Figure 2.2).

En 1990, Cui et al. [8] montrent que les longueurs d'onde situées entre 510 et 560 nm, qui correspondent à la couleur verte dans le spectre visible, permettent la mesure la plus efficace en réflectance et sont celles utilisées dans les systèmes modernes [40] afin de mesurer le rythme cardiaque uniquement.

Les études portant sur les caractéristiques de la source lumineuse et du capteur associé, ont permis de déterminer les longueurs d'onde utiles à l'observation du signal PPG. La Figure 2.3 permet de visualiser le taux d'absorption des deux types d'hémoglobine en fonction des longueurs d'onde. Dans le cas où l'hémoglobine est chargée en oxygène, elle est appelée oxyhémoglobine et est notée HbO_2, et dans le cas où l'hémoglobine est dissociée de l'oxygène, elle est appelée déoxyhémoglobine et est notée Hb.

En fonction des besoins de la mesure, d'autres longueurs d'onde sont employées. Ainsi, deux gammes de longueurs d'onde sont habituellement employées pour l'estimation de la saturation en oxygène dans le sang (SpO_2), une première située entre 600 et 750 nm (correspondant au rouge) et une seconde entre 850 et 1000 nm (correspondant au proche infrarouge) [84]. Comme les réponses en absorption de l'oxyhémoglobine et de la déoxyhémoglobine ne sont pas égales pour une même longueur d'onde, il est possible d'établir la concentration en oxygène dans le sang SpO_2 (exprimée en %) :

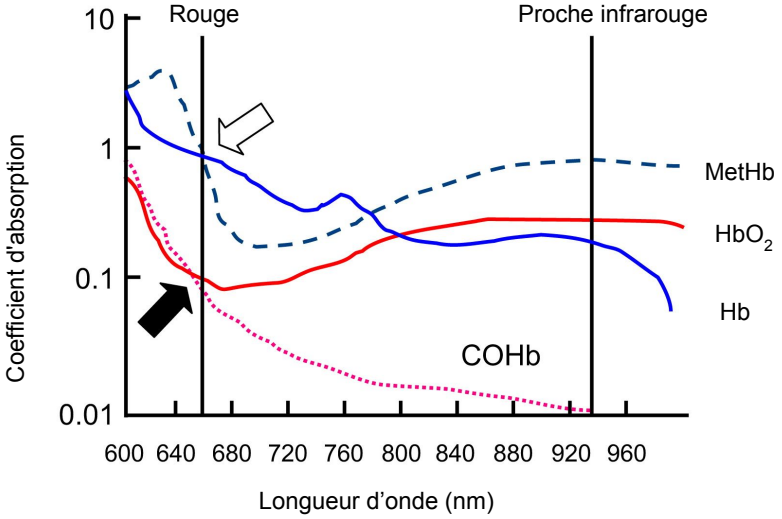


FIGURE 2.3 – Courbes d'absorption de l'oxyhémoglobine (HbO_2), de la déoxyhémoglobine (Hb), de la carboxyhémoglobine ($COHb$) et de la méthémoglobine ($MetHb$) en fonction de la longueur d'onde [69].

$$SpO_2 = \frac{HbO_2}{HbO_2 + Hb}. \tag{2.1}$$

Dans le cadre de l'oxymétrie au contact, on travaille avec une source polychromatique rouge et infrarouge et on fait le ratio des ratios des réponses obtenues avec le capteur pour établir cette concentration [84] :

$$R = \frac{(AC/DC)_1}{(AC/DC)_2}, \tag{2.2}$$

où le ratio $(AC/DC)_1$ est le rapport des composantes alternatives et continues du signal PPG (cf. Figure 2.2), modulé pour les longueurs d'onde comprises entre 600 et 750 nm (rouge). Le ratio $(AC/DC)_2$ correspond à la même opération considérant les longueurs d'onde comprises entre 850 et 1000 nm (infrarouge). Une fois ce rapport obtenu, on peut établir une fonction reliant le SpO_2 et R de manière analytique grâce au modèle de Beer-Lambert [54] ou par calibration avec par exemple la fonction suivante :

$$SpO_2 = \frac{k_1 - k_2 \cdot R}{k_3 - k_4 \cdot R}, \tag{2.3}$$

où les valeurs k_i représentent des paramètres qui sont obtenues à la calibration de l'oxymètre de pouls. Les dérivés d'hémoglobine qui ne sont plus capables de fixer l'oxygène comme la méthémoglobine, noté MetHb, et la carboxyhémoglobine, noté COHb, contribuent à la modulation du signal PPG et peuvent donc altérer la mesure. Leurs concentrations sont négligées dans les appareils qui utilisent uniquement deux sources lumineuses rouge et infrarouge. Dans le cas de personnes en mauvais état de santé, ou qui souffrent d'un empoisonnement au monoxyde de carbone, la mesure peut être faussée.

Tous les appareils modernes présents dans le cadre médical fournissent au minimum le rythme cardiaque ainsi que la saturation en oxygène et sont donc équipés de source lumineuse rouge et proche

infrarouge. Au contraire, les appareils destinés au grand public comme les montres connectées sont en général équipés d'une unique source lumineuse verte pour la mesure cardiaque. Dans leurs travaux parus en 2011, Maeda et al. [51] évaluent les différences au niveau de la mesure en estimant le rythme cardiaque avec les longueurs d'onde correspondantes au vert ou celles correspondantes au proche infrarouge. Leurs conclusions montrent que les mesures effectuées à l'aide d'une source de lumière verte sont les plus précises dans le cas d'une mesure au contact, et présentent une robustesse accrue aux perturbations. Ce résultat est en accord avec les travaux menés jusqu'ici, considérant les niveaux d'absorption plus important dans les longueurs d'onde 510 - 560 nm qui permettent des amplitudes plus importantes dans la mesure.

La technologie a continué d'évoluer pour faciliter sa mise en oeuvre et accroitre sa fiabilité [35]. En 2009, Vogel et al. [46] publient leurs travaux sur le développement d'un dispositif de mesure PPG intra-auriculaire sans fil pour la prévention des risques liés aux maladies cardiovasculaires. Il est muni d'une source polychromatique rouge et infrarouge et permet la mesure continue du rythme cardiaque et de la concentration en oxygène dans le sang. Ce développement marque l'ouverture à de nouvelles problématiques pour la mesure au contact considérant un grand niveau de perturbations supplémentaires dans la mesure des signaux qui peuvent être induits par la marche, la mastication ou plus généralement les mouvements réalisés par le sujet d'observation.

2.2/ Limites de la photopléthysmographie au contact

Bien que très efficace, la mesure photopléthymographique présente des limitations importantes. Tout d'abord la mesure optique estime le rythme des pulsations du volume sanguin et non pas celle du rythme cardiaque. En effet, le signal PPG étant issu de la modulation de la lumière par les variations du volume sanguin dans les capillaires et artérioles de la peau, c'est bien la pulsation du volume de sang qui est mesurée. Le rythme cardiaque, qui correspond aux contractions du muscle cardiaque est mesuré à l'aide d'un ECG avec comme données d'entrée l'activité électrique du muscle cardiaque. Or, si ces mesures sont habituellement identiques, en cas de troubles circulatoire, le cœur peut ne pas correctement propager le sang dans le réseau circulatoire périphérique. Dans ce document toutefois nous utilisons la dénomination "rythme cardiaque" par commodité en lieu et place de "rythme des pulsations du volume sanguin". Comme les mesures effectuées dans nos expérimentations ne se font qu'à partir de capteur optique, il n'y a pas de confusion possible, cependant nous préférons préciser qu'il s'agit d'un abus de langage. Aussi, il n'est pas possible d'envisager un remplacement complet des méthodes de surveillance cardiaque actuelles par des méthodes optiques puisque les informations mesurées ne sont pas les mêmes. Dans le cas d'une surveillance critique ou pour la détection de pathologies, l'ECG reste la référence [23]. La méthode optique souffre également énormément des perturbations induites par les mouvements, ce qui est problématique pour certaines applications, comme par exemple, dans le cas d'une surveillance d'une activité sportive. Dans le cadre d'expérimentations, l'ECG, fixée en ceinture sur la poitrine, demeure la référence de terrain privilégiée [71, 136].

Comme présentées en introduction, les mesures PPG et ECG souffrent de la nécessité de contact avec la peau. Tout d'abord la position et le placement des électrodes ou de pinces peuvent causer gênes et inconfort au patient, tout comme l'utilisation d'adhésif pour maintenir ces électrodes qui ajoute douleur et stress. Les très jeunes enfants, typiquement avant 27 semaines après naissance, peuvent souffrir de lésions cutanées lors de l'utilisation de ces capteurs en raison d'une plus forte adhérence entre l'adhésif et la peau que celle existante entre le derme et l'épiderme [19]. On retrouve une situation similaire dans le cas de patients souffrant de brulures ou encore par exemple pour certaines personnes âgées. De plus pour les oxymètres de pouls, la taille du capteur étant

souvent mal adaptée aux prématurés, une grande majorité des alertes sont dues à la mauvaise connexion voire à la perte de contact entre le capteur et l'enfant suite à des mouvements. La pression exercée par le mécanisme de maintien peut également provoquer des lésions ou des gênes.

Du point de vue de la mesure, les études ont montrées qu'une pression modérée du capteur sur la peau pouvait améliorer la qualité du signal dans le cas de sonde au contact en réflectance. Cependant, une pression excessive, qui survient très facilement dans le cas de patient fragile, peut créer une occlusion temporaire des capillaires sous-cutanés et ainsi nuire à la mesure [12].

Teng et al. [28] ont apporté une première observation de l'impact de la pression d'un oxymètre de pouls utilisé au doigt. Ces test ont été réalisés sur 15 patients. Pour une force appliquée variant de 0.2 Newton (N) à 1.0 N, la composante continue DC augmente et la composante alternative AC du signal PPG est largement atténuée. Cela se traduit par une augmentation initiale puis une chute importante du ratio AC/DC. Pour différents sujets, la perte d'information peut apparaître dès 0.2 N. En moyenne les meilleures valeurs sont atteintes pour une force comprise entre 0.2 N et 1 N (cf. Figure 2.4), ce qui correspond à une force exercée faible pour maintenir un capteur.

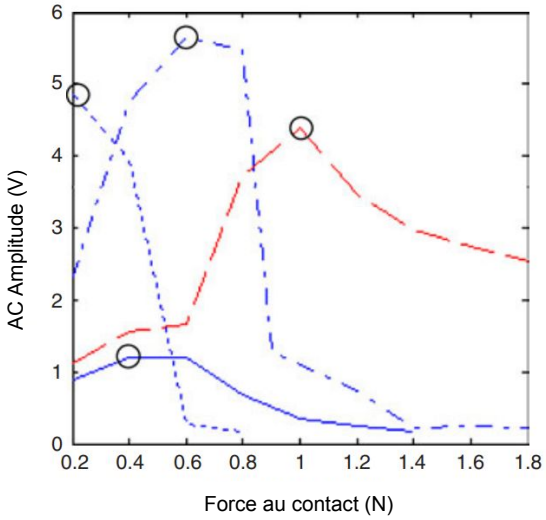


FIGURE 2.4 – Exemple d'amplitude AC, interprétable comme une mesure de qualité du signal, pour 4 sujets en fonction de la force au contact. Les cercles indiquent une amplitude maximale [28].

En conclusion, les limitations principales de la méthode PPG sont très majoritairement liées à la nécessité de contact inhérente à la mesure. L'apport d'une solution sans contact apparaît par conséquent significatif. De plus, l'utilisation d'une méthode de type PPG ne permet pas actuellement des mesures aussi précises qu'avec un ECG. Une méthode de type rPPG apparaît alors comme une solution à privilégier afin d'effectuer une surveillance cardiaque non critique, tout en assurant le confort du patient et du personnel médical.

2.3/ DU POINT AUX PIXELS : LA PHOTOPLÉTHYSMOGRAPHIE SANS CONTACT

Les premières contributions significatives dans le domaine de la mesure photopléthysmographique à distance sont de Such et al. [13,14]. Le capteur ponctuel est remplacé par une matrice, une caméra CCD proche infrarouge. Les applications de la technologie sont alors plus orientées vers la dermatologie et l'estimation du niveau de perfusion des tissus. En 2000, Wu et al. [18] introduisent une nouvelle méthode afin de visualiser les variations de perfusion et en déduire des dysfonctionne-

ments du système vasculaire en cartographiant le réseau veineux hémodynamique et la perfusion du réseau artériel. En 2002, Huelsbusch et Blazek [22] améliorent la méthode en utilisant la transformée en ondelettes afin d'isoler des caractéristiques correspondant aux rythmes biologiques. Zheng et al. [38, 44] introduisent pour la première fois l'utilisation de plusieurs types de capteurs, passants du couteux CCD au plus abordable CMOS pour visualiser le niveau de perfusion des tissus. Ils proposent la première implémentation multispectrale, 660 et 880 nm, à destination de la dermatologie en 2008 [43]. La technologie précédemment nommée PPGI devient alors rPPG.

Basés sur le modèle de la PPG par refléctance, Takano et al. [36] et Verkruysse et al. [42] mettent en évidence la capacité à mesurer le signal photopléthysmographique à distance, cette fois à destination de la mesure cardiaque. Ce sont les premières implémentations à distance qui permettent la mesure de rPPG telle qu'elle est faite aujourd'hui. Dans leurs approches, les sources lumineuses sont remplacées par la lumière ambiante ou bien par des sources de lumières polychromatiques. Le capteur devient une caméra et la source d'information passe de la mesure d'un point unique à matricielle.

En 2010 Poh et al. [47] proposent une méthode d'observation à l'aide d'une caméra RGB et mettent en place la chaîne algorithmique classique utilisée actuellement dans la littérature. La Figure 2.5 montre l'installation nécessaire à la mesure rPPG. Le sujet est assis à une distance d'environ un mètre d'un PC portable, le capteur d'entrée est la caméra RGB intégrée de l'ordinateur, le flux vidéo est enregistré sans compression pour éviter toute perte d'information. Nous détaillons ici l'expérimentation comme elle permet d'introduire la chaîne algorithmique employée dans toutes les méthodes modernes.

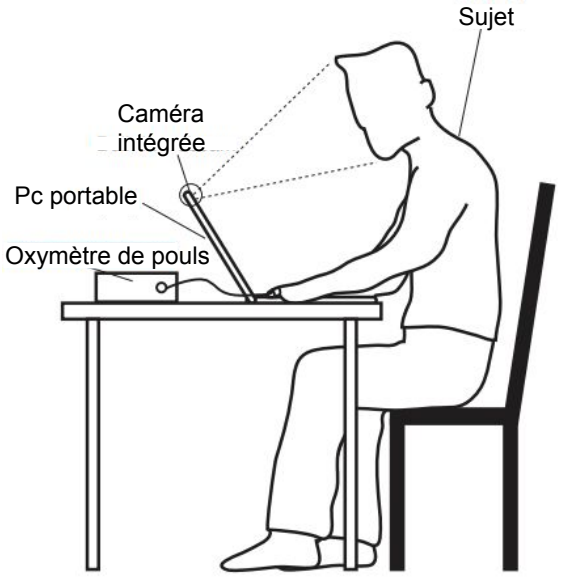


FIGURE 2.5 – Installation pour la mesure rPPG telle que présenté par Poh et al. [47].

L'installation ne comprend pas l'apport d'une source lumineuse : c'est la lumière ambiante qui est utilisée. La vérité terrain est établie à l'aide d'un oxymètre de pouls. La validation de la technologie se fait par l'estimation du rythme cardiaque comparée entre le signal PPG acquis au contact et le signal rPPG acquis à distance. Pour chaque image du flux vidéo, une détection de visage [20] est réalisée afin de segmenter une zone d'intérêt. Les pixels considérées dans cette dernière sont moyennés pour chaque composante chromatique, ici Rouge, Vert et Bleu à chaque nouvelle image. Les signaux $x_1(t)$, $x_2(t)$ et $x_3(t)$ sont respectivement acquis pour les canaux rouge, vert et bleu sur une fenêtre temporelle glissante de 30 secondes. Les données sont ajoutées successivement avec un

chevauchement des fenêtres temporelles de 96.7%, ce qui correspond à un incrément de 1 seconde. Une fois les traces acquises, les signaux $x_i(t)$ sont normalisées :

$$x'_i(t) = \frac{x_i(t) - \mu_i}{\sigma_i}. \tag{2.4}$$

Avec $x_i(t)$ les signaux issus des canaux R, G et B, μ_i leurs valeurs moyennes et σ_i leurs écarts types. Les signaux $x'_1(t)$, $x'_2(t)$ et $x'_2(t)$ sont alors centrés en 0 et ont une variance unitaire. Afin de déterminer le signal rPPG compris dans les trois signaux, Poh et al. utilisent une analyse en composantes indépendantes (ICA). La méthode ICA, détaillée plus loin (3.2.1), est une méthode de traitement du signal qui permet d'isoler des sources indépendantes à partir de plusieurs signaux d'entrée. En sortie, la méthode fournie autant de signaux qu'il y a de signaux d'entrée. Pour simplifier l'expérimentation, Poh et al. considèrent dans ces premiers travaux que le signal est forcément contenu dans la seconde composante résultante. L'estimation du rythme cardiaque est faite en utilisant une transformation de Fourier considérant que le signal rPPG est pseudo-périodique et devrait donc avoir une importante énergie pour la principale fréquence qui le constitue. La recherche de maxima est bornée à la bande passante [0.75, 4] Hz, ce qui correspond à la plage de rythmes cardiaques [45, 240] bpm. Le rythme cardiaque est donné pour le pic de plus grande énergie dans cette bande de fréquence. Les résultats montrent une grande précision dans l'estimation du rythme cardiaque et ouvrent la porte aux nouvelles méthodes et des utilisations originales de cette technologie.

2.3.1/ Impact du capteur

A l'instar des développements initiaux dans le cadre de la PPGI, plusieurs configurations ont été testées afin de valider l'impact de l'usage de différents types de capteurs. En 2011, Poh et al. [53] et Sun et al. [56] publient leurs travaux sur l'utilisation de caméras entrée de gamme, type webcam PC, et la capacité de ces caméras à détecter les variations subtiles de couleur de la peau induites par le signal rPPG. Les résultats annoncés alors sont très prometteurs et ouvrent la voie aux méthodes actuelles de mesures. L'acquisition est faite par un capteur standard CMOS polychromatique RGB avec des fréquences d'acquisition de 25 à 30 Hz. Le signal extrait permet de mesurer le rythme cardiaque avec une bonne précision. Les caméras haut de gamme CCD sont également utilisées en raison de leur faible bruit intrinsèque au capteur [123] mais ne sont pas un pré-requis pour obtenir une mesure fiable et leur utilisation reste minime [63].

Bien que mesurable avec des caméras bas de gamme, le signal photopléthysmographique peut aisément être corrompu en fonction des conditions d'acquisition et des règlages de l'appareil. Les fonctions automatiques telles que la correction automatique des blancs, réglage du gain automatique, réglage du temps d'exposition automatique ou de l'auto-focus sont autant de perturbations dans l'acquisition qui, si elles permettent une image dont l'aspect visuel est plus plaisant pour l'utilisateur, peuvent altérer les variations minimes que l'on souhaite extraire dans le cas de la mesure rPPG [85]. En effet, les modifications induites par les corrections automatiques d'une caméra ne sont pas constantes sur de longues périodes d'observation, aussi les perturbations du signal RGB peuvent masquer les faibles variations que l'on souhaite mesurer.

Les formats de compression vidéo des encodeurs courants tendent aussi à faire disparaitre ces variations. Comme les variations sont temporelles et qu'elles sont minimes, elles peuvent être simplement éliminées à cause de la compression du signal vidéo ou perturbées par les artefacts de compression [112]. Certaines caméras ne permettent pas une acquisition de l'image brute. Suivant la résolution de l'image et la fréquence d'acquisition (4K à 30Hz pour certaines caméras récentes),

la bande passante nécessaire à la transmission d'une telle quantité d'information est incompatible avec l'interface de communication disponible employée sans compression. Ainsi il est courant de voir des systèmes qui encodent l'information RGB en sortie du capteur, avec un codeur vidéo matériel H.264 par exemple, avant de le décoder pour le fournir à l'utilisateur. Malheureusement, l'information rPPG est altérée par l'encodage. Enfin la fréquence d'acquisition de la caméra doit rester stable, autant que possible, pour éviter l'apparition de bruit fréquentiel lors de l'analyse du signal rPPG [136]. Une méthode simple pour éviter ce genre de problème réside dans l'utilisation d'une interpolation linéaire du signal, la valeur temporelle est enregistrée lors de l'acquisition pour permettre un rééchantillonage régulier afin d'améliorer la qualité de l'analyse. Bien que les variations de couleurs à mesurer soient faibles, une dynamique de 8 bits par canaux est suffisante pour permettre la mesure [123]. La résolution spatiale de l'image n'est pas non plus un facteur primordial tant que le capteur parvient à capturer suffisamment de pixels de peau afin de compenser le bruit de quantification de la caméra. En 2015, Blackford et al. [91] réalisent une acquisition sur 25 sujets à l'aide de 9 caméras disposées en arc de cercle tous les 20°, à une distance d'1.5 m et à hauteur du visage. L'éclairage est contrôlé et les variables sont la fréquence d'échantillonnage et la résolution spatiale de la zone d'intérêt. Pour une fréquence d'échantillonnage variant de 30 à 120 images par seconde et une résolution spatiale de 658x492 pixels à 329x246 pixels, soit un ratio de 1/4. Les résultats montrent des différences non significatives. Aussi l'auteur conclut que l'acquisition à l'aide d'une caméra unique possédant une fréquence images d'au moins 30 images par seconde, et ce quelle que soit la résolution spatiale, est suffisant pour réaliser une mesure dans les meilleures conditions. Ainsi en respectant ces critères un grand nombre de caméras commerciales, de différentes qualités, peuvent être considérées : les webcams [53, 119, 123], les téléphones portables [59] ou encore les Google Glass [97].

2.3.2/ Capteur matriciel : mesures multi-points pour la pression artérielle

L'utilisation d'un capteur matricielle permet d'effectuer assez directement une mesure multi-points. Ainsi, Sun et al. [63] mettent en évidence une mesure du "pulse transit-time" (PTT) qui est une information temporelle liée à la propagation spatiale du signal photopléthysmographique. Pour ce faire, ils exploitent la mesure simultanée faite à partir d'un signal rPPG avec une acquisition vidéo du visage et une autre depuis un capteur PPG au doigt. Cette valeur permet l'obtention de la vitesse de propagation spatiale de l'onde de pouls qui est théoriquement proportionelle à la pression artérielle. Pour parvenir à réaliser ce genre de mesure Sugita et al. [104] utilisent une caméra haute vitesse, 140 images par seconde, afin d'obtenir le maximum de précision dans l'estimation du PTT. Les zones d'intérêt sélectionnées sont : la paume de la main droite, le front, et les pommettes. Le signal est ainsi quasi synchrone entre le front et la pommette comme les zones sont spatialement proches. La main étant beaucoup plus loin du cœur, la propagation est potentiellement observable. L'estimation du PTT est remplacée par le calcul de la différence de phase instantanée depuis les différents sites avec une transformée de Hilbert. La vérité terrain est définie par un capteur au contact attaché au doigt. Les trois comparaisons de phases sont faites entre :

— PD_1 : front et pommette
— PD_2 : paume main droite et pommette
— PD_3 : paume main droite et front

La pression artérielle systolique est utilisée comme comparaison. Elle est artificiellement augmentée en demandant aux sujets de retenir leur respiration.

Les résultats, présentés Figure 2.6, montrent qu'un niveau d'intercorrélation de quasi 70% peut être obtenu entre les mesures réalisées respectivement au niveau de la main et du front. Cette

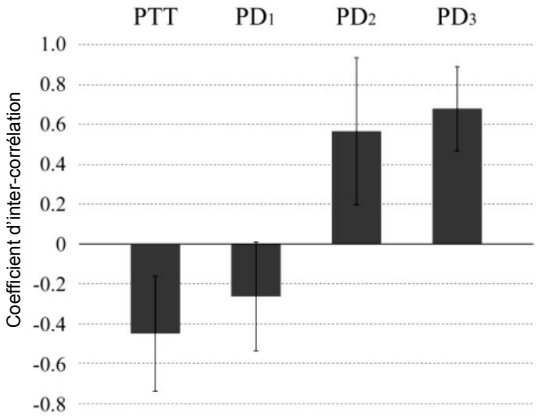


FIGURE 2.6 – Moyenne des coefficients d'inter-corrélations de la pression artérielle systolique et des variations de phases PD_i [104].

expérimentation démontre la faisabilité d'une telle mesure sans contact à l'aide de la technologie rPPG. Cependant, il n'est toujours pas clairement établi que la mesure effectuée est uniquement fonction de la modulation par l'interaction avec l'hémoglobine. En effet, les battements cardiaques et la circulation du sang dans les artères provoquent une déformation élastique de celles-ci que l'on appelle *onde de pouls*. Ces déformations peuvent contribuer ou affecter la modulation observable [133]. Une des conséquences directes de ce phénomène est que la phase à l'origine des signaux rPPG mesurés peut varier de façons non négligeables pour des zones de peau spatialement proches. Comme les méthodes de rPPG utilisent un calcul de la valeur moyenne des pixels de peau observée afin de s'affranchir du bruit de quantification de la caméra, les variations de phase peuvent nuire à la qualité de l'information. Comme le signal rPPG est toujours issu du calcul de la moyenne des pixels dans une zone d'intérêt, la phase à l'origine du signal résultant n'est pas clairement définie. Dans l'attente de plus amples expérimentations, il apparait nécessaire de rester prudent sur les capacités des méthodes actuelles sur ce type de mesure.

2.3.3/ LONGUEURS D'ONDE POUR LA RPPG

A l'instar de la PPG, le signal rPPG est plus facilement mesurable pour les longueurs d'onde situés entre 510 et 560 nm, qui correspondent à la couleur verte du spectre visible [40, 51]. Cependant, dans le cas de l'utilisation de capteurs et de sources lumineuses polychromatiques RGB, il apparait que le signal rPPG est également mesurable dans les canaux rouge et bleu. En 2011, Corral et al. [52] ont ainsi étudié la qualité de la réponse rPPG en fonction de la longueur d'onde.

Ils ont ainsi mesuré le signal rPPG au niveau du front de 7 sujets différents. La source lumineuse utilisée est une lampe quartz-halogène qui émet des radiations dans le spectre visible et proche infrarouge. Pour limiter la zone d'observations, un carton noir est placé devant le visage des sujets de sorte que seule la zone de peau soit visible. L'élément qui sert à concentrer la lumière est une lentille destinée à la photographie, de focale f=80mm et $f/3.5$. L'extrémité d'une fibre optique placée dans le plan focal de la lentille, collecte la lumière émise au niveau du front puis la lumière est acheminée jusqu'à un spectromètre (Ocean Optics QE6500) qui permet la mesure pour la bande passante [200, 980] nm. La lentille optique est placée à 2 m du sujet (cf. Figure 2.7).

Le signal est enregistré durant une durée de 60 secondes pour chacun des volontaires avec une fréquence d'échantillonnage de 10 spectres complets par seconde. La vérité de terrain est établie avec un oxymètre de pouls du commerce. Pour extraire le signal cardiaque, le signal brut d'entrée

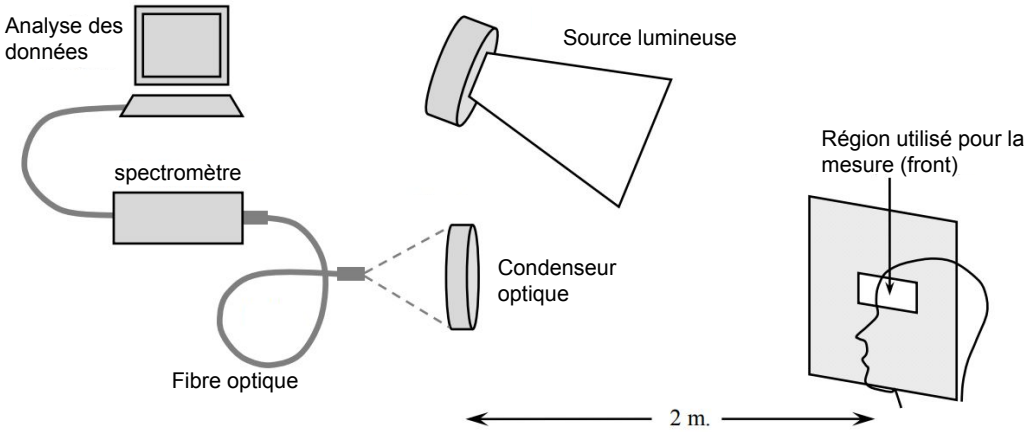


FIGURE 2.7 – Dispositif expérimental utilisé pour déterminer la qualité de la réponse du signal rPPG pour l'estimation du rythme cardiaque et du rythme réspiratoire en fonction de la longueur d'onde [52].

est filtré avec un filtre passe haut de fréquence de coupure $f_c = 0.416$ Hz (25 bpm).

L'analyse des signaux pour les 783 longueurs d'onde d'entrées situées entre 380 et 980 nm permet alors d'établir la correspondance entre la qualité de l'estimation rPPG et la longueur d'onde (cf. Figure 2.8). Les conclusions sont que le signal rPPG doit être mesuré entre les longueurs d'onde 480 et 610 nm ou entre 800 et 925 nm pour permettre un ratio signal/bruit suffisant.

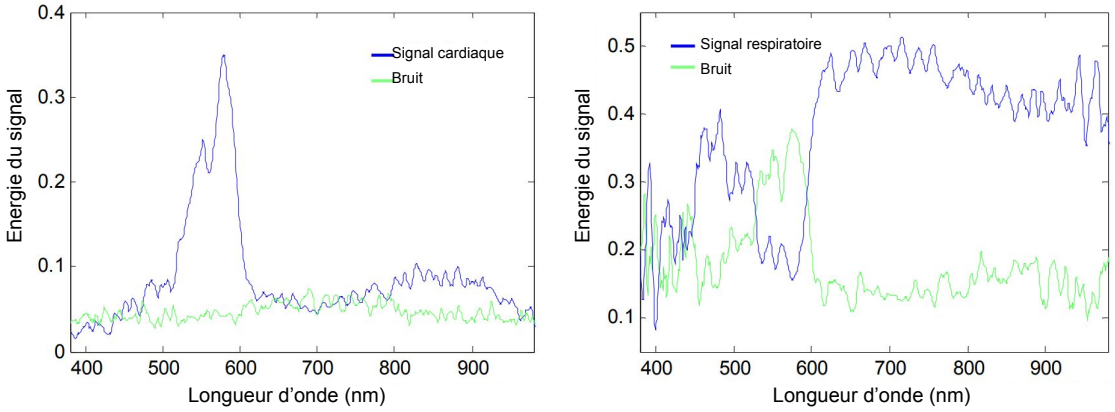


FIGURE 2.8 – Énergie du signal cardiaque (gauche) et du signal respiratoire (droite) en comparaison à l'énergie du bruit mesurée en fonction de la longueur d'onde [52].

Ces travaux sont fondamentaux car ils permettent une bien meilleure compréhension de l'impact du choix des longueurs d'onde à mesurer sur la qualité de la mesure. Il est ainsi possible de corréler cette information avec la réponse RGB d'une caméra et la réponse en réflectance de la peau afin de déterminer la contribution de chaque canal au signal rPPG (cf. Figure 2.9). Cette étude démontre la robustesse d'une mesure de rPPG basée sur l'utilisation du canal vert.

En 2014, McDuff et al. [83] implémentent la mesure rPPG à l'aide d'une caméra RGBCO qui comprend les canaux Cyan et Orange en plus des traditionnels Rouge, Vert et Bleu. L'origine de l'augmentation de la qualité de la mesure réside dans le fait que le nombre d'éléments de mesure

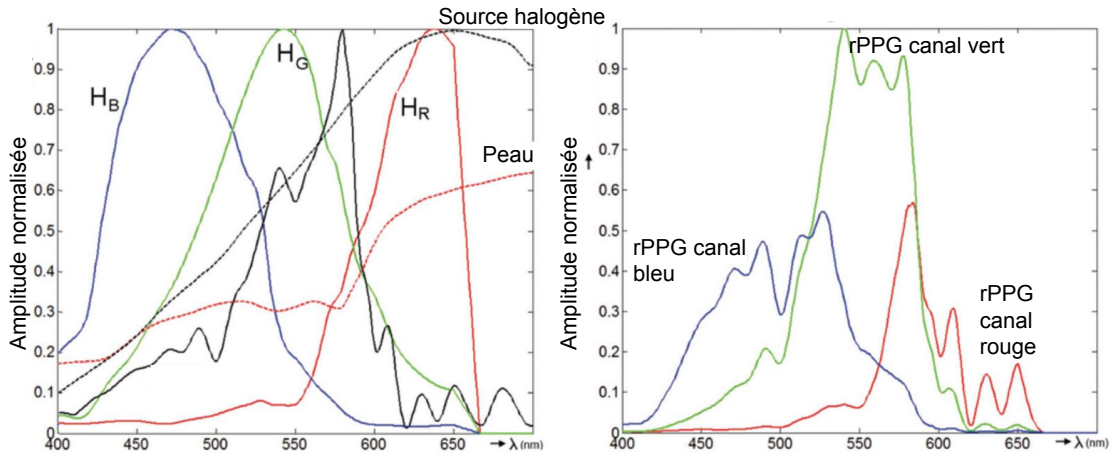


FIGURE 2.9 – Exemple de l'estimation de la qualité de la réponse rPPG pour les 3 différents canaux R, G et B d'une caméra polychromatique UI-2220SE-C, vis à vis de la réponse en réflectance de la peau, de la courbe de réponse établie Figure 2.8 et avec une source de lumière halogène [80]. La convolution des différentes courbes (à gauche) permet d'établir l'amplitude normalisée du signal rPPG en fonction des canaux de la caméra utilisé pour la mesure (à droite).

du signal rPPG augmentent avec le nombre de canaux. Tout comme avec l'utilisation initial d'un unique canal pour effectuer la mesure était sous optimale comme le signal rPPG est également mesuré dans les canaux rouge et bleu, l'utilisation d'un plus grand nombre de canaux permet de réduire d'avantage le bruit intrinsèque au capteur en moyennant les informations réparties sur tous les canaux. En effet, la statistique est alors plus favorable et par conséquent le bruit de quantification impacte moins la mesure [129].

Comme l'information pulsative ne contribue pas de manière identique au sein des différents canaux couleurs de la caméra, de nombreuses études ont été menées afin d'exploiter cette différence pour isoler l'information rPPG. Ainsi à titre d'exemple, une méthode très simple, mais étonnamment performante, consiste à réaliser la différence entre les canaux verts et rouges [39]. Cette méthode permet d'obtenir de bons résultats dans l'estimation du rythme cardiaque en comparaison de méthodes bien plus complexes. On peut expliquer cela en se référant à la réponse du signal rPPG en fonction des canaux, présentée Figure 2.9. Hulsbusch et al. [39] établissent que les perturbations induites par les variations lumineuses ou les mouvements sont en partie compensés par la différence de ces canaux puisque le signal rPPG est très faible dans le canal rouge alors que les perturbations sont supposées avoir une contribution égale dans tous les canaux.

Afin de réaliser la mesure rPPG, les travaux disponibles dans la littérature utilisent le modèle dichromatique et expriment le signal rPPG en fonction de l'information lumineuse spéculaire et diffuse ainsi que l'impact du mouvement sur ces composantes. Le Chapitre 3 permet de présenter les modèles d'interaction lumière-tissus employés ainsi que les méthodes de rPPG de l'état de l'art.

3

Concepts et méthodes rPPG existantes

Sommaire

3.1 Modèles d'interactions lumière-tissu pour la rPPG 24
3.2 Algorithmes rPPG existants 26
3.2.1 Algorithmes par séparation de sources 27
3.2.2 Algorithmes par projection 29
3.3 Problématiques et objectifs de la thèse 34

Les méthodes de rPPG peuvent être classées en deux grandes catégories. Premièrement, l'utilisation dans la chaîne algorithmique des méthodes par séparation de sources (*Blind Source Separation* BSS) comme l'analyse en composantes principales (*Principal Component Analysis PCA*) [49] ou l'analyse en composantes indépendantes (*Independent Component Analysis ICA*) [53] a permis de réaliser une mesure rPPG de bien meilleure qualité que la solution initiale consistant à n'utiliser que le canal vert des caméras RGB. Deuxièmement, des travaux de recherche mettent en avant des méthodes exploitant des propriétés caractéristiques de la peau et du signal rPPG. Par exemple, la méthode *CHROM* [71] permet la projection du triplet $\{R, G, B\}$ sur un plan de l'espace RGB qui améliore grandement la qualité du signal. Il s'agit actuellement de la méthode la plus utilisée dans la littérature avec la méthode *ICA*. De plus, on peut souligner que les méthodes comme la différence des canaux $(G - R)$ [39] ou la sélection d'un canal unique [42] sont également des méthodes de projection sur un plan spécifique de l'espace RGB.

Même si un grand nombre de publications font référence à la robustesse des méthodes aux mouvements (par exemple [80, 106, 127]), il n'y a à ce jour auncune méthode permettant une mesure quelque soit le type de mouvement. De plus, l'observation de la peau présente plusieurs défis. Typiquement mesuré depuis le visage, le signal rPPG n'est pas reparti de façon homogène. Or quasiment toutes les méthodes de rPPG utilisent la valeur moyenne des pixels de peau à chaque instant. Ainsi, des pixels qui ne contiennent aucune information utile sont utilisés pour le calcul de cette moyenne ce qui réduit la qualité de l'information. En considérant l'intégralité du corps, la composition des tissus peut varier en fonction de la quantité d'eau, de gras ou de la vascularisation. Ainsi, les zones comme le visage, les mains, les lobes des oreilles ou la plante des pieds ont par exemple, un niveau de pulsation mesurable plus important que les bras, le coup ou la poitrine [102].

Dans ce chapitre, les méthodes de rPPG de la littérature sont d'abord présentées, en les positionnant par rapport à un modèle d'interaction lumière peau. Les méthodes par séparation de sources puis par projection sont ainsi détaillées. Une troisième section est dédiée à la définition des verrous scientifiques. Nous y analysons dans un premier temps les limitations des méthodes existantes du point de vue algorithmique puis, dans un second temps, le cout calculatoire des différentes méthodes est considérée dans l'objectif du développement d'un capteur rPPG. La définition de ces verrous permettra de déterminer l'orientation des travaux de thèse.

Du fait du très récent développement de ce champ de recherche, les contributions proposées dans la littérature portent principalement sur la recherche d'une combinaison linéaire optimale des canaux RGB. Il est possible de souligner ici que nos contributions décrites à partir du chapitre 4, concernent la segmentation de la région d'intérêt et aux performances et introduiront chacune les références nécessaires.

3.1/ Modèles d'interactions lumière-tissu pour la rPPG

Ce n'est que récemment qu'une modélisation plus avancée du phénomène photopléthysmographique sans contact a été considérée [96, 136]. Ainsi, les approches initiales se sont principalement focalisées sur l'étude de la qualité du signal en fonction de la longueur d'onde mesurée [52]. Ces études ont permis de favoriser le canal vert [42] pour une meilleure qualité de mesure. Ensuite, les combinaisons simples des canaux [39] ont améliorées les performances et permis une robustesse accrue aux perturbations, et enfin la reconstruction à l'aide de méthodes du traitement du signal existante comme l'analyse en composantes principales, PCA [49] et l'analyse en composantes indépendantes ICA [32, 47, 53] et ses variantes [64, 140, 142] bien plus présentes dans la littérature, ont permis des mesures de très bonnes qualités.

Dans leurs travaux originaux [47], Poh et al. considèrent uniquement que le signal rPPG et le bruit sont séparables. Le modèle dichromatique courant prend en compte la décomposition de l'information lumineuse en fonction du temps $S(t)$ comme la somme de l'information spéculaire, que l'on notera $v_s(t)$, et de l'information diffuse $v_d(t)$:

$$S(t) = v_s(t) + v_d(t). \tag{3.1}$$

Ces méthodes permettent une reconstruction du signal rPPG d'une qualité suffisante pour estimer le rythme cardiaque et sont souvent employées pour isoler le signal cardiaque des perturbations dues aux mouvements et aux changements des conditions d'illumination de la scène. En 2015, Feng et al. [96] proposent une modélisation simple pour expliquer les perturbations induites par le mouvement du sujet dans la mesure rPPG. Ils définissent $C_i(t)$ comme suit :

$$C_i(t) = \alpha_i\beta_i(S_0 + \gamma_i S_0 Pulse(t) + R_0), \tag{3.2}$$

avec $C_i(t)$ le canal de couleur rouge, vert ou bleu, $Pulse(t)$ le signal rPPG normalisé idéal. S_0 est l'intensité de la lumière diffuse dans la zone d'intérêt considérée avec une illumination blanche. R_0 est l'intensité de la lumière spéculaire réfléchie par la peau. L'indice i définit l'un des trois canaux du capteur RGB. Enfin α_i, β_i et γ_i sont des constantes adaptées aux différents canaux. Les mouvements sont une source de perturbations qui affectent tous les canaux de la même manière :

$$C_i(t) = \alpha_i\beta_i(S_0 + \gamma_i S_0 Pulse(t) + R_0)M(t), \tag{3.3}$$

avec $M(t)$ la modulation due aux mouvements. Aussi l'opération de différence des canaux peut être réécrite :

$$D(t) = \frac{C_i(t)}{\alpha_i\beta_i} - \frac{C_j(t)}{\alpha_j\beta_j} = (\gamma_i - \gamma_j)S_0 Pulse(t)M(t), \tag{3.4}$$

avec i et j deux indices différents pour les canaux R, G et B. L'opération de différence des canaux permet la suppression de la modulation par le mouvement spéculaire $R_0M(t)$. Feng et al. déduisent que le suivi temporel, tracking, de la zone d'intérêt est une opération primordiale puisqu'il permet la compensation de la modulation du signal rPPG par le mouvement. Cette première modélisation illustre clairement l'impact du mouvement sur le signal rPPG mesuré.

Toujours basé sur le modèle dichromatique, Wang et al. [136] proposent un modèle un peu différent afin d'établir de nouvelles méthodes de rPPG. Au niveau pixel, la peau fournit une information qui peut être modélisée sous la forme d'une fonction temporelle variant dans l'espace chromatique RGB, ainsi le signal mesuré peut être décomposé sous la forme :

$$C_k(t) = I(t) \cdot (v_s(t) + v_d(t)) + v_n(t). \tag{3.5}$$

Le signal $C_k(t)$ représente l'information RGB, en représentation vectorielle, pour chaque instant, du k-ieme pixel. Elle est fonction de $I(t)$, l'intensité lumineuse qui varie en fonction des caractéristiques de la source lumineuse mais également des distances respectives entre la source lumineuse et la surface de la peau ainsi que de la distance de la peau à la caméra. Elle est modulée par les fonctions $v_s(t)$ et $v_d(t)$ qui représentent respectivement l'information spéculaire et diffuse du modèle dichromatique. Les mouvements ainsi que la pulsation induite par les variations du volume sanguin sont à l'origine de ces deux fonctions. Enfin la fonction $v_n(t)$ représente le bruit de quantification induit par le capteur optique. Ce dernier terme est négligé en considérant le moyennage des pixels de peau.

Le modèle est ensuite affiné, considérant que l'information spéculaire ne contient aucune information de pulsation utile à la mesure PPG. La composante temporelle de cette fonction est due aux mouvements ainsi qu'aux changement de géométrie de la surface de peau, ce qui peut amener des variations d'intensité lumineuse considérant notamment la projection des ombres sur la surface observée :

$$v_s(t) = u_s \cdot (s_0 + s(t)), \tag{3.6}$$

avec u_s le vecteur couleur du spectre lumineux de la source, s_0 et $s(t)$ respectivement les valeurs moyennes et alternatives de l'information spéculaire. L'information diffuse est elle en revanche associée à la propagation de la lumière dans les tissus et à la rétrodiffusion associée :

$$v_d(t) = u_d \cdot d_0 + u_p \cdot p(t), \tag{3.7}$$

avec u_d et d_0 respectivement le vecteur couleur de la peau et l'intensité de la composante moyenne. u_p l'énergie du signal rPPG suivant les canaux, en se référant aux études préalables sur la qualité du signal en fonction de la longueur d'onde section 2.3.3. Contrairement au modèle présenté plus haut (Équation 3.2), cette modélisation ne considère pas de modulation par le mouvement pour la composante diffuse mais uniquement vis-à-vis de l'intensité lumineuse. Cette approximation laisse supposer que la combinaison linéaire des canaux doit permettre de déterminer le signal rPPG dans tous les cas. La substitution des termes définis plus haut et la combinaison des composantes continues $u_c \cdot c_0 = u_s \cdot s_0 + ud \cdot d_0$ permet la réécriture suivante :

$$C_k(t) = I_0 \cdot (1 + i(t)) \cdot (u_c + c_0 + u_s \cdot s(t) + u_p \cdot p(t)), \tag{3.8}$$

avec $I(t)$ défini comme la combinaison d'une composante continue I_0 et variable dans le temps $I_0 \cdot i(t)$, avec 1 le vecteur $(1, 1, 1)^T$. Avec cette réécriture il est désormais possible de définir le rôle

des algorithmes de rPPG de la façon suivante : la méthode doit permettre de rendre indépendante la mesure de pulsation $p(t)$ de l'intensité lumineuse et ses variations $I_0 \cdot (1 + i(t))$, qui peuvent être également modulé par les mouvements. Il faut ensuite isoler la composante spéculaire $s(t)$ afin de s'affranchir de toutes les sources de perturbations. Les méthodes rPPG reposant sur des modèles et caractéristiques liés à la peau seront présentés section 3.2.2.

Les méthodes de rPPG utilisent le triplet RGB à chaque instant issu du calcul de la moyenne des pixels de peau sur la zone d'intérêt considérée. Cette opération permet de réduire l'impact du bruit de quantification de la caméra $v_n(t)$ qui peut alors être négligé. Cette approximation permet également de considérer que les vecteurs couleurs définies ci-dessus sont indépendants de la position des pixels de peau. Ainsi, le signal $C(t)$ est essentiellement défini par la valeur moyenne des composantes R, G, et B à chaque instant :

$$C(t) \approx u_c \cdot I_0 \cdot c_0 + I_0 \cdot c_0 \cdot i(t) + u_s \cdot I_0 \cdot s(t) + u_p \cdot I_0 \cdot p(t). \tag{3.9}$$

Cette approximation est possible en considérant le fait que les modulations AC des signaux sont beaucoup plus faibles que leurs composantes continues. Ce qui fait que le produit des modulations (a titre d'exemple $p(t) \cdot i(t)$) peut être négligé puisque d'amplitude très faible. Cette dernière approximation permet de définir le signal temporel $C(t)$ comme un mélange linéaire des signaux $i(t)$, $s(t)$ et $p(t)$. Les bons résultats de la méthode ICA pour extraire le signal rPPG sont totalement cohérents avec cette modélisation.

La même méthode de différence des canaux, vue en section précédente, permet alors une simplification bien plus avantageuse.

$$D(t) = C_i(t) - C_j(t) = I_0 \cdot c_0 \cdot (i_i(t) - i_j(t)) + u_s \cdot I_0 \cdot (s_i(t) - s_j(t)) + u_p \cdot I_0 \cdot (p_i(t) - p_j(t)) \tag{3.10}$$

Avec i et j deux canaux différents parmi les trois R, G et B. Comme les variations temporelles induites par les perturbations lumineuses ou les mouvements affectent sensiblement tous les canaux de la même façon, la différence des composantes $i_i(t) - i_j(t)$ et $s_i(t) - s_j(t)$ devient négligeable devant les autres termes. De plus, en considérant la contribution rPPG suivant les longueurs d'onde, en sélectionnant correctement les canaux, $p_i(t) - p_j(t) \approx p_i(t)$ dans le cas G - R. Le signal final ainsi obtenu a une contribution très amoindrie des signaux $i(t)$ et $s(t)$:

$$D(t) \approx I_0 \cdot c_0 \cdot \epsilon_1(t) + u_s \cdot I_0 \cdot \epsilon_2(t) + u_p \cdot I_0 \cdot p(t), \tag{3.11}$$

avec les signaux $\epsilon_k(t)$, pour k valant 1 ou 2 respectivement pour l'information résiduelle des signaux $i(t)$ et $s(t)$, faible devant $p(t)$.

3.2/ Algorithmes rPPG existants

Historiquement, Verkruysse et al. [42] sont les premiers à montrer la faisabilité de la méthode rPPG en mesurant un signal au niveau du front d'une personne immobile. Leurs travaux montrent la modulation dans les trois canaux R, G et B et déterminent que le canal vert contient le plus d'information. Ils n'introduisent pas de méthode spécifique si ce n'est un filtrage passe-bande du signal afin d'éliminer toutes les modulations qui ne correspondent pas aux plages de valeurs possibles spécifiques au rythme cardiaque.

Hulsbusch et al. [39] introduisent plus tard la méthode de différence entre les canaux vert et rouge. Comme exposé dans la section (3.2.1), cette méthode permet d'atténuer les perturbations dues au mouvement et variations de l'illumination de la scène afin d'obtenir une meilleure estimation du signal rPPG.

L'utilisation du canal vert ou la différence des canaux vert et rouge présentées ci-dessus permettent la mesure mais ne sont pas optimales. En effet, en considérant un seul canal parmi les trois disponibles, il est impossible de supprimer l'influence des perturbations puisque les trois canaux sont porteurs de modulations induites par les mouvements et les variations des conditions d'illuminations. En revanche, la différence directe de deux canaux permet de supprimer les perturbations dans la mesure mais cette méthode n'est pas optimale car elle impacte la qualité de la mesure rPPG en atténuant le signal. A partir de ces premières tentatives, les méthodes de traitement du signal (ICA, PCA) par séparation de sources aveugles ont tout d'abord été employées pour isoler le signal rPPG. Récemment, des méthodes de projection directe du vecteur temporel $C(t)$ sur un plan adapté de l'espace chromatique ont été mise au point et sont les plus performantes.

3.2.1/ Algorithmes par séparation de sources

Les méthodes par séparation de sources estiment le signal rPPG en séparant les sources mélangées dans les canaux RGB en maximisant la variance des composantes (*PCA*) ou leurs indépendances (*ICA*). En 2010, Poh et al. [47] utilisent la méthode de traitement du signal *ICA* afin d'exploiter tous les canaux et de reconstruire l'information rPPG la plus complète possible. Dans leurs approches les auteurs considèrent que les contributions au signal rPPG sont réparties dans les canaux RGB et constituent autant d'éléments de mesure qui sont mélangés par une opération linéaire pour chaque observation (canal de la caméra) :

$$x_i(t) = \sum_{j=1}^{3} a_{ij} \cdot C_j(t), \tag{3.12}$$

avec $C_j(t)$ les signaux des canaux R, G, et B et $x_i(t)$ les signaux sources qui sont mélangés par les facteurs a_{ij}. Cette formulation est très proche de celle introduite par Wang et al. [136] (Équation 3.10). Par contre, dans ce cas, les signaux sources ne sont pas clairement identifiés. Dans le cas de l'utilisation de la méthode *ICA*, on considère en général que les signaux résultants, autre que le signal rPPG, ne sont pas utiles. De plus les coefficients sont indépendants les uns des autres, la Figure 3.1 permet d'illustrer l'utilisation de la méthode *ICA* appliquée à la mesure rPPG. Dès lors, toutes les méthodes peuvent être formulées suivant cette hypothèse de combinaison linéaire de sources indépendantes.

Le problème est de la forme :

$$X(t) = W \cdot C(t), \tag{3.13}$$

avec $X(t)$ les signaux sources, qui correspondent au signal rPPG avec les différentes perturbations, W la matrice pour isoler les signaux mélangés dans les observations $C(t)$. La différence majeure entre PCA et ICA est l'hypothèse concernant les signaux sources. ICA considèrent que les sources sont indépendantes alors que PCA considèrent qu'elles sont non corrélées, ce qui a pour effet de différencier les deux approches dans leur façon d'établir W. PCA utilise la matrice de covariance construite d'après les signaux RGB, $Cov = C \cdot C^T$, afin d'en déterminer la décomposition en vecteurs propres :

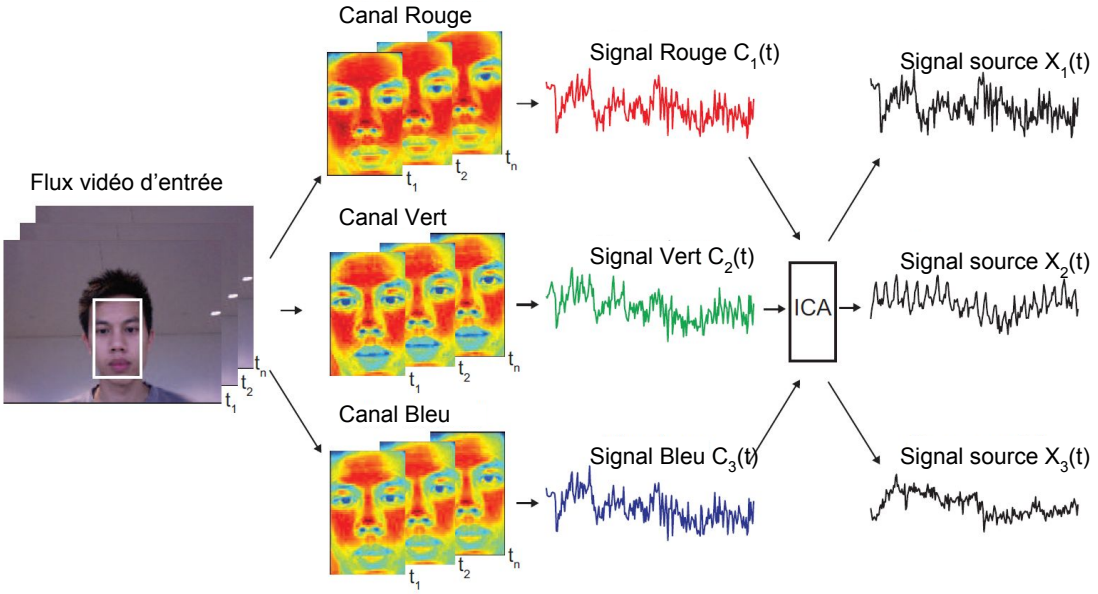


FIGURE 3.1 – Le flux vidéo d'entrée est segmenté afin d'isoler le visage du sujet, les pixels sont moyennés dans les trois canaux afin de construire les signaux $x_i(t)$ (*e.g.* les observations). La méthode ICA permet d'extraire autant de signaux de sortie qu'il y a d'observations, ici trois. Le signal rPPG est alors contenu dans l'un des trois signaux [47].

$$Cov = C \cdot C^T = U \cdot \Sigma \cdot V, \tag{3.14}$$

où U et V sont les matrices de vecteurs propres pour l'espace propre où la matrice Cov est diagonale de spectres $diag(\Sigma)$. Cette opération permet de maximiser la variance du signal avec la valeur propre la plus importante. Cependant, pour permettre d'isoler le signal rPPG, celui-ci doit être prépondérant vis-à-vis du bruit ou d'un autre signal présent, dû à des mouvements par exemple. Il n'y a de fait aucune garantie que le signal résultant de l'utilisation de PCA soit le signal rPPG. Dans le cas de ICA, les hypothèses sont que les sources sont statistiquement indépendantes pour déterminer W.

Comme la qualité de la construction de la matrice W est dépendante du nombre d'échantillons de mesure, il est préférable de travailler avec un signal d'entrée de longue durée, 30 secondes pour Poh et al. [47]. De plus, parmi les signaux sources reconstruits, il n'y a aucune indication sur l'index du signal qui représente la mesure rPPG, ce qui pose un problème évident dans le cas d'une mesure automatisée. Ce problème a été solutionné de deux façons. En considérant que le signal rPPG est le seul qui montre une périodicitée, il est possible d'identifier le signal résultant en observant leurs covariances ou décomposition en fréquences. Cependant, il est également possible d'intégrer cette notion directement dans le processus de reconstruction des signaux. La seconde solution consiste à utiliser une variante de la méthode *ICA* : *Constrained ICA (cICA)*. Cette méthode permet de reconstruire les signaux sources en considérant des contraintes. Dans le cas de la rPPG, la reconstruction priorise les signaux pseudo-périodiques, à l'aide d'une mesure de covariance par exemple [140] ou d'inter-corrélation avec des signaux périodiques (sinusoides [64]).

De plus, la complexité algorithmique inhérente à la construction des matrices de covariance puis de la décomposition en vecteurs propres ou valeurs singulières (SVD) ou bien la reconstruction des signaux avec *ICA* est bien moins optimal qu'une projection directe des canaux RGB sur un plan donné et est ainsi moins adapté à une solution temps réel. Enfin, les résultats obtenus dans

la littérature montrent une performance moindre en utilisant *ICA* ou *PCA* en comparaison avec d'autres méthodes dédiées à la rPPG : *e.g. CHROM*, *POS* ou *2SR*.

3.2.2/ Algorithmes par projection

Les méthodes par projection présentent l'avantage de permettre une formulation analytique directe basée sur un modèle. Les méthodes qui utilisent une association ou une sélection de canaux sont les suivantes :

- Green [42], 2008 : Le signal vert est sélectionné parmi les trois canaux RGB.
- Green - Red [39], 2008 : La combinaison des canaux permet de réduire l'impact des perturbations induites par les mouvements.
- CHROM [71], 2013 : La combinaison linéaire des canaux est ici faite pour projeter les canaux R, G et B de la caméra suivant deux vecteurs formant un plan perpendiculaire à un vecteur universel de couleur de la peau.
- PBV [80], 2014 : Le vecteur Pulse Blood Volume est un vecteur représentant la direction dans l'espace chromatique RGB des variations du signal rPPG. La projection des canaux suivant ce vecteur permet alors de s'affranchir des autres sources d'information et de bruit.
- 2SR [126], 2016 : Cette méthode utilise la projection successive du vecteur peau principal, basée sur la décomposition en vecteurs propres, sur les deux autres vecteurs de la base propre issus d'une pseudo matrice de covariance avec une image de référence. Il est ainsi possible d'observer les variations de direction du vecteur représentant la peau dans le temps. Cette méthode est intéressante puisqu'elle est uniquement basée sur les informations fournies par les données, ce qui en fait une méthode personnalisée pour chaque sujet.
- POS [136], 2017 : Cette méthode est similaire à CHROM puisque les canaux R, G et B sont projetés suivant deux vecteurs afin de construire un plan qui cette fois permet de se rendre indépendant de la composante de spéculaire.

La méthode CHROM est certainement la méthode la plus employée dans la littérature. Cela s'explique par sa faible complexité en calcul et son efficacité pour isoler le signal rPPG même en cas de perturbations mineures. Les différences de performance entre les quatre dernières méthodes ne sont pas significatives. Bien que plus récente, la méthode POS permet une mesure d'une qualité équivalente à celle de CHROM dans nos expérimentations (Chapitre 5) et ne se démarque pas nécessairement dans la littérature [130, 136].

Wang et al. [136] lient leurs travaux au modèle (présenté dans l'équation 3.9). Les méthodes PBV et CHROM sont similaires dans leurs approches. Elles reposent sur des connaissances propres à la mesure rPPG établies sur un ensemble de données varié afin d'obtenir une modélisation compatible avec un maximum de sujets. Cela afin de rendre le signal $C(t)$ indépendant de la contribution moyenne de la peau en réflectance défini par la composante continu du signal sur chacun des canaux. Pour ce faire, les signaux RGB sont normalisés en les divisant par leurs valeurs moyennes. On obtient alors le signal $\bar{C}(t)$ défini par :

$$\begin{aligned} \bar{C}(t) &\approx u_c \cdot I_0 \cdot c_0, \\ N \cdot \bar{C}(t) &= N \cdot u_c \cdot I_0 \cdot c_0 = 1. \end{aligned} \tag{3.15}$$

A partir de l'équation 3.9, on détermine une matrice de normalisation N. Après réduction, on obtient la fonction normalisée suivante :

$$\tilde{C}_n(t) \approx 1 \cdot i(t) + N \cdot u_s \cdot I_0 \cdot s(t) + N \cdot u_p \cdot I_0 \cdot p(t), \tag{3.16}$$

avec $1 = [1, 1, 1]^T$ le vecteur direction de l'intensité lumineuse. La méthode PBV consiste à projeter ce signal normalisé suivant une unique direction définie comme proportionnelle au signal utile $p(t)$:

$$\hat{p}(t) = \tilde{C}_n^T(t) \cdot z = k \cdot p(t), \tag{3.17}$$

avec z le vecteur de projection 3 x 1 et k le facteur de proportionnalité avec $k \in \Re^*$. De plus, on considère que le signal $p(t)$ et donc $\hat{p}(t)$ ne sont pas corrélés avec les autres signaux sources $s(t)$ et $i(t)$:

$$\mathbb{E}\{p(t) \cdot i(t)\} = \mathbb{E}\{p(t) \cdot s(t)\} \approx 0, \tag{3.18}$$

avec $\mathbb{E}$ l'espérance mathématique. On peut alors réécrire :

$$\mathbb{E}\{\tilde{C}(t) \cdot \hat{p}(t)\} = k \cdot \mathbb{E}\{\tilde{C}(t) \cdot p(t)\} \approx k \cdot N \cdot u_p \cdot I_0 \cdot \mathbb{E}\{p(t) \cdot p(t)\}. \tag{3.19}$$

Dans leurs travaux, Haan et al. [80] déterminent le vecteur pour la méthode PBV sur une base de données privée. Ce vecteur doit être établi en fonction de la scène pour maximiser la qualité du signal. Dans leur approche théorique, ils définissent ce vecteur afin de permettre l'approximation :

$$u_{pbv} \approx N \cdot u_p \cdot I_0. \tag{3.20}$$

Cette hypothèse est nécessaire à l'application de la méthode. L'information d'entrée de la méthode est la matrice de covariance temporelle 3x3 des signaux $\tilde{C}_n(t)$, on réécrit alors l'Équation 3.19 :

$$\mathbb{E}\{\tilde{C}(t) \cdot \hat{p}(t)\} \cdot z = U_{pbv} \cdot k \cdot \mathbb{E}\{p^2(t)\}, \tag{3.21}$$

et le vecteur projection z peut être défini par :

$$z = \mathbb{E}\{\tilde{C}(t) \cdot \hat{p}(t)\}^{-1} \cdot U_{pbv} \cdot k \cdot \mathbb{E}\{p^2(t)\}, \tag{3.22}$$

au lieu de la moyenne globale sur l'ensemble de pixels de peau, la méthode PBV utilise une matrice de covariance 3x3 défini par :

$$\tilde{\Sigma} = \overline{\{\tilde{C}_n(t) \cdot \tilde{C}_n^T(t)\}}, \tag{3.23}$$

avec $\{\bar{\cdot}\}$ l'opération de moyenne temporelle pour obtenir les valeurs de covariance pour chaque instant. Comme k est choisi tel que $k = 1/\|u_{pbv}\|$, on obtient finalement que le vecteur de projection z soit défini proportionnel à l'inverse de la matrice de covariance suivant le vecteur u_{pbv} :

$$z \propto \tilde{\Sigma}^{-1} \cdot u_{pbv}. \tag{3.24}$$

La méthode PBV a cependant plusieurs limitations. La qualité de la mesure peut diminuer brutalement si l'hypothèse 3.20 n'est pas respectée, ce qui survient si le vecteur n'est pas établi pour

la scène et que les différences entre la scène de référence (illuminations, couleur de peau, ...) sont trop importantes. De plus comme la détermination du vecteur z est définie par un calcul d'inverse, le rang de la matrice $\tilde{\Sigma}$ doit être égale à 3 pour qu'elle soit inversible. Il est théoriquement possible de parvenir aux conditions idéales où $i(t) = s(t) = 0$. Dans ce cas, n'importe quelle valeur peut être employée pour z puisque la seule contribution est celle du signal rPPG. Cependant, elle n'est valable que pour une expérimentation particulière puisque le vecteur u_{pbv} est déterminé par les conditions d'illuminations de la scène et le sujet de l'expérimentation. Il s'agit de la limitation principale de cette méthode. Dans leurs travaux, Hann et al. [80] définissent la valeur pour le vecteur u_{pbv} comme une valeur moyenne sur un ensemble de test pour permettre la mesure sur l'ensemble de la base de données de référence avec $u_{pbv} = [0.33, 0.77, 0.53]^T$.

La méthode CHROM [71] permet la mesure avec de plus faibles contraintes. Dans un premier temps, le signal $\tilde{C}_n(t)$ est projeté suivant une matrice M défini comme : $M^{-1} \cdot u_{skin} = 1$, où le vecteur $u_{skin} = [0.77, 0.51, 0.38]^T$ est un vecteur normalisé de la couleur de peau défini comme une moyenne sur un ensemble de test à grande échelle. Comme précédemment, avec la matrice de normalisation N, on défini la matrice M pour éliminer les composantes sources de perturbations dans $\tilde{C}_n(t)$ tel que :

$$M \cdot \tilde{C}_n(t) = M \cdot 1 \cdot i(t) + M \cdot N \cdot u_s \cdot I_0 \cdot s(t) + M \cdot N \cdot u_p \cdot I_0 \cdot p(t). \tag{3.25}$$

Le vecteur de la réflectance spéculaire $N \cdot u_s \cdot I_0$ est considéré proportionnel à la lumière blanche, en considérant que les effets de brillance dut à l'information spéculaire sont dans la direction du vecteur 1, on a l'approximation $M \cdot N \cdot u_s \cdot I_0 \approx \nu \cdot 1$ avec $\nu \in \Re^*$. Enfin, $M \cdot \tilde{C}_n(t)$ est projeté suivant un plan orthogonal à 1 pour être indépendant des variations spéculaires :

$$S(t) = P_c \cdot M \cdot \tilde{C}_n(t) \approx P_c \cdot M \cdot 1 \cdot i(t) + P_c \cdot M \cdot N \cdot u_p \cdot I_0 \cdot p(t), \tag{3.26}$$

avec $P_c = \begin{pmatrix} 1 & -1 & 0 \\ 0.5 & 0.5 & -1 \end{pmatrix}$ une matrice 2x3 défini manuellement dans [71] comme une matrice de projection orthogonale dans l'espace RGB. La suppresion du terme spéculaire, avec $P_c \cdot M \cdot N \cdot u_s \cdot I_0 \approx P_c \cdot nu \cdot 1 = 0$, est alors possible. La projection finalement employée pour réaliser la mesure rPPG est définie par la matrice $P_c \cdot M = \begin{pmatrix} 3 & -2 & 0 \\ 1.5 & 1 & -1.5 \end{pmatrix}$, on obtient alors :

$$\tilde{p}(t) = S_1(t) - \alpha \cdot S_2(t), \alpha = \frac{\sigma(S_1)}{\sigma(S_2)}, \tag{3.27}$$

avec S_i les deux signaux issus de la projection $P_c \cdot \tilde{C}_n(t)$ et $\sigma(S_i)$ leurs déviations standard. L'utilisation de cette matrice et la reconstruction du signal rPPG $\tilde{p}(t)$ de cette façon met en évidence un phénomène de phase / antiphase pour les variations induites par les mouvements et celles induites par le signal rPPG. L'auteur explique dans les travaux lié à POS [136] que cela est dû au choix de la matrice P_c qui sera détaillé plus loin. Cette méthode, bien que particulièrement efficace n'est pas exempte de limitations. L'utilisation du vecteur u_{skin} pour définir la matrice de projection P_c suppose que l'hypothèse d'un vecteur universelle de peau soit vérifiée. Le vecteur u_{skin} est construit comme un vecteur moyen afin de définir la direction de l'information spéculaire, ce faisant plus les caractéristiques du sujet s'éloignent de la moyenne plus l'information spéculaire résiduelle sera importante.

La méthode POS [136] est très similaire à CHROM dans son utilisation, à nouveau une matrice de projection est utilisée afin de construire un signal depuis une combinaison linéaire des canaux RGB. L'étude menée pour déterminer ce nouveau plan de projection permet cependant la mise

en place d'une méthode pour l'évaluation de la quantité d'informations, au travers des composantes pulsatives et spéculaires. Dans un premier temps, le signal $C_n(t)$ est projeté suivant un plan perpendiculaire au vecteur $1 = [1, 1, 1]$ tel que :

$$\begin{aligned} S(t) &= P_p \cdot C_n(t), \\ &\approx P_p \cdot N \cdot u_s \cdot I_0 \cdot s(t) + P_p \cdot N \cdot u_p \cdot I_0 \cdot p(t), \end{aligned} \tag{3.28}$$

avec P_p la matrice de projection 2x3 vers le plan perpendiculaire à 1. Les vecteurs, en ligne, $P_{p,1}$ et $P_{p,2}$ sont définis comme orthogonaux. Le plan de projection est ansi perpendiculaire à la direction des variations de l'intensité lumineuse. A titre de comparaison, dans CHROM le plan de projection est construit pour être perpendiculaire à la direction des variations de l'information spéculaire, avec l'hypothèse d'un vecteur peau universelle. Ce changement de construction du plan de projection permet aux auteurs, considérant l'hypothèse initiale du modèle que les variations induites par le mouvement sont traduites par des modulations du signal $i(t)$ égales dans les trois canaux, de s'affranchir de la plus importante quantité de perturbations dans la mesure.

En se basant sur l'étude des variations du vecteur z, Wang et al. déterminent finalement une matrice de projection permettant la séparation des informations spéculaires et pulsatives $P_p = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix}$. La construction des signaux $S_1(t) = G_n(t) - B_n(t)$ et $S_2(t) = G_n(t) + B_n(t) - 2 \cdot R_n(t)$ permet la projection des trois signaux normalisés $R_n(t)$, $G_n(t)$ et $B_n(t)$ sur un plan maximisant l'information pulsative. les auteurs définissent $\tilde{p}(t)$ comme la fusion des signaux $S_1(t)$ et $S_2(t)$:

$$\tilde{p}(t) = S_1(t) + \alpha \cdot S_2(t), \alpha = \frac{\sigma(S_1)}{\sigma(S_2)}. \tag{3.29}$$

La différence de signe entre la méthode CHROM et la méthode POS vient du fait que les vecteurs de la matrice de projection P_c ont des signes différents pour l'information avec le vecteur z, c'est ce qui crée l'inversion de phase entre les deux signaux $S_1(t)$ et $S_2(t)$. Cela n'est pas le cas avec la matrice de projection P_p. La méthode POS permet de s'affranchir des modèles utilisés par les méthodes CHROM et PBV. La mesure peut ainsi être potentiellement améliorée par rapport aux méthodes précédemment décrites. Cependant, les données de test employées ne sont pas ouvertes à la communauté scientifique, et les implémentations faites dans la littérature ou dans le cadre de cette thèse montre que les différences de performances entre les méthodes CHROM et POS ne sont pas significatives [130].

En marge des modèles utilisés jusqu'ici, une autre méthode complètement pilotée par les données (*data-driven*) a été mise au point par Wang et al. en 2016 [126]. La méthode 2SR permet la mesure du signal rPPG en considérant les variations de la composante principale du vecteur de peau établi à partir des données d'entrée. Au lieu de considérer la moyenne des pixels de peau dans une zone d'intérêt, la méthode 2SR travaille du point de vue de la distribution de ces pixels avec :

$$A = \frac{V^T \cdot V}{N}. \tag{3.30}$$

La matrice 3x3, notée A, est la matrice de covariance du vecteur d'entrée V où les lignes de V sont des pixels uniques et les colonnes sont les canaux RGB. N est le nombre de pixels de peau utilisés pour construire le vecteur V. La décomposition en vecteurs propres de la matrice A permet d'isoler la composante principale définie comme étant la couleur de la peau :

$$A = U \cdot \sigma \cdot V = \lambda_1 \cdot u_1 \cdot u_1^T + \lambda_2 \cdot u_2 \cdot u_2 T + \lambda_3 \cdot u_3 \cdot u_3^T. \tag{3.31}$$

Comme la matrice A est symétrique par construction, le vecteur V est la transposée et l'inverse du vecteur U, avec λ_i la i-ieme valeur propre et u_i le i-ieme vecteur propre. Comme les vecteurs u_i sont unitaires, les variations temporelles de la distribution et donc de la matrice A se traduisent nécessairement par une rotation du sous-espace vectoriel. En considérant deux sous espaces à deux instants distincts, pour la première image U_τ et à l'instant courant U_t, on établit la rotation entre ces deux sous espaces comme :

$$R = U_t^T \cdot U_\tau = \begin{pmatrix} u_1^t & u_2^t & u_3^t \end{pmatrix} \cdot \begin{pmatrix} u_1^\tau & u_2^\tau & u_3^\tau \end{pmatrix}. \tag{3.32}$$

Le produit scalaire est égal au produit des normes multiplié par le cosinus de l'angle entre les deux vecteurs. Comme dans ce cas les vecteurs propres ont une norme unitaire et on peut définir :

$$\cos(\Theta_{ij}) = \frac{{u_i^t}^T \cdot u_j^\tau}{\|u_i^t\| \cdot \|u_i^\tau\|}, \tag{3.33}$$

avec i et j les indices des colonnes et Θ_{ij} l'angle de rotation entre les deux vecteurs. Les propriétés des vecteurs propres, orthogonalités et normes unitaires, permettent de simplifier l'écriture de R. De plus, les variations de distribution du vecteur V sont minimes donc la mesure doit être faite considérant l'approximation des petits angles $\sin(\theta) \approx \theta$. Aussi, une fois tous les angles testés pour effectuer la mesure, les auteurs sélectionnent les vecteurs qui montrent la pulsation de plus grandes amplitudes afin de mesurer les variations temporelles de la distribution des pixels de peau :

$$R' = {u_1^t}^T \cdot \begin{pmatrix} u_2^\tau & u_3^\tau \end{pmatrix}. \tag{3.34}$$

En plus de considérer les rotations des sous-espaces, les auteurs utilisent l'information de l'énergie des valeurs propres, qui sont la variance des vecteurs propres, afin d'exploiter les modulations de la variance de la distribution des pixels de peau également impactés par le signal rPPG. Wang et al. établissent le rapport des énergies pour les sous-espaces comme :

$$S = \begin{pmatrix} \sqrt{\lambda_1^t / \lambda_2^\tau} \\ \sqrt{\lambda_1^t / \lambda_3^\tau} \end{pmatrix}. \tag{3.35}$$

Considérant que la mesure est faite sur une fenêtre temporelle, la référence du sous-espace pour les vecteurs propres U_τ changent pour chaque nouvelle fenêtre. Aussi il est nécessaire de projeter les vecteurs à nouveau dans l'espace RGB classique :

$$S \cdot R' = \sqrt{\frac{\lambda_1^t}{\lambda_2^\tau}} \cdot {u_1^t}^T \cdot u_2^\tau \cdot {u_2^\tau}^T + \sqrt{\frac{\lambda_1^t}{\lambda_3^\tau}} \cdot {u_1^t}^T \cdot u_3^\tau \cdot {u_3^\tau}^T. \tag{3.36}$$

Un signal SRt', dont les deux premières composantes sont en opposition de phase, est construit en concaténant les mesures répétés des variations de la distribution en variance et rotations $S \cdot R'$. De même que pour CHROM la construction du signal final h est établi en les fusionnant :

$$h = SRt'_1 - \frac{\sigma(SRt'_1)}{\sigma(SRt'_2)} \cdot SRt'_2. \tag{3.37}$$

Dans leurs travaux, Wang et al. [126] construisent un signal final H défini comme la somme de tous les signaux h successifs, moins leur moyenne, pour les différentes fenêtres temporelles avec un pas d'une seconde entre chaque nouveau sous espace de référence. Cette méthode permet une personnalisation de la mesure pour chaque sujet mais nécessite un très bon suivi temporel de la zone d'intérêt pour ne pas altérer la distribution en incorporant des pixels qui ne représente pas la peau, ou en éliminant une zone de la région d'intérêt entre deux images suite à un mouvement par exemple. Cette méthode souffre cependant d'une complexité algorithmique plus importante comme la construction de la matrice de covariance ainsi que la décomposition en vecteurs propres pour chaque nouvelle image entraine un cout en calcul plus important que les méthodes *POS*, *CHROM* ou *PBV*.

3.3/ Problématiques et objectifs de la thèse

Les méthodes de rPPG de la littérature ont un certain nombre de limitations qu'il est nécessaire d'identifier. En considérant le schéma fonctionnel suivant (cf. Figure 3.2) comme une approche standard de la méthode rPPG, il est possible d'isoler les différences fondamentales des méthodes existantes.

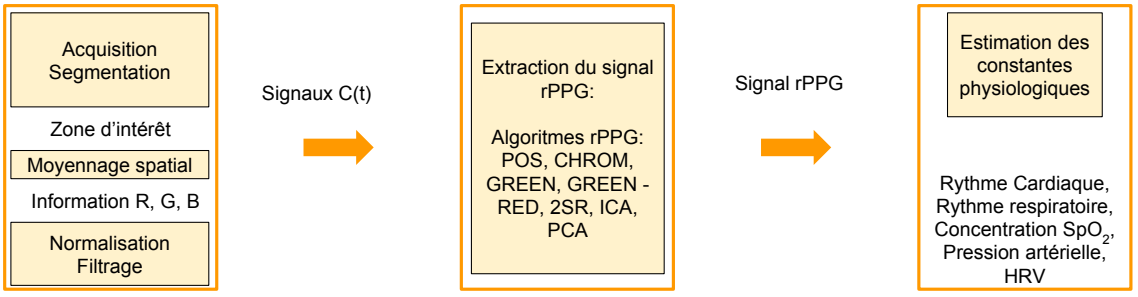


Figure 3.2 – Schéma fonctionnel pour la mesure rPPG tel que réalisée dans la littérature. En encadré, les fonctions de la chaîne du processus de mesure rPPG, et en flottant les informations issues des différentes fonctions.

La première étape de la chaîne algorithmique pour la mesure rPPG est l'acquisition du flux vidéo et la segmentation. Celle-ci est généralement effectuée avec une détection du visage dans la première image puis le suivi de cette détection est effectué pour les images suivantes. Des méthodes de segmentation de la peau sont également utilisées afin d'éliminer les zones qui ne contiennent pas d'informations utiles. Typiquement, les cheveux, les yeux, ou tout autre zone dont l'information spéculaire est trop importante (brillance qui peut saturer le capteur) et qui sont donc éliminés par la segmentation de peau. L'information fournie par les canaux R, G et B est issue (mis à part dans le cas de la méthode 2SR) du calcul de la moyenne de tous les pixels considérés par l'étape de segmentation. L'acquisition, répétée pour chaque nouvelle image, permet la construction du signal $C(t)$, tel que décrit plus avant (cf. section 3.2). La normalisation et le filtrage de ces signaux permettent de s'affranchir des perturbations qui ne sont pas en lien avec les modulations que l'on souhaite extraire.

Les différents algorithmes existants sont alors utilisés afin de permettre l'extraction du signal rPPG

et d'en déduire les informations utiles telles que les rythmes cardiaques ou respiratoires ou encore la saturation SpO_2. Les principales différences entre les travaux présents dans la littérature sont centrées autour des méthodes employées afin de construire le signal rPPG à partir du signal d'entrée $C(t)$. Cependant, chacune de ces méthodes repose sur plusieurs hypothèses qui peuvent être complexes. Dans le cas idéal, la mesure s'effectue sous des conditions d'éclairages constants, sans mouvement du sujet et avec une période d'intégration de 20 secondes. Qui plus est, toutes les méthodes de construction du signal rPPG supposent que le signal $C(t)$ est continu sur la période d'intégration et désigne toujours la même zone de peau. La première étape de segmentation est ainsi primordiale afin de permettre la bonne mesure du signal rPPG. Même s'il est courant de voir des publications dans la littérature qui présentent des méthodes robustes aux mouvements [32, 57, 80, 96, 106, 127, 128], les scénarios sont souvent limités et la mise en application de ces méthodes pour des mesures en direct est peu efficace en raison de leurs complexités et spécificités. Les différentes méthodes de segmentation de la zone d'intérêt sont couteuses en terme de calculs. À titre d'exemple, pour la détection du visage on limite la recherche à une taille de visage minimale pour la première image, puis on cherche dans un voisinage de la première détection pour les nouvelles images afin de réduire les couts de calcul. Ce besoin de performance, qui est également fonction de la complexité de la méthode de rPPG employée, justifie l'intérêt de développer les méthodes par projection qui sont bien moins couteuses en puissance de calcul.

La mesure rPPG est directement impactée par l'environnement de mesure. Les améliorations et traitements automatiques ainsi que la compression du flux vidéo de la caméra doivent être désactivés et les conditions d'éclairages doivent varier le moins possible. Si ces contraintes sont importantes et limitent le nombre de scénarios possibles pour la mesure rPPG sans même considérer le sujet de la mesure, la plus lourde contrainte reste l'étape de segmentation. En effet, la segmentation de la zone d'intérêt et son suivi temporel impactent toute la chaîne algorithmique en introduisant des perturbations dont il est difficile de s'affranchir. Les méthodes de segmentation courantes s'appuient sur la détection du visage comme il est admis que cette zone permet d'obtenir un signal rPPG de bonne qualité. Cependant, si cela permet de s'affranchir d'une contrainte importante, la mesure est de fait supervisée. La conséquence de ce choix algorithmique est que le visage doit être accessible, ce qui limite les scénarios et rend le suivi complexe puisque les orientations du visage modifient les zones de peau visibles par la caméra. La segmentation de peau est également un élément problématique puisque souvent inadaptée à une grande variété de couleurs de peau. Un classifieur entraîné pour des teints de peaux claires échoue souvent avec des teints de peau plus foncés et inversement [74], ce qui rend cette méthode peu efficace afin de proposer une méthode générale de mesure rPPG.

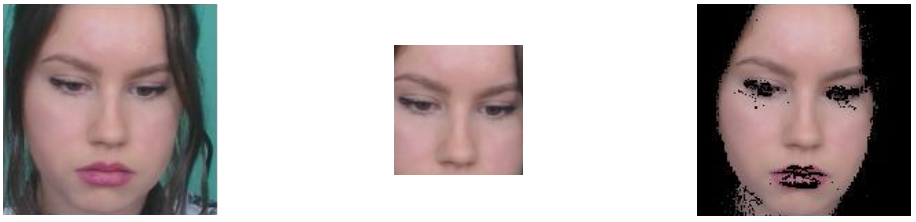


FIGURE 3.3 – Segmentation de la zone d'intérêt pour la mesure rPPG. Détection et segmentation du visage puis segmentation de la peau.

Les contraintes appliquées au sujet de la mesure sont inhérentes aux scénarios de la mesure. Elles sont lourdes dans le cas d'une surveillance cardiaque pour des exercices sportifs et négligeables dans le cas d'une surveillance cardiaque à un bureau. Si le premier cas nécessite la mise en place d'une chaîne algorithmique complexe afin de compenser les perturbations induites par les mouvements, la surveillance dans des cas statiques est nettement simplifiée et les résultats des différentes

méthodes de l'état de l'art valide l'efficacité de la méthode rPPG.

Afin de parvenir à rendre temps réel la mesure rPPG, la question des performances des différents segments de la chaîne algorithmique doit être soulevée. En progressant dans la chaîne fonctionnelle, on constate qu'une fois encore la première étape d'acquisition / segmentation est le point limitant de la méthode. Les copies en mémoire nécessaires lors de l'acquisition du flux vidéo ainsi que la recherche des visages par convolutions [20] sont des opérations qui peuvent consommer beaucoup de temps de calcul. Même si les systèmes modernes fonctionnent souvent avec plusieurs cœurs de calculs (on pense aux smartphones ou aux cartes PC type Raspberry par exemple), il est nécessaire de valider le fonctionnement des méthodes avec le moins de recours à du parallélisme possible, ne serait-ce que parce que les systèmes sont rarement entièrement dédiés à une seule application et que le degré de parallélisme à l'exécution est ainsi limité par l'ordonnancement du système. Les méthodes de rPPG par projection sont très efficaces et ont des temps de calcul très rapide, le recours à des méthodes comme PCA, ICA ou encore 2SR et la décomposition en valeur propre/singulaire associée augmente significativement le nombre d'opérations nécessaires et par conséquent le temps de calcul. Enfin les opérations de filtrage ou de décomposition en transformée de Fourier sont des opérations à faible complexité en tenant compte que l'on ne traite qu'un signal 1-D d'une longueur de 600 points dans le cas d'une acquisition avec une fréquence images de 30 fps et pour 20 secondes de signal.

La mesure rPPG *non supervisée*, telle que définie dans [106], permet d'effectuer la mesure sans avoir à identifier le visage ou même à segmenter la peau dans la scène. Cette méthode permet de s'affranchir d'une contrainte importante présente dans la majorité de la littérature qui est la nécessité de travailler avec le visage. Afin de permettre ce type d'approche, il est nécessaire de définir une référence à identifier dans le flux vidéo. Pour ce faire, Wang et al. [106] utilisent une décomposition du flux vidéo en voxels, qui sont des clusters de pixels cohérents dans les 6 dimensions (espace, couleur et temps) [76] et avec une décomposition pyramidale afin de détecter des zones de peau de différentes tailles. L'identification des zones intéressantes s'effectue sans connaissances a priori sur la couleur. C'est l'information temporelle des différents voxels et leurs caractéristiques similaires qui permet de regrouper les voxels similaires. La seule hypothèse est que les éléments de l'image qui vont fournir une information intéressante ont un signal périodique. La réponse fréquentielle d'un tel signal a une fréquence fondamentale de grande énergie, ce qui se traduit par un rapport signal sur bruit positif. En associant entre eux les voxels similaires, à l'aide des signaux temporels qui en résultent, et avec les propriétés inhérentes à la décomposition en voxels, les auteurs montrent la faisabilité d'une identification automatique des sujets et d'une segmentation de l'image cohérente en accord avec les contours entre les objets dans la scène. Cependant, le regroupement et l'identification des voxels se basent sur des méthodes de *sparse PCA* [66] et *sparse subspace clustering* [72] qui requiert un processus itératif faisant appel à de multiples décompositions en valeur propre/singulaire et *décomposition QR* dont la complexité est trop importante. Une fois encore, la phase de segmentation est bien trop lourde puisque la décomposition en voxels n'est pas réalisable en temps réel sur une plateforme conventionnelle. Le recours à des accélérateurs matériels purement parallèle de type GPU permet ce genre de segmentation mais impose une consommation énergétique et un cout bien plus important.

Au vu des éléments présents dans la littérature, la mesure rPPG et le développement d'un capteur photopléthysmographique sans contact se heurtent à deux verrous scientifiques majeurs. Premièrement, la quasi-totalité de la littérature et des méthodes proposées se repose sur la détection du visage et la segmentation de la peau ce qui est une solution sous optimale dans la sélection de la zone d'intérêt. En effet, des perturbations importantes sont induites par les yeux, la bouche et les cheveux qui ne sont pas nécessairement éliminés correctement. Également, l'information rPPG n'est pas uniformément mesurable sur le visage et la segmentation de peau est souvent mal

adaptée pour toutes les teintes de peau. Deuxièmement, dès lors que l'on souhaite s'affranchir de la détection du visage, les couts en calcul des méthodes existantes sont incompatibles avec une implantation temps réel de la mesure rPPG. Ayant pour but de montrer la faisabilité du développement d'un capteur rPPG temps réel, cette thèse sera axée autour de la problématique suivante : comment réaliser une amélioration de la qualité de l'estimation du rythme cardiaque sans besoin de supervision et tout en restant compatible avec une implémentation temps réel ?

Afin de répondre à cette problématique, les travaux de thèse sont développés suivant deux directions. Dans un premier temps : valider une méthode innovante permettant une optimisation de la qualité du signal sans segmentation supervisée de la région d'intérêt. Dans un second temps : proposer des méthodes et optimisations pour réduire la complexité algorithmique afin de satisfaire la contrainte temps réel.

II

Mesure rPPG : algorithme et validation

4

Détection non-supervisée de la région d'intérêt pour la mesure rPPG

Sommaire

4.1 Motivations **42**
4.1.1 Travaux connexes sur la détection non supervisée de la région d'intérêt 43
4.1.2 Segmentation vidéo par superpixels temporels : présentation de TSP . . 45
4.2 Segmentation en superpixels temporels pour la rPPG **47**
4.2.1 Métrique de qualité d'un signal rPPG avec un modèle en fréquence . . 48
4.2.2 Fusion des signaux rPPG 51
4.2.3 Modélisation d'un signal rPPG avec des modèles de Markov cachés . . 52
4.2.4 Métrique de qualité d'un signal rPPG avec les modèles de Markov cachés 55

La segmentation est une étape clé dans de nombreux algorithmes de vision, et tout particulièrement dans le cas de la photopléthysmographie sans contact (rPPG) puisque seule la peau fournit une information exploitable pour la mesure. De plus, il a été montré que le signal rPPG n'est pas distribué spatialement de façon homogène sur la peau. La plupart du temps, les algorithmes pour la mesure rPPG utilisent, comme information d'entrée, les pixels de l'image moyennés dans une zone d'intérêt issue d'une détection du visage possiblement cumulée avec une détection de peau. Cependant, cette méthodologie basée sur un apprentissage supervisée de descripteur d'apparence ou de couleur a plusieurs limitations qui seront détaillées dans ce chapitre.

Dans ce chapitre, nous proposons une approche simple afin de sélectionner les régions de tissus vivants, et ceci de manière *implicite* (*i.e.* qui ne nécessite pas d'utiliser une méthode spécifique à la détection de la peau) en se basant sur leurs pulsatilité. Les images d'entrée sont décomposées en plusieurs superpixels temporels depuis lesquels sont extraits les signaux rPPG. Une mesure de qualité du signal est appliquée pour chaque signal ainsi généré afin de réaliser une fusion pondérée des signaux. Cela permet d'améliorer la qualité du signal pour l'estimation du rythme cardiaque. Puisque le signal avec une pulsatilité plus importante, mesurée avec un rapport signal sur bruit plus important, fourni une information de meilleure qualité, les zones où l'information est prédominante sont favorisées.

Afin de proposer une alternative à l'analyse fréquentielle des signaux rPPG pour en déterminer leur qualité, nous proposons une nouvelle métrique de qualité basée sur un indice de vraisemblance avec un modèle de Markov caché. Cette nouvelle approche permet d'obtenir une estimation des zones d'intérêt plus précise et permet également de diminuer la complexité algorithmique et donc le coût en calcul de cette étape du traitement.

4.1/ Motivations

Les méthodes présentes dans l'état de l'art partagent un pipeline commun [32, 47, 71, 126] : une région d'intérêt est déterminée pour la première image de la vidéo puis suivie pour les suivantes, les canaux R, G et B de la caméra sont combinés pour déterminer le signal rPPG. Le signal mesuré est ensuite filtré puis analysé dans le domaine fréquentiel afin d'en extraire les données physiologiques du sujet (rythme cardiaque, rythme respiratoire, etc.). Les méthodes de rPPG utilisent la valeur des pixels moyennée dans la zone d'intérêt pour construire les signaux R, G et B temporels, ici noté $C(t)$. La sélection de la zone d'intérêt (ROI) est une première étape critique pour obtenir des signaux $C(t)$ de bonne qualité et doit contenir autant de pixels de peau que possible. Plusieurs approches ont été utilisées afin de sélectionner la zone d'intérêt dans un flux vidéo à utiliser pour la mesure rPPG. Dans les travaux préliminaires, la sélection était manuelle pour des vidéos sans mouvement [42, 57]. La méthode courante consiste à utiliser une détection du visage [20] et des méthodes de suivi [88], possiblement affinées avec une segmentation de la peau [106].

Cependant, les méthodes disponibles dans l'état de l'art nécessitent une étape de segmentation qui repose sur un apprentissage *supervisé*. En 2015, Wang et al. [106] proposent une méthode permettant de réaliser la mesure rPPG sans besoin de supervision. En 2017 [136], ils utilisent l'information rPPG afin de réaliser une segmentation des tissus vivants dans la vidéo plus robuste que les méthodes conventionnelles basées sur la couleur de la peau. Dernièrement, Tulyakov et al. [125], ont introduit une approche basée sur des méthodes de complétion de matrice dans laquelle plusieurs signaux issus de plusieurs ROI indépendantes sont fusionnés suivant une procédure d'optimisation. En effet, la distribution spatiale de l'information rPPG permet l'estimation de la qualité et de l'erreur entre les différents signaux temporels. La plupart des méthodes introduites jusqu'ici utilisent une segmentation supervisée alors même qu'elles permettent d'optimiser la sélection des zones d'intérêt à l'intérieur de la zone du visage. La principale problématique est ainsi éludée et la suite de la chaîne algorithmique est commune à toutes les méthodes : les pixels dans la zone d'intérêt sont généralement moyennés et le processus est répété pour chaque nouvelle image. Le résultat de ce processus est la création de multiple signaux $C^i(t)$, avec i la zone d'intérêt considérée. La fusion des données, afin d'obtenir un signal rPPG unique par individu, est fonction des méthodes proposées. Wang et al. [106] se contentent de fusionner les zones entre elles tandis que Tulyakov et al. [125] utilisent la redondance de l'information pour corriger les incohérences dans les signaux mesurés avant de fusionner les différentes zones d'intérêt.

Il a été montré dans plusieurs études que la qualité de la sélection de la zone d'intérêt a un impact direct sur la qualité du signal rPPG [49]. Premièrement, parce que l'utilisation d'un faible nombre de pixels ne permet pas de s'affranchir du bruit de quantification de la caméra. On peut observer par exemple que la qualité du signal rPPG diminue en sous-échantillonnant la zone d'intérêt. Cela peut être expliqué par la réduction du bruit de quantification du capteur par un facteur égal à la racine carrée du nombre de pixels utilisés pour calculer la valeur moyenne de la ROI [98]. Deuxièmement, la qualité du signal est également impactée par la quantité de pixels utilisés pour le calcul de la moyenne qui ne sont pas des pixels de peau [126]. Tous les algorithmes de rPPG souffrent d'une baisse de performance significative dès lors que l'on définit mal la ou les zones d'intérêt. Ces deux remarques sont assez intuitives mais il est en fait plutôt difficile d'obtenir une zone d'intérêt bien définie, qui est stable dans le temps, sans avoir recours à des calculs complexes.

La méthode détaillée dans [109, 130] et les travaux connexes [106] proposent une approche fondamentalement différente basée sur l'utilisation de superpixels ou voxels. Cette segmentation est utilisée pour détecter les zones de l'image qui contribuent significativement à l'information rPPG. De plus, tel que montré par Kamshilin et al. [99], le signal rPPG n'est pas spatialement distribué de façon homogène. Certaines zones de peau fournissent plus d'information que d'autres. Par

exemple, on observe que le rapport signal sur bruit est plus important pour un signal rPPG mesuré sur le front ou les pommettes que sur les autres zones du visage. La Figure 4.1 permet de visualiser la distribution spatiale de l'information rPPG. A droite, la qualité du signal rPPG défini par son rapport signal sur bruit permet d'identifier les zones de peau qui fournissent une information de bonne qualité et qui sont donc à favoriser.

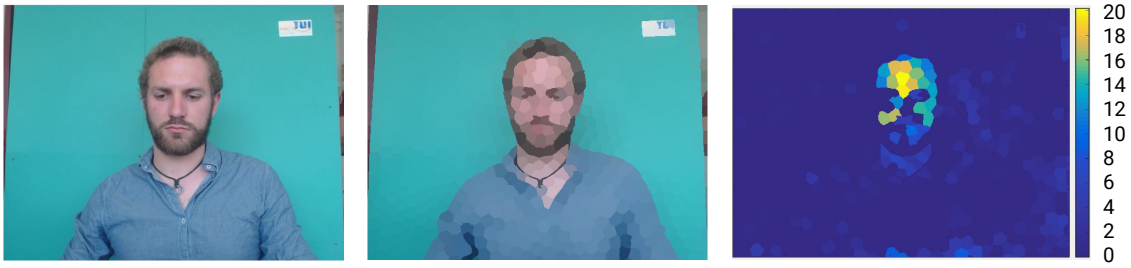


FIGURE 4.1 – Qualité du signal rPPG pour chaque superpixel temporel avec de gauche à droite : image d'entrée, segmentation en superpixels temporels et qualité du signal rPPG (la qualité de la mesure est meilleur dans les tons jaunes/oranges, elle est exprimée en dB).

Pour passer outre ces limitations, nous proposons une nouvelle méthode qui sélectionne implicitement les zones d'intérêt qui représentent des zones de peau vivante et qui vont favoriser la qualité de l'information rPPG. Nous utilisons le terme *implicite* afin de souligner l'approche indirecte de la segmentation des régions d'intérêt. La sélection des zones d'intérêt est basée sur l'hypothèse que les zones de peau vivante, seules, permettent d'observer un signal qui montre une pulsation, périodique ou pseudo périodique. A la différence des approches classiques qui considèrent que tout le visage contribue au signal rPPG. Le flux vidéo est décomposé en plusieurs superpixels temporels desquels sont extraits des signaux rPPG potentiels. La qualité de ces signaux est évaluée en calculant le rapport signal sur bruit (*SNR*) en accord avec un modèle en fréquence. Cela permet de fusionner tous les signaux de la vidéo en utilisant le *SNR* comme pondération afin de favoriser les signaux de meilleures qualités. Cette approche peut être utilisée avec toutes les méthodes de rPPG existantes, puisque la décomposition en superpixels temporels a pour conséquence d'uniquement augmenter le nombre de ROI sans modifier la chaîne algorithmique.

4.1.1/ TRAVAUX CONNEXES SUR LA DÉTECTION NON SUPERVISÉE DE LA RÉGION D'INTÉRÊT

Les travaux de Wang et al. [106] publiés en 2015 et nommés VPS pour *Voxel Pulse Spectral* constituent la contribution la plus proche de notre travail. Bien que fondamentalement différente sur l'approche, ces deux méthodes effectuent une sélection a posteriori des zones d'intérêt en fonction de paramètres liés aux signaux temporels des différentes zones d'intérêt.

Dans leur approche, Wang et al. décomposent le flux vidéo en voxels afin de trouver des zones spatialement, chromatiquement et temporellement cohérentes qui montrent une information de pulsation. Cette approche a initialement été proposée par Gibert et al. [73] où les images sont découpées par une grille fixe où chaque zone est évaluée afin de déterminer la qualité de l'information rPPG mesurée. Bien évidemment, l'utilisation d'une grille fixe pour segmenter l'image est très limitée puisque le sujet doit être parfaitement immobile durant toute la durée de la mesure. De plus les frontières entre les zones ne correspondent pas aux frontières dans l'image et considèrent des pixels très différents dans le calcul de la valeur moyenne des ROIs. L'objectif est d'isoler les zones intéressantes de l'image en maximisant le nombre de pixels considérés. En effet, toutes les zones de peau qui ressortent dans l'identification des voxels qui contribuent à l'information rPPG

sont agglomérées afin d'augmenter le nombre de pixels de peau considérés dans la construction du signal $C(t)$ et le calcul du signal rPPG.

L'utilisation de voxels hiérarchiques consiste en une segmentation en groupes cohérents de pixels du flux vidéo avec plusieurs échelles. Les pixels qui partagent ainsi une information chromatique et spatiale proches dans le domaine temporel sont groupés entre eux. L'étape de segmentation est effectuée avec plusieurs résolutions de voxels afin de s'adapter automatiquement aux objets de la scène avec des tailles variables. L'étape de segmentation en voxels [76] est définie comme un processus de minimisation d'une énergie chromatique et d'une énergie spatiale dans un intervalle de temps court. La méthode de segmentation *k-mean* qui permet le regroupement des pixels en k groupes est appliquée en considérant les informations spatiales, chromatiques et temporels. Dans leurs travaux, les auteurs utilisent trois échelles de résolutions pour la segmentation en voxels du flux vidéo, ainsi la segmentation est effectuée trois fois pour chaque intervalle de temps étudié ce qui augmente le cout en calculs.

Cette approche a plusieurs avantages du point de vue de la segmentation pour la rPPG :
- Elle permet d'obtenir des zones spatio-temporelles cohérentes.
- Elle rend possible la détection de personnes avec des variations d'échelles dans l'image.
- Elle permet d'obtenir une bonne correspondances des frontières entre voxels sur celles des objets de la scène.

Le calcul de la moyenne des pixels est pondéré par la distance des pixels aux frontières des voxels. Cela permet de minimiser d'éventuels problèmes de définition des frontières considérant les mouvements possibles du sujet. Le signal rPPG est calculé depuis le signal $C(t)$ ainsi généré avec la méthode CHROM (*c.f.* section 3.2.2).

Afin de permettre l'identification des contributions par voxels et leurs fusions, une matrice de similarité est construite en prenant en compte les caractéristiques des signaux rPPG issus de la méthode CHROM. Les auteurs considèrent que les signaux issus d'un seul et même sujet sont cohérents entre eux de plusieurs façons. Ils ont par exemple la même fréquence fondamentale et ainsi leur inter-corrélation doit être élevée, de plus les signaux ne doivent pas présenter de déphasage important. Leur inter-corrélation ne doit pas montrer de distorsion importante et enfin, en considérant les signaux temporels comme des vecteurs multidimensionnels, l'angle entre les vecteurs doit être faible.

La matrice de similarité ainsi construite est raffinée en utilisant une décomposition *incremental sparse PCA*. Cette méthode permet d'interpréter la matrice d'entrée comme une combinaison linéaire de vecteurs et valeurs propres $\lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T + ...\lambda_n x_n x_n^T$ avec $x_i \in X$ où X est un ensemble de vecteurs orthogonaux dans un espace de dimension n. En décomposant la matrice de similarité avec l'ensemble X, les différents sujet de la mesure sont isolés dans des vecteurs propres indépendants. Puisque la matrice de similarité est une matrice *sparse* (avec beaucoup d'entrées égales à 0), en considérant que les éléments de l'arrière-plan ne sont pas similaires entre eux, la méthode *Sparse PCA* [66] peut être utilisée pour la décomposer.

En projetant les vecteurs propres *sparse* sur les voxels hiérarchiques, ils obtiennent une carte d'identification des objets de la scène basée sur la décomposition en voxels. Chaque échelle de voxels utilisée contribue à l'identification des zones montrant une pulsation plus importante et qui sont similaires entre elles. Les voxels sont fusionnés par la minimisation d'une fonction de cout : $\gamma E_1 + (1 - \gamma)E_2$, avec E_1 défini pour minimiser la différence de l'énergie entre les vecteurs propres et E_2 défini pour minimiser la différence de l'énergie entre les voxels qui se chevauchent spatialement entre les différentes échelles de segmentation.

L'utilisation de cette décomposition permet de réaliser la détection non supervisée de zones qui montrent une pulsation temporelle dans la vidéo et dans le même temps de différencier et isoler

les sujets dans la scène s'il en existe plusieurs comme dans l'exemple Figure 4.2.

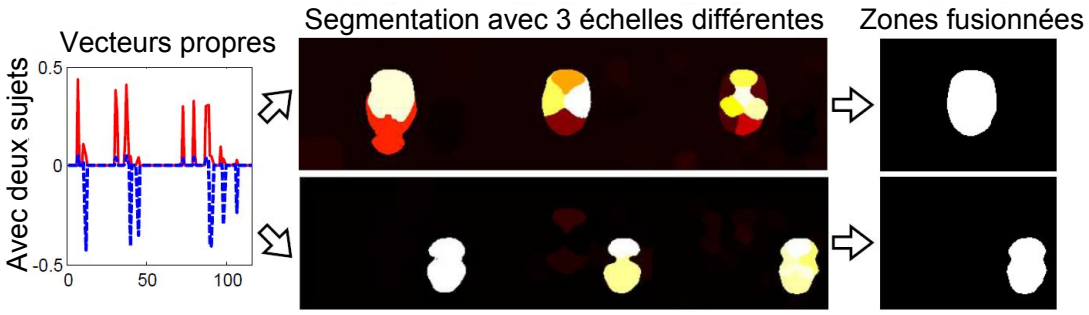


FIGURE 4.2 – Les vecteurs propres similaires entre eux en énergie et pour les différents niveaux de segmentation sont fusionnés afin de déterminer la zone d'intérêt finale [106].

La complexité et le cout en calcul des outils utilisés ici sont bien au-delà de ce qui est envisageable dans le cas d'une intégration pour une utilisation en temps réel. En effet, la décomposition *Sparse PCA* est un processus itératif où des ensembles de vecteurs propres sont évalués afin de minimiser une fonction de cout. De plus, la segmentation initiale en voxels est une opération qui nécessite plusieurs secondes de traitement par image avec les méthodes disponibles dans la littérature, ou bien l'utilisation de plateformes dédiées type GPU [111] afin d'obtenir les performances nécessaires pour une utilisation temps réel.

Cependant, le principe de la détection implicite des zones d'intérêt en fonction des caractéristiques temporelles et fréquentielles des signaux issus des voxels est une idée particulièrement intéressante. Les deux étapes du processus qui sont problématiques dans notre cas sont la décomposition initiale de la vidéo en voxels et l'utilisation de la matrice de similarité. De plus, la méthode ainsi employée ne permet pas de maximiser la qualité du signal. En effet, la zone d'intérêt finale permet de regrouper le plus de pixels de peau possible qui ont des contributions similaires et ainsi réduire davantage le bruit de quantification. Cependant, les zones qui ont une contribution supérieure, en terme de qualité du signal, ne sont pas favorisées dans le processus de fusion.

4.1.2/ SEGMENTATION VIDÉO PAR SUPERPIXELS TEMPORELS : PRÉSENTATION DE TSP

Afin de palier au problème de complexité de la méthode de segmentation en voxels, nous avons décidé d'utiliser une segmentation en superpixels. La segmentation en superpixels ne considèrent pas la dimension temporelle ce qui réduit grandement la complexité de la méthode en comparaison à la segmentation par voxels. De plus, la segmentation est effectuée pour une seule résolution. En 2013, Chang et al. [70] publient leurs travaux sur l'utilisation de superpixels temporellement propagées, méthode nommée *TSP*, afin de segmenter des éléments dans un flux vidéo. Le résultat est la construction d'un ensemble de superpixels temporels qui sont définis sur un ensemble d'images de la vidéo. En cas de discontinuité dans la propagation temporelle, la propagation est interrompue et le superpixel temporel est réinitialisé. Cela est nécessaire dans le cas d'un mouvement dans la scène ou d'un changement de luminosité important par exemple.

La construction des superpixels temporels se base sur un processus itératif qui propage la segmentation en superpixels d'une façon cohérente d'une image à la suivante. De la même manière qu'avec la segmentation en voxels, la décomposition en superpixels peut se résumer à la minimisation d'une fonction E, fonction des énergies E_c et E_s dans l'image, de la forme :

$$E = \sum_{i \in P(t)} (1 - \lambda) E_c(p, k) + \lambda E_s(p, k), \tag{4.1}$$

avec λ la compacité, k le nombre de superpixels et $P(t)$ l'ensemble des pixels de l'image à l'instant t. La compacité permet de modifier la prépondérance relative des deux termes d'énergie E_c et E_s. Ainsi, une compacité plus faible amène à la formation de superpixels plus étendus dans l'image mais plus cohérent dans leur segmentation chromatique. En général, une compacité plus élevée rend la segmentation visuellement plus agréable pour l'utilisateur.

Dans TSP, la segmentation en superpixels est celle de la méthode *SLIC* [58] et se fait dans l'espace chromatique *CIE l*a*b**. La segmentation est effectuée en deux étapes. Premièrement, le regroupement des pixels en superpixels est similaire avec la méthode *k-mean*. Les énergies E_c et E_s sont définies comme la distance Euclidienne entre les pixels évalués et les valeurs moyennes des superpixels appelées *seeds* :

$$E_c(i, k) = \|(l, a, b)_i - (l, a, b)_{\{k\}}\|, \tag{4.2}$$

$$E_s(i, k) = \|(x, y)_i - (x, y)_{\{k\}}\|, \tag{4.3}$$

avec i un pixel dans l'image, $\{k\}$ l'ensemble des *seeds* et $\| \cdot \|$ l'opérateur de distance Euclidienne. Deuxièmement, la cohérence entre les superpixels est accrue en éliminant les pixels isolés et en s'assurant que chaque superpixel est un groupe de pixels contigus et connectés avec d'autres superpixels. Afin de rendre leur algorithme robuste et cohérent vis à vis des perturbations induites par les mouvements dans la scène, les auteurs utilisent un calcul de flot optique entre les images consécutives dans la vidéo pour faire correspondre les mouvements des superpixels entre les images aux mouvements des pixels. Cela permet de faire coïncider les mouvements des superpixels avec les mouvements dans la vidéo. Dans le cas d'un mouvement de la caméra, c'est tout l'arrière-plan de l'image qui est translaté par exemple, cette méthode permet d'être robuste à ce type de perturbation.

Dans le cas de la méthode *TSP*, les auteurs mettent à jour les superpixels à chaque nouvelle image en échangeant les pixels aux frontières des superpixels entre des superpixels adjacents afin de minimiser la variance de l'histogramme des superpixels. Ainsi, à cause des mouvements de la caméra et des occlusions dans la scène, il est nécessaire de pouvoir éliminer des superpixels et d'en définir des nouveaux. Un superpixel *mort* est défini comme un superpixel qui existait dans l'image précédente mais qui n'est plus défini dans l'image courante, ce qui est équivalent à dire que la propagation de la valeur de la *seeds* n'est pas faite entre les deux images. Les choix algorithmiques réalisés dans TSP n'imposent pas un nombre de superpixels constant pour les images successives. Afin d'éviter que les superpixels ne convergent tous en un superpixel unique, ils définissent une fonction de vraisemblance qui diminue avec la taille des superpixels et accroit de fait les chances de création de nouveaux superpixels temporels dans l'image.

Les auteurs comparent leur méthode avec deux méthodes de l'état de l'art en termes de segmentation par voxels / supervoxels. Comme ces méthodes proposent des segmentations en voxels, elles sont utilisées sur la séquence vidéo entière, ou sur des segments de la vidéo. Les résultats montrent que la méthode permet de meilleurs résultats en termes de segmentation des objets de la scène sur l'ensemble de la vidéo. Les superpixels sont capables de mieux s'adapter en cas de mouvements car définis image par image et ainsi varient moins en taille et permettent de maintenir une meilleure cohérence dans la définition des groupes de pixels que les voxels. De plus les méthodes de segmen-

tation par voxels sont inadaptées au traitement d'un flux vidéo en acquisition continue puisque, à moins d'adaptation par l'utilisateur, elles sont employées sur des segments temporels finis.

Ces résultats nous poussent à considérer la segmentation en superpixels et leur propagation temporelle comme une meilleure alternative en termes d'étape de segmentation pour la rPPG. Cependant, dans le cas de la méthode *TSP*, le calcul du flot optique entre les images est une opération très coûteuse en calcul. Aussi, l'utilisation de la méthode *TSP* est limitée à la validation de la segmentation non supervisée de la région d'intérêt et de l'optimisation de la qualité du signal rPPG. La méthode de propagation des superpixels entre les images est bien évidemment un autre des défis de la méthode qu'il nous est nécessaire d'aborder séparément. Cela étant, l'intérêt de la méthode, de notre point de vue, est avant tout la capacité à traiter chaque image de façon indépendante pour l'étape de segmentation. Comme la segmentation en superpixels a un cout en calcul proche des contraintes du temps réel sur des unités de traitement modernes [62], elle est plus intéressante de notre point de vue.

4.2/ Segmentation en superpixels temporels pour la rPPG

Le schéma fonctionnel de la méthode est présenté Figure 4.3. L'algorithme peut être décomposé suivant 4 étapes principales :

1. Le flux vidéo est découpé en superpixels temporels à l'aide de la méthode *TSP*.
2. Un signal rPPG *potentiel* est extrait depuis les signaux $C^i(t)$ pour chaque i-ième superpixel temporel.
3. La qualité du signal est évaluée grâce au rapport signal sur bruit de sa décomposition spectrale.
4. Une moyenne pondérée permet de construire le signal rPPG final qui est utilisé pour l'estimation du rythme cardiaque.

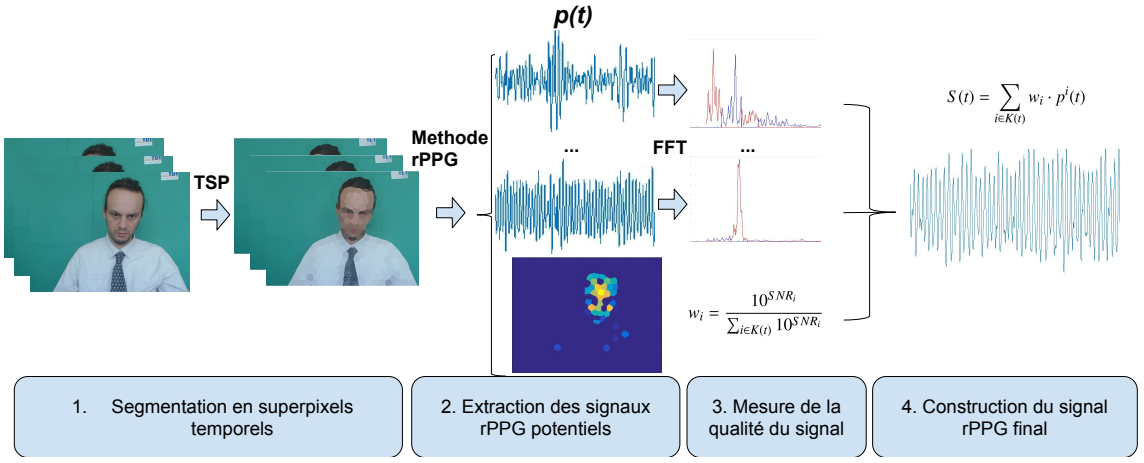


Figure 4.3 – Schéma fonctionnel de la méthode proposée. (1) les images de la vidéo en entrée sont segmentées en superpixels temporels. (2) les signaux rPPG *potentiels* $p^i(t)$ sont extraits depuis les multiples signaux $C^i(t)$ fournis par les superpixels. (3) Une mesure de la qualité du signal est effectuée afin de déterminer les superpixels qui contribuent le mieux à l'information rPPG, les poids w_i sont définis à partir de la mesure de *SNR*. (4) Une moyenne pondérée des signaux est effectuée afin d'obtenir un signal final.

L'intérêt de la méthode est qu'il n'est pas nécessaire d'utiliser un algorithme supervisé de segmentation de la zone d'intérêt. De plus, le signal reconstruit favorise les zones qui fournissent une bonne information rPPG afin de s'affranchir des zones de bruit de l'image, comme l'arrière-plan par exemple. Enfin, notre approche utilise des méthodes de traitement du signal à faible complexité algorithmique.

Afin de valider la méthode proposée, nous l'avons utilisée avec 5 méthodes de rPPG couramment utilisées dans la littérature. Elles sont expliquées dans la section 3.2. La méthode *GREEN* [42] sert de base pour l'évaluation des performances, la composante verte des signaux $C(t)$ est directement utilisée comme signal rPPG. La méthode *GREEN - RED* [39] est également utilisée comme il s'agit de la première contribution à base de combinaison linéaire des canaux chromatiques de la caméra. Cette méthode permet une estimation du rythme cardiaque étonnamment bonne considérant sa simplicité. La méthode PCA [49] est la seule méthode de séparation de sources implémentée dans ces travaux. Cette méthode permet de construire un signal qui maximise la variance des signaux en diagonalisant une matrice de covariance $C(t) \cdot C(t)^T$, comme ceux-ci mesurent tous le signal rPPG, si la principale variation commune dans les canaux est celle du signal rPPG alors le processus de maximisation de la variance dans PCA permet d'extraire ce signal. La méthode ICA n'est pas implémentée pour la raison que le processus est bien plus complexe que les autres approches et que les publications récentes montrent des résultats meilleures pour les méthodes *CHROM* et *POS*. Enfin les méthodes *CHROM* [71] et *POS* [136] qui consistent à projeter le triplet $C(t)$ sur un plan qui permet d'annuler respectivement la composante de réflexion spéculaire de la peau ou la composante spéculaire globale défini par la direction $\{1, 1, 1\}^T$ dans l'espace chromatique RGB sont également implémentées comme ce sont les méthodes de référence de la littérature en termes de précision dans l'estimation du rythme cardiaque.

4.2.1/ Métrique de qualité d'un signal rPPG avec un modèle en fréquence

Notre méthode repose sur l'hypothèse que seule la peau vivante doit fournir une information de pulsation dans la vidéo. Cela peut se traduire en plusieurs informations essentielles qui permettent l'identification des superpixels temporels et signaux associés comme signaux rPPG de bonne qualité. Tout d'abord, dans le processus d'analyse du signal qui est effectué dans la méthode pour chaque superpixel i, un signal $C^i(t)$ est construit avec la valeur moyenne du superpixel temporel pour chaque instant t.

Afin d'identifier les superpixels qui contribuent au signal rPPG (qui segmentent des zones de peau) de ceux qui introduisent du bruit dans l'image nous évaluons tous les superpixels temporels comme des signaux rPPG *potentiels*. Pour limiter les perturbations dans les signaux, on établit une deuxième hypothèse qui est que le signal rPPG est directement modulé par le rythme des pulsations du volume sanguin. Cela se traduit par une représentation fréquentielle dont toute l'information utile doit être contenue dans la bande passante $[0.7, 4]$ Hz ce qui est équivalent à considérer que le cœur bat entre 42 et 240 battements par minute (bpm). Pour les méthodes *GREEN*, *GREEN - RED* et *PCA*, afin d'éliminer un maximum de perturbations, les signaux $C^i(t)$ sont normalisés :

$$\tilde{C}^i(t) = \frac{C^i(t) - \mu_c^i}{\sigma_c^i}, \tag{4.4}$$

avec μ^i la valeur moyenne du signal $C^i(t)$ et σ^i son écart type, on élimine également les variations de très faibles fréquences à l'aide d'une méthode de *detrending* [24]. Pour les méthodes *CHROM* et *POS*, les signaux sont normalisés en les divisant par leurs valeurs moyennes et le *detrending* n'est pas appliqué. Enfin les signaux sont filtrés avec un filtre passe bande de Butterworth dans la

bande passante $[0.7, 4]$ Hz. Une fois les étapes de prétraitement des signaux effectués, les signaux sont combinés suivant les différents algorithmes de rPPG implémentés afin de fournir un ensemble de signaux rPPG *potentiels* notée $p^i(t)$.

La Figure 4.4 permet de visualiser la représentation fréquentielle des signaux issus respectivement d'un superpixel temporel sur le visage et d'un superpixel temporel issu de l'arrière-plan.

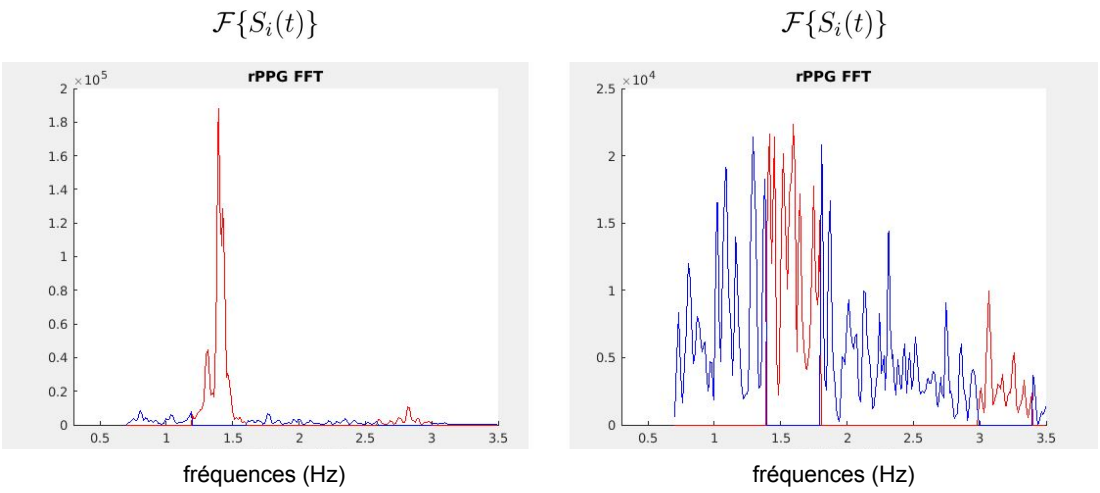


FIGURE 4.4 – Exemple de décomposition fréquentielle pour des signaux rPPG *potentiels* respectivement issus d'un TSP de peau et d'un TSP de l'arrière-plan. Avec en rouge l'information utile du signal et en bleu le bruit.

Dans le domaine fréquentiel, le signal synchrone avec le rythme des pulsations du volume sanguin montre une énergie importante concentrée dans son fondamental, à la fréquence de pulsation cardiaque. Possiblement, son second harmonique et une quantité d'information de modulation autour de son fondamental sont également constitutifs du signal rPPG tel que l'on peut le mesurer dans de très bonnes conditions d'observation. Afin d'estimer la qualité du signal rPPG, nous définissons un modèle en fréquence qui nous permet de calculer le rapport signal sur bruit des signaux (*SNR*). Ce rapport est défini comme le rapport des énergies caractéristiques du signal sur les énergies restantes dans la décomposition fréquentielle et pour une bande de fréquences donnée. Il est exprimé en décibel (dB).

La qualité de l'information de pulsation du i-ième signal $p^i(t)$ est donnée par :

$$SNR_i = 10 \log_{10} \frac{\int_{f_1}^{f_2} h^i_{signal}(f) |\mathcal{F}\{p^i(t)\}|^2 df}{\int_{f_1}^{f_2} h^i_{noise}(f) |\mathcal{F}\{p^i(t)\}|^2 df} \tag{4.5}$$

avec $\mathcal{F}\{p^i(t)\}$ la transformée de Fourrier du signal rPPG pour le i-ième signal. La fonction double porte h permettant d'isoler l'énergie des deux premiers harmoniques est définie comme :

$$\begin{aligned} h^i_{signal}(f) &= [\delta(f - f^i_0) + \delta(f - 2f^i_0)] * \prod(\pm f_r) \\ h^i_{noise}(f) &= 1 - h^i_{signal}(f) \end{aligned} \tag{4.6}$$

avec δ la fonction delta de Dirac, f^i_0 la fréquence du fondamental, convoluée avec la fonction *rect*, notée $\prod$ de semi largeur $f_r = 0.35Hz$ dans nos implémentations. La valeur du SNR sera plus

importante pour un signal de bonne qualité pour un superpixel temporel de peau que pour celui d'un superpixel issu de l'arrière-plan.

Cela se vérifie en faisant coïncider les superpixels temporels dans la vidéo et les valeurs de SNR issus des signaux rPPG *potentiels*. La Figure 4.5 permet de visualiser ce rapprochement. Les superpixels à l'instant t sont colorés en fonction de la valeur du SNR des signaux qui en sont produits. Les valeurs les plus positives tendent vers le jaune tandis que les valeurs nulles et négatives sont bleues.

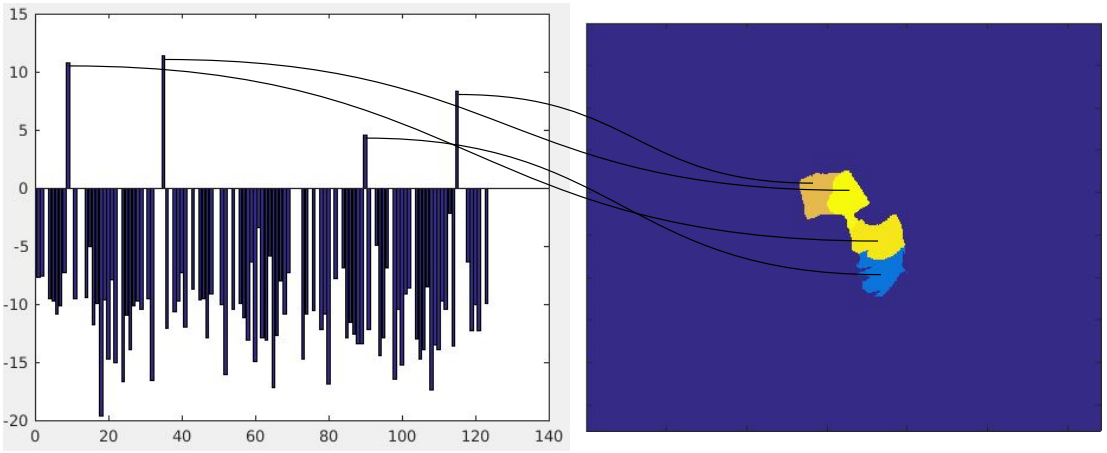


FIGURE 4.5 – Exemples de valeur de SNR en dB et leurs superpixels temporels correspondants.

Ce processus d'évaluation de la qualité respective des superpixels permet d'identifier de façon simple le niveau de contribution de chacun. Cela étant, il reste à fusionner l'information rPPG afin d'obtenir une mesure de meilleure qualité.

4.2.2/ Fusion des signaux rPPG

La différence de contribution exprimée grâce au calcul du rapport signal sur bruit est à la base de l'optimisation de la qualité de l'information rPPG. Une fois le niveau de contribution individuelle identifiée, l'information doit être regroupée de sorte à améliorer la qualité de l'estimation du rythme cardiaque finale. Pour fusionner les différentes contributions, nous nous plaçons dans le domaine temporel et ce sont les signaux $p^i(t)$ qui sont utilisés comme information d'entrée. Le niveau de contribution des signaux dans le domaine temporel est défini par la fonction suivante :

$$c_i = 10^{SNR_i/10}. \tag{4.7}$$

La distribution de l'information rPPG sur l'intégralité de l'image peut être quasi nulle si l'on tient compte que la mesure est faite sur la peau vivante qui peut être masquée (par des vêtements par exemple) et qui dans le cas du visage peut représenter une faible zone de l'image. Une première solution envisagée consiste à supprimer les contributions dont le SNR est négatif (puisque les signaux ne contribuent pas de façon importante à l'information rPPG). Cependant cela revient à négliger une grande partie de l'information dans la vidéo et de plus il est régulier de mesurer un signal rPPG avec une faible qualité mais quand même pouvoir estimer le rythme cardiaque correctement. Cette solution n'est pas la plus adaptée.

Une autre façon de faire consiste à modifier les valeurs c_i afin de favoriser les superpixels temporels qui contribuent le plus et de minimiser le plus possible ceux qui ne contribuent pas. Pour ce faire nous avons supprimé le facteur 10 afin de maximiser les écarts entre les poids, nous définissons alors une nouvelle fonction $\tilde{c}_i = 10^{SNR_i}$.

Une fois les contributions individuelles déterminées, nous construisons le signal rPPG final $S(t)$ en effectuant une moyenne pondérée des signaux $p^i(t)$:

$$S(t) = \sum_{i \in K(t)} w_i \cdot p^i(t), \tag{4.8}$$

avec w_i les poids normalisés :

$$w_i = \frac{\tilde{c}_i}{\sum_{i \in K(t)} \tilde{c}_i} = \frac{10^{SNR_i}}{\sum_{i \in K(t)} 10^{SNR_i}}, \tag{4.9}$$

et $K(t)$ l'ensemble des superpixels à l'instant t. Une fois le signal $S(t)$ établi, le rythme cardiaque est déterminé par la fréquence fondamentale du signal dans le domaine fréquentiel. À cause des possibles perturbations dans la vidéo, la qualité de la segmentation temporelle et ainsi la qualité des signaux $C^i(t)$ est amenée à varier très rapidement. Aussi, les valeurs temporelles de SNR peuvent subir des perturbations hautes fréquences. Afin de réduire les perturbations de la mesure induite par ces variations instantanées, nous introduisons un chevauchement dans le calcul de la valeur des SNR sur une fenêtre glissante de 10 secondes. Cela permet de corriger de façon significative l'estimation du rythme cardiaque puisque les variations de la qualité des superpixels sont lissées par le calcul de la moyenne. Cette approximation permet de ne pas définir de taux de variations maximales entre les estimations du rythme cardiaque comme couramment utilisé dans la littérature [47].

4.2.3/ Modélisation d'un signal rPPG avec des modèles de Markov cachés

L'estimation de la qualité du signal rPPG, et plus particulièrement de l'ensemble des signaux *potentiels* avec l'utilisation des superpixels temporels, impose un grand nombre de transformées dans le domaine fréquentiel afin de réaliser le calcul des *SNR*. En prévision d'une intégration temps réel de notre solution, la méthodologie doit également évoluer sur cet aspect. Le principal intérêt dans l'utilisation d'un modèle en fréquence est que la forme du signal temporel n'est pas prise en compte. Puisqu'une modulation en signal carré permettrait d'obtenir un SNR positif également. Dans le cas de la rPPG cependant, la forme temporelle du signal peut être une indication de la qualité de l'information rPPG sans avoir recours à une transformée de Fourrier. La problématique ici est de déterminer comment modéliser l'apparence du signal. En effet, il serait évident d'utiliser une mesure d'inter-corrélation entre les signaux mesurés et un signal de référence. Cependant, le rythme cardiaque n'est pas constant au cours de la mesure, ce qui implique une multitude de test avec des fréquences différentes [64]. Cette solution est bien évidemment sous optimale puisque le nombre d'opérations augmente de manière très significative et cela avec une complexité minimale pour l'inter-corrélation en $o(n^2)$, avec n le nombre de points du signal.

Dans leurs travaux de 2015, Fan et al. [95] utilisent un modèle de Markov caché (HMM) afin d'estimer le signal PPG. Dans leurs travaux, les auteurs utilisent la caméra d'un smartphone comme capteur d'entrée. L'outil de mesure se présente comme une application téléphonique, il est demandé à l'utilisateur de masquer le capteur optique de façon périodique tandis que le flash de l'appareil est actif. Cela permet d'observer des modulations qui vont permettre d'estimer le rythme cardiaque. Leur méthode repose sur la modélisation statistique temporelle du signal rPPG. Ils définissent le cycle cardiaque et le signal mesurable associé en fonction des événements biologiques qui entrainent la modulation. Ainsi le signal rPPG est toujours défini par le cycle systole puis diastole comme montré Figure 4.6.

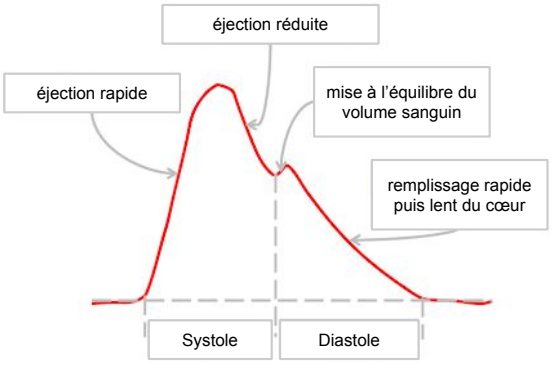


Figure 4.6 – Forme caractéristique du signal rPPG [95].

Les auteurs s'appuient sur les modèles de Markov cachés pour identifier les régularités temporelles des différentes étapes dans le cycle cardiaque (information qui est caché) et les observations disponibles (signal mesuré). Après avoir entrainé le modèle, il est possible de segmenter les observations en séquences d'états grâce au décodage de Viterbi qui permet d'identifier la séquence d'états la plus probable ayant engendré les observations en fonction du modèle entrainé. Le rythme cardiaque est ensuite estimé en calculant le temps nécessaire à chaque cycle depuis une reconstruction du signal globale en alignant toutes les mesures intermittentes.

La forme caractéristique du signal rPPG permet de diviser chaque période en 4 états successifs correspondant aux 4 étapes d'un cycle cardiaque courant : l'éjection rapide du sang (systole),

l'éjection réduite, la mise à l'équilibre volumétrique dans les artères et enfin le remplissage pour la prochaine pulsation comme résumé dans la Figure 4.6. La correspondance avec les états cachés de Markov peut être établi avec pour chaque état caché une étape du cycle cardiaque comme présenté en Figure 4.7.

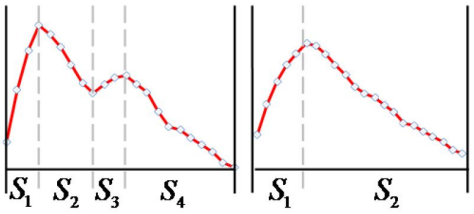


FIGURE 4.7 – Etats du modèle de Markov pour une modélisation en 2 ou 4 états [95]. Chaque état correspond à une étape du cycle cardiaque.

Cependant, il faut tenir compte des conditions de mesure qui ne sont pas toujours optimales et des perturbations potentielles pendant la mesure. Aussi, les auteurs utilisent deux modèles afin de s'adapter au mieux à la qualité du signal pendant la mesure. Un modèle avec 4 états si la mesure est de très bonne qualité et un modèle plus simple avec 2 états si la mesure est bruitée et ne permet pas d'accéder au même niveau de détails dans la forme temporelle du signal rPPG. Qui plus est, il est inutile de considérer des modèles avec plus de 4 états. En effet l'entrainement devient plus complexe et la quantité de données nécessaires pour entrainer des modèles complexes peut devenir important.

Afin d'évaluer le taux de vraisemblance entre le signal mesuré et le modèle entraîné, 4 types d'observation sont définies (*cf.* Figure 4.8).

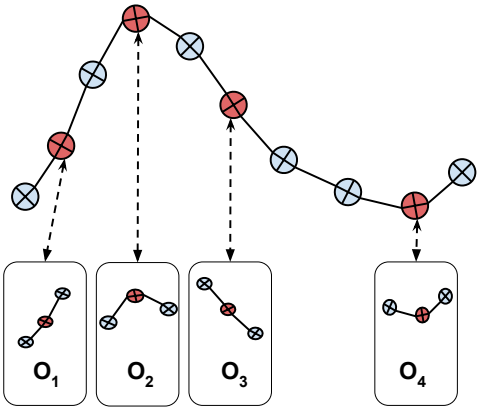


FIGURE 4.8 – Observations possibles sur le signal mesuré.

Avec un signal temporel et en considérant 3 points de mesure successifs, il est possible d'observer :

1. Une croissance continue,
2. Un maximum local,
3. Une décroissance continue,
4. Un minimum local.

Les deux modèles mis en place sont alors des modèles cycliques dont la représentation schématique est donnée Figure 4.9.

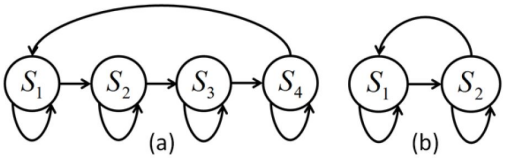


FIGURE 4.9 – Etats et transitions possibles pour les deux modèles de Markov caché utilisés pour la mesure rPPG [95].

On note N le nombre d'états possibles dans le modèle, $S = S_1, ..., S_N$, les N états et s_t l'état à l'instant t. On note également M le nombre d'observations possibles, dans notre cas $M = 4$, et $O = \{O_1, O_2, O_3, O_4\}$ les symboles observables. La distribution initiale est notée $\pi = \{\pi_i\}$ et :

$$\pi_i = P\{s_1 = S_i\}, 1 \leqslant i \leqslant N. \tag{4.10}$$

La distribution des probabilités de transition entre les états est définie par $A = \{a_{ij}\}$, avec :

$$a_{ij} = P\{s_{t+1} = S_j | s_t = S_i\}, 1 \leqslant i, j \leqslant N. \tag{4.11}$$

Comme le cycle des états doit se faire de l'état 1 à l'état N, sans possible retour en arrière (excepté pour le cas a_{N1} pour démarrer un nouveau cycle), toutes les probabilités de transition a_{ij} sont égales à 0 pour la paire de valeur (i, j) où $i > j$. Enfin, la distribution dans l'état j des symboles observables est notée $B = \{b_j(k)\}$, qui définit l'émission des observations, avec :

$$\begin{aligned} b_j(k) = P\{O_k \text{ à l'instant } t | s_t = S_j\} \\ 1 \leqslant j \leqslant N \\ 1 \leqslant k \leqslant M \end{aligned} \tag{4.12}$$

Le modèle de Markov caché utilisé ici peut être caractérisé par trois mesures de probabilité : A, B et π. Par soucis de compacité, les modèles seront notés $\lambda_1 = \{A_1, B_1, \pi_1\}$ pour le modèle à 4 états et $\lambda_2 = \{A_2, B_2, \pi_2\}$ pour le modèle à deux états.

L'estimation des paramètres de transitions et d'émissions $\lambda = \{A, B, \pi\}$ des deux modèles est effectuée avec la méthode Baum-Welch, la distribution initiale π est déterminée d'après le temps moyen des cycles dans l'ensemble de données d'entrainement. Pour la distribution des probabilités de transitions entre états A, la probabilité de rester dans le même état est défini avec une haute probabilité (par exemple 80%) et la probabilité de transitions entre états est plus faible (les 20% restant). Sans oublier que les probabilités de transition vers un état antérieur sont nulles à part pour lancer un nouveau cycle.

La sélection entre les modèles à deux ou quatre états est effectuée avec le critère d'information Bayésien (BIC) défini par :

$$BIC_\lambda = -2 \cdot \ \ln(\ \Pr\{O|\lambda\}) + k \cdot (\ln(n) - \ln(2\pi)), \tag{4.13}$$

où $\ln(\ \Pr\{O|\lambda\})$ est le logarithme népérien de la probabilité cumulée des observations O vis à vis du modèle λ. Ainsi il est possible de sélectionner le modèle le plus vraisemblable en fonction des observations fournies. L'utilisation de la métrique BIC afin de déterminer le modèle le plus

vraisemblable pour une séquence d'observations est une nouvelle approche intéressante. En effet, dans le cadre de la mesure rPPG, nous avons besoin d'une métrique de qualité afin de favoriser les zones d'intérêt qui contribuent le plus à l'information rPPG. La mesure de vraisemblance, même si elle n'est pas prévue pour cela à l'origine, est tout à fait compatible avec cette utilisation.

4.2.4/ Métrique de qualité d'un signal rPPG avec les modèles de Markov cachés

Comme la mesure est réalisée à distance, le signal est bien souvent détérioré, et il est rare de parvenir à observer le léger décrochage à mi-cycle du signal rPPG (*i.e.* appelé nœud dichrotique). De fait, nous travaillons avec un modèle à deux états. L'avantage de cette méthode est qu'elle permet d'évaluer la vraisemblance d'un modèle en fonction de ses paramètres et des observations données. Cela n'impose aucune analyse fréquentielle et la complexité algorithmique est en $o(n)$ avec n le nombre de points du signal. Afin de déterminer la qualité relative des signaux $p^i(t)$ nous utilisons deux modèles à deux états. Le premier modèle est le modèle entrainé sur un ensemble de signaux rPPG noté λ_s. Le second modèle est un modèle équiprobable noté λ_n, qui permet une modélisation d'un bruit blanc. A l'image de la définition du rapport signal sur bruit, la mesure de vraisemblance par rapport à un modèle équiprobable nous permet de définir une référence pour notre métrique de qualité.

Comme la valeur de $\ln(\Pr\{O|\lambda\})$ est dépendante de la longueur de la séquence des observations et que les écarts relatifs entre les estimations du BIC pour chaque signal vont évoluer en conséquence et peuvent ainsi être très faible, il est difficile de définir une fonction de poids pour isoler les superpixels temporels qui contribuent le plus. Nous avons donc remplacé cette solution par l'intégration d'un seuil sur la mesure de BIC des signaux rPPG *potentiels*. Afin de favoriser les signaux qui sont le plus similaire au modèle entrainé, seuls les $x\%$ meilleurs doivent être conservés. Initialement, l'indice *BIC* n'est pas prévu pour faire une comparaison de différents signaux pour un même modèle. Cette utilisation dérivée doit par conséquent être validée. Également, il faut déterminer le bon seuil à utiliser afin de parvenir à réaliser la mesure rPPG avec la meilleure qualité possible.

Il est cependant possible de visualiser la correspondance entre les valeurs fournies par la mesure de SNR et celles fournies par la mesure du *BIC*. La Figure 4.10 permet de visualiser côte à côte les deux métriques de qualité. Ainsi, il semble évident que dans les deux cas le visage est correctement sélectionné et que les zones qui contribuent le plus (pommettes et front) ont un meilleur score.

Afin d'établir l'indice de qualité des différents superpixels temporels, l'indice BIC est calculé pour les deux modèles et pour chaque signal. L'indice de qualité ν est défini comme :

$$\nu = 1 - \frac{BIC_{\lambda_s}}{BIC_{\lambda_n}}, \tag{4.14}$$

Afin de ne conserver que les $x\%$ meilleurs contributions, on définit une valeur $\tilde{\nu}$ qui représente l'ensemble des valeurs ν tel que :

$$\tilde{\nu} = \{\nu, \nu > \max_i(\nu_i) * \frac{x}{100}\}, \tag{4.15}$$

avec i le i-ième superpixel temporel et avec $\max_i(\nu_i)$ la contribution la plus importante dans l'ensemble des superpixels temporels. Le détail de la validation de cette méthode et l'identification de la valeur de x sont présentés en section 5.2.3.

Les poids w_i sont normalisés depuis les valeurs ν :

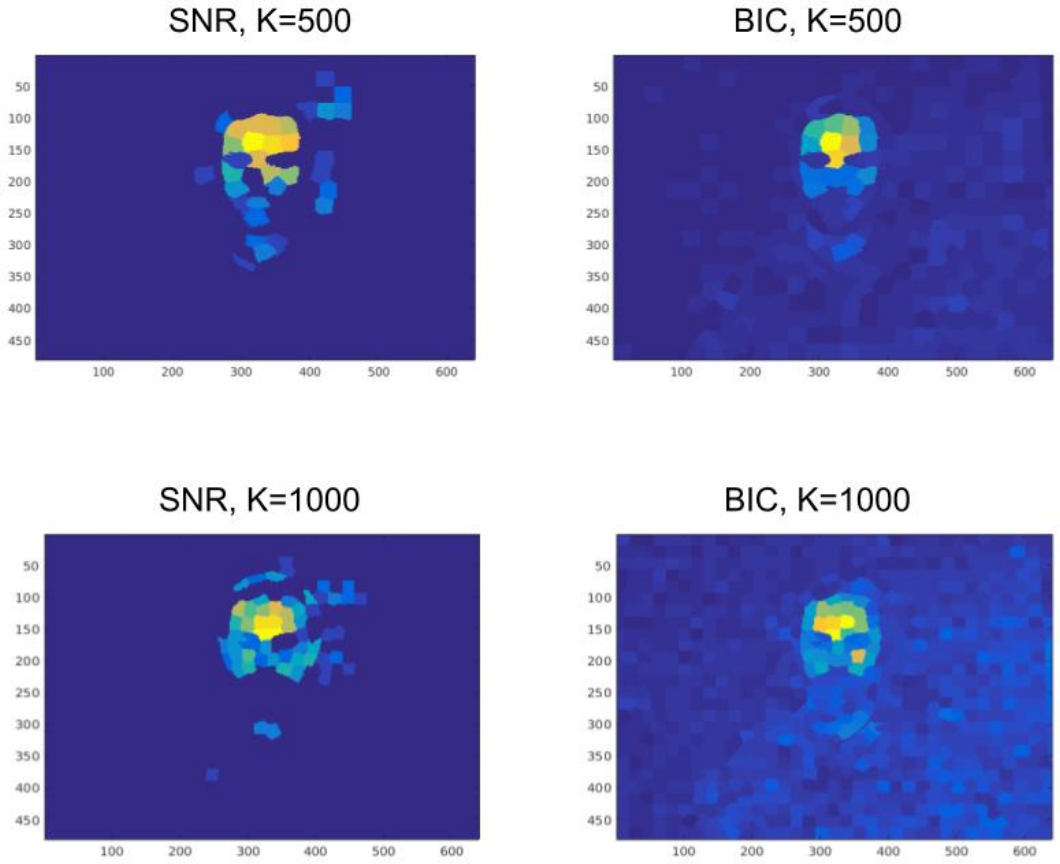


FIGURE 4.10 – Segmentation pour deux résolutions de superpixels temporels différentes $K = 500$ et $K = 1000$ avec correspondance entre la valeur de *SNR* et *BIC*. Plus la teinte tend vers le jaune plus la qualité du signal est élevée.

$$w_i = \frac{\tilde{v}_i}{\sum_{i \in K(t)} \tilde{v}_i}, \tag{4.16}$$

avec $\tilde{v}_i$ l'ensemble des estimations de qualité v des signaux rPPG *potentiels* satisfaisant l'Équation 4.15.

Finalement, la fusion est effectuée à l'aide des états cachés déterminés par la méthode de Viterbi notée $H_i(t)$. Cela permet d'obtenir pour chaque superpixel temporel la séquence d'états la plus probable ayant mené aux observations fournies. Cette méthode agit comme un filtre et permet d'atténuer des perturbations. Le signal final $S(t)$ est construit avec la moyenne pondérée des séquences d'états des différents superpixels temporels pondérés par les poids w_i :

$$S(t) = \sum_{i \in K(t)} w_i \cdot H_i(t). \tag{4.17}$$

Avec H_i et w_i respectivement les états cachés determinés par la méthode de Viterbi et les poids issus du calcul du BIC pour le signal rPPG *potentiel* du superpixel temporel i.

5

Validation expérimentale de la mesure rPPG avec une détection non supervisée de la région d'intérêt

Sommaire

5.1 Protocole expérimental **58**
5.1.1 Données de test : UBFC-rPPG 58
5.1.2 Méthodes implémentées 60
5.1.3 Métriques 61
5.2 Résultats **62**
5.2.1 Validation préliminaire de la segmentation non supervisée : cas d'usage simple 62
5.2.2 Validation de la segmentation non supervisée : cas d'usage réaliste . . . 64
5.2.3 Validation de la métrique de qualité du signal basée sur les modèles de Markov cachés 70
5.3 Discussions **72**

Dans ce chapitre, nous validons expérimentalement notre méthode de segmentation automatique des tissus vivants pour la sélection des zones d'intérêt en utilisant 5 différentes méthodes de rPPG de l'état de l'art : Green [42], Green-Red [39], PCA [49], CHROM [71] et POS [99]. Nous présentons également le choix des métriques employées afin de valider les méthodes proposées. Toutes les contributions sont évaluées au travers de la précision dans l'estimation du rythme cardiaque.

Nous avons évalué nos contributions en utilisant un nouvel ensemble de données publique mis au point par l'équipe de recherche. Les différentes méthodes de l'état de l'art ont été utilisées afin de valider le procédé de segmentation en comparaison avec les méthodes de segmentation couramment employées dans l'état de l'art. Les résultats de nos expérimentations montrent que notre méthode est plus performante que les méthodes de segmentation supervisées.

Également, nous avons validé la méthode avec les modèles de Markov cachés et l'utilisation de l'indice BIC en obtenant de meilleurs résultats avec cette méthode qu'avec l'utilisation du SNR comme métrique de qualité. Cette méthode permet de mesurer le rythme cardiaque alors même que le signal est potentiellement parasité par un bruit haute fréquence. Ce qui est le cas pour certaines des vidéos utilisées dans notre ensemble de données et dont des exemples sont donnés dans ce chapitre.

5.1/ Protocole expérimental

Afin de comparer notre méthode de segmentation de la façon la plus complète possible avec une segmentation supervisée, nous avons implémenté 5 algorithmes de rPPG différents. Cela nous permet de valider notre approche indépendamment de la méthode de rPPG employée dans la chaîne algorithmique. Les implémentations discutées dans ce chapitre ont toutes été réalisées avec Matlab et sont par conséquent bien moins efficace d'un point de vue temps de calcul qu'avec une implémentation dédiée en C++ par exemple. Ainsi, toutes les méthodes implémentées sont évaluées seulement du point de vue de la qualité de l'estimation du rythme cardiaque qu'elles permettent. Les tests ont été réalisés sur une plateforme PC standard équipé d'un processeur Intel i7-4790 @ 3.60 GHz.

5.1.1/ Données de test : UBFC-rPPG

Les méthodes ont été testées sur la base de données UBFC-rPPG enregistrée dans le laboratoire. La base de données est composée de vidéos enregistrées sans compression depuis une caméra PC *Logitech C920 HD*. Les vidéos sont enregistrées avec une fréquence de 30 images par seconde, avec une résolution spatiale de 640x480 dans l'espace chromatique RGB courant et avec une dynamique de 8 bits. Nous utilisons un oxymètre de pouls au doigt *Contec Medical CMS50E* afin d'établir la vérité terrain nécessaire à l'évaluation de la qualité de l'estimation cardiaque. Le signal PPG issu de l'oxymètre est synchronisé avec l'acquisition des images provenant de la caméra. Comme la méthode d'estimation du rythme cardiaque effectuée au sein de l'oxymètre de pouls n'est pas connue, nous utilisons le signal PPG et estimons le rythme cardiaque de la même façon entre les deux signaux PPG et rPPG.

La première partie de la base de données, appelée *dataset simple*, est composée de 7 vidéos (pour un total de 16500 images). Les participants sont assis en face de la caméra à une distance d'environ

Figure 5.1 – Exemple d'images issues des vidéos constituant l'ensemble de données *dataset simple* de la base UBFC-rPPG.

1 mètre avec le visage visible. Les personnes devaient demeurer le plus immobile possible avec les yeux fermés durant la durée de l'enregistrement (en moyenne 1 minute 30). En pratique, dans plusieurs séquences, il est possible d'observer quelques mouvements au début de la vidéo. Tous les enregistrements ont été effectués en intérieur dans un environnement de travail à l'arrière-plan varié et en utilisant la lumière ambiante, soleil et luminaires, comme source lumineuse sans aucun contrôle de l'intensité de la lumière ou de ses variations, un exemple d'images est présenté Figure 5.1. Ce premier ensemble de données nous permet d'avoir une base de données de référence très simple pour le développement des méthodes de rPPG.

Le second ensemble de données, appelé *dataset réaliste*, est bien plus important. Il est constitué de 43 vidéos d'une durée de 2 minutes en moyenne. Les participants sont assis en face de la caméra à environ 1 mètre avec le visage visible. Il leur a été demandé de jouer à un jeu au rythme croissant qui permet d'observer une accélération du rythme cardiaque. Cette stimulation permet d'obtenir une base de données plus réaliste d'un utilisateur à son poste de travail par exemple. De plus la variation cardiaque, si elle est plus ou moins marquée chez les différents participants, permet de mettre à l'épreuve les méthodes testées. Tous les enregistrements ont été effectués en intérieur et en utilisant la lumière ambiante comme source lumineuse. Cette fois ci l'arrière-plan est uni et n'implique pas de possibles éléments perturbateurs comme dans le premier ensemble de données (passage de personnes ou mouvements variés en arrière plan par exemple).

FIGURE 5.2 – Exemple d'images issues des vidéos constituant l'ensemble de données *dataset réaliste*.

Ces deux ensembles de données sont publics et disponibles en téléchargement après demande sur la page du projet [1]. Notre méthode a dans un premier temps été validée grâce au *dataset simple* [109] puis dans un deuxième temps sur le *dataset réaliste* [130].

1. https://sites.google.com/view/ybenezeth/ubfcrppg

5.1.2/ Méthodes implémentées

Notre méthode de segmentation implicite des tissus vivants grâce à la détection de signaux rPPG est directement comparée avec trois méthodes couramment employées pour la segmentation supervisée dans le cas de la mesure rPPG. Ces méthodes sont nommées *face*, *crop* et *skin* telles qu'elles sont employées respectivement dans [42], [47] et [106].

La méthode *face* repose sur la détection et le suivi du visage du sujet. Dans notre implémentation, la détection du visage et son suivi ont été effectuées en utilisant les méthodes de Viola-Jones [20] et de Kanade-Lucas-Tomasi [9] telles qu'implémentées dans la boîte à outils Matlab *computer vision*. Avec cette méthode, les pixels de toute la zone d'intérêt déterminés par la détection du visage sont moyennés afin de construire le signal $C(t)$. La méthode *crop* définit une zone d'intérêt centrée sur la détection du visage et dont la largeur est égale à 60% de la zone d'intérêt délimitée par la détection du visage. Cette méthode permet d'affiner la sélection des pixels utilisés pour construire la moyenne RGB. Enfin, la méthode *skin* utilise la méthode de détection de peau comme proposée par Conaire et al. [34] sur la zone d'intérêt déterminée par la détection du visage. Cette dernière approche permet d'éliminer l'arrière-plan et toutes les zones qui ne correspondent pas à de la peau dans un intervalle de couleurs. La Figure 5.3 permet de voir un exemple de segmentation effectué par les trois méthodes.

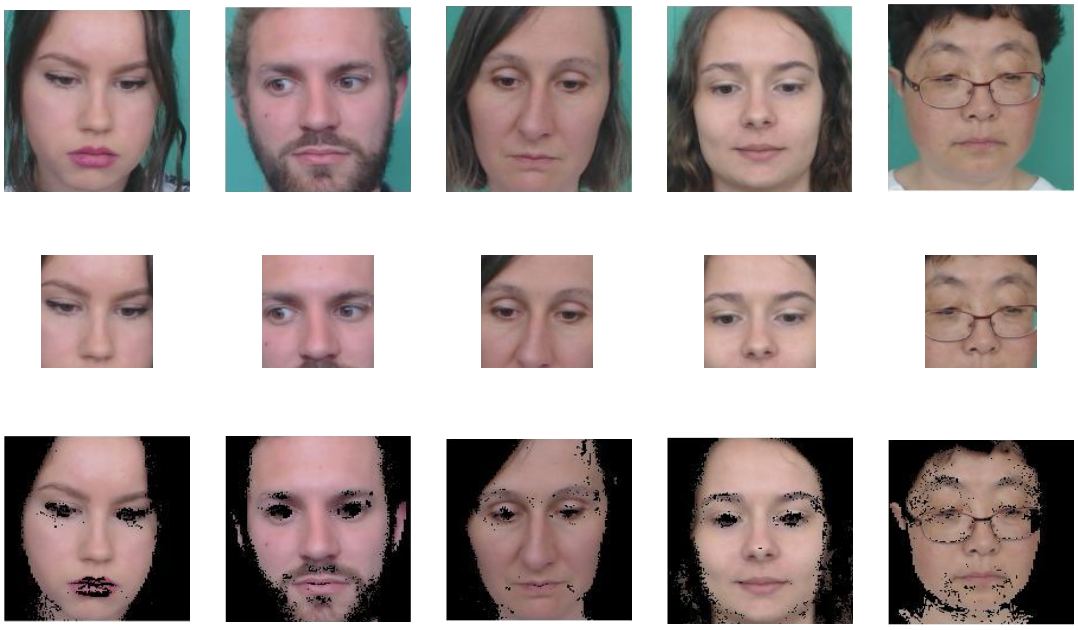


Figure 5.3 – Exemple de zones d'intérêt déterminées par les trois méthodes de segmentation *face* en ligne 1, *crop* en ligne 2 et *skin* en ligne 3.

Ces trois méthodes sont utilisées pour construire le signal $C(t)$ en moyennant tous les pixels des zones d'intérêt à chaque instant. Instinctivement, on suppose que les méthodes *crop* et *skin* doivent permettre de meilleurs résultats puisque la majorité des éléments qui ne sont pas de la peau sont éliminés.

La phase de traitement du signal est la même que celle décrite en 4.2.1 : les signaux sont normalisés puis filtrés dans la bande passante qui nous intéresse avant d'être utilisés dans les différents algorithmes pour extraire le signal rPPG. Toutes les méthodes suivent la même chaîne algorith-

mique de sorte qu'il n'y ait aucune différence dans le traitement des données entre les méthodes de segmentation. Pour toutes les vidéos, le signal rPPG est calculé dans une fenêtre glissante de 20 ou 30 secondes. Le rythme cardiaque est calculé en considérant la fréquence du fondamental de la transformée de Fourrier du signal rPPG. La même procédure est appliquée au signal PPG fourni par le capteur au doigt et aux signaux issus des différentes méthodes de segmentation et à notre méthode.

5.1.3/ Métriques

Nous avons utilisé cinq métriques différentes dans nos travaux. La multiplication des métriques permet d'avoir un regard plus objectif sur l'impact des méthodes et de d'identifier leurs points forts et faiblesses. Les métriques suivantes ont été utilisées afin de comparer les méthodes implémentées :

— Le facteur de corrélation de Pearson r est le facteur d'inter-corrélation entre l'estimation du rythme cardiaque depuis la mesure rPPG et l'estimation du rythme cardiaque depuis le signal PPG de la vérité terrain.
— L'erreur quadratique moyenne (Root mean square error : RMSE) est l'erreur quadratique calculée entre les valeurs de rythme cardiaque mesuré et la vérité terrain.
— La précision de la mesure à 2.5 et 5 battements par minute (bpm) est également calculée puisque la fiabilité des appareils de mesures pour le commerce médical requiert une quantification de la précision de la mesure. Elle est calculée en considérant le pourcentage des mesures ayant une erreur absolue inférieure à 2.5 et 5 bpm avec la vérité terrain.
— La valeur moyenne du rapport signal sur bruit (SNR) est calculée afin de donner une indication de qualité global pour la méthode. Pour ce calcul, la bande passante est définie par la bande $[0.7; 3.5]$ Hz et la largeur du Dirac est définie à 0.1 Hz.
— La représentation de Bland-Altman qui mesure le taux de correspondance entre les valeurs mesurées et la vérité terrain. Les lignes représentent la valeur moyenne et la limite à 95% de la correspondance entre les données testées.

Par soucis de lisibilité, la représentation de Bland-Altman ne sera employée que dans la validation préliminaire de la méthode proposée. Le grand nombre de vidéos comprises dans le *dataset réaliste* et la grande quantité d'estimation du rythme cardiaque résultante rendent cette représentation confuse dans le cas d'un grand nombre d'éléments de mesure pris en compte.

5.2/ Résultats

La validation de la méthode proposée de segmentation implicite de tissus vivants pour la mesure rPPG est faite au travers de deux phases expérimentales. Dans un premier temps, nous validons notre méthode sur un cas d'usage simple en utilisant l'ensemble de données *dataset simple*. Cette première expérimentation nous permet de montrer que notre méthode de segmentation améliore la qualité de la mesure comparée à une segmentation par détection du visage. Dans un second temps, nous réalisons une phase expérimentale afin de tester notre approche dans un cas d'usage plus réaliste. Enfin, une dernière expérimentation nous permet de valider l'usage de la métrique de qualité *BIC* via une comparaison avec la métrique de qualité *SNR* appliquée à notre méthode.

5.2.1/ Validation préliminaire de la segmentation non supervisée : cas d'usage simple

Dans cette étude préliminaire nous avons limité notre approche en utilisant uniquement la méthode CHROM [71] de l'état de l'art. Ce choix est réalisé en raison de la popularité de la méthode dans la littérature en raison de son efficacité pour extraire un signal rPPG de bonne qualité et de sa faible complexité algorithmique. De plus nous utilisons uniquement l'ensemble de données *dataset simple*. La méthode de segmentation proposée est comparée à une approche plus classique utilisant une détection du visage couplée à une segmentation des pixels de peau comme illustré Figure 5.4.

Figure 5.4 – Exemple de segmentation du visage et de la peau pour la mesure rPPG depuis le *dataset simple*.

Cette étude se divise en deux phases expérimentales. Dans notre première expérience avec cette méthode, nous avons évalué la robustesse de notre segmentation en faisant varier la quantité de superpixels temporels dans la vidéo. Ainsi, la résolution en TSP varie pour l'ensemble de valeurs $\{50, 200, 400, 600, 800, 1000\}$. La Figure 5.5 présente les résultats pour la précision de la mesure avec une marge d'erreur respectivement pour 2.5 et 5 bpm. Les résultats sont stables de 95% à 99% de précision avec une erreur maximale de 5 bpm et de 89% à 95% de précision avec une erreur maximale de 2.5 bpm. À mesure que le nombre de pixels dans chaque TSP diminue, l'erreur de quantification augmente mais semble être compensée par l'utilisation de la moyenne pondérée pour la fusion des signaux. Pour l'expérience suivante nous avons travaillé avec une résolution fixe de 200 TSP comme l'impact de la résolution sur cet ensemble de méthode et en utilisant la méthode CHROM n'est pas significatif.

Dans la deuxième expérience nous avons comparé les estimations du rythme cardiaque entre notre méthode et la méthode de référence (segmentation du visage et de peau nommée CHROM dans le tableau suivant). Les Figures 5.6 et 5.7 montrent les résultats obtenus en termes de corrélation et de représentation de Bland-Altman pour les deux méthodes.

Avec une résolution en superpixels temporels de 200 TSP, nous obtenons un facteur de corrélation

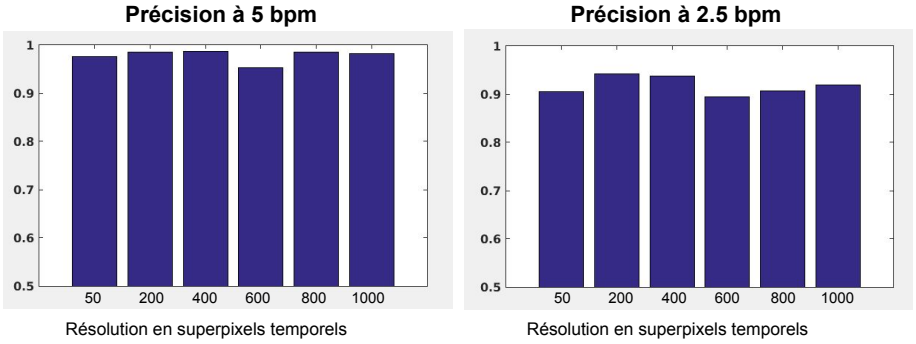


FIGURE 5.5 – Précision de la mesure rPPG avec une erreur maximale de 2.5 bpm et 5 bpm pour une résolution en superpixels temporels variant de 50 à 1000 TSP.

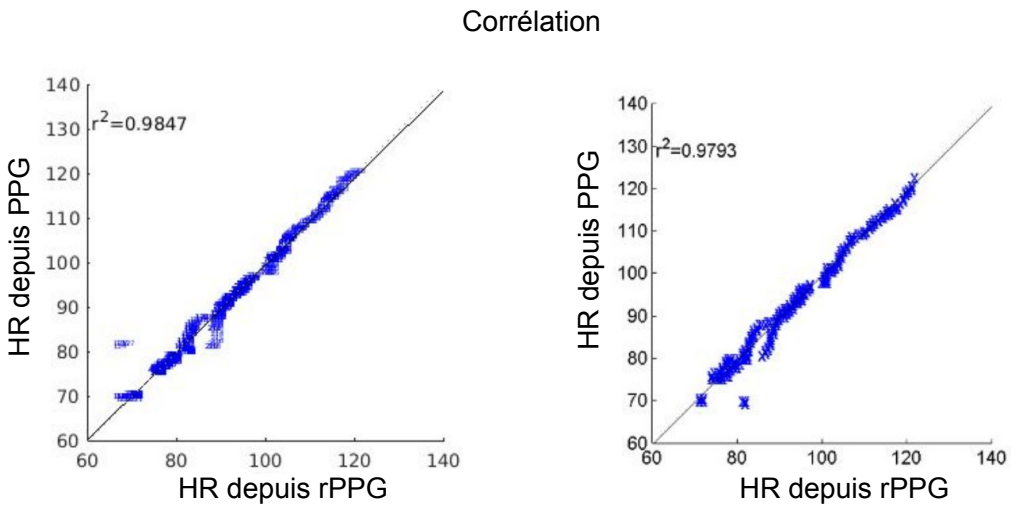


FIGURE 5.6 – Corrélation entre les données issues de la mesure rPPG et de la vérité terrain. Avec à gauche les résultats de notre méthode et à droite les résultats de la méthode de référence. L'abscisse représente les données de la mesure rPPG et en ordonnée la vérité terrain.

d'une valeur de 0.9847 comparé à 0.9793 pour la méthode de référence. La représentation de Bland-Altman montre que l'erreur moyenne est de -0.36 bpm sur l'ensemble des estimations avec 95% des valeurs comprises entre 3.4 bpm et 2.7 bpm d'erreur absolue. La méthode de référence montre une erreur moyenne de -0.55 bpm avec 95% des estimations comprises entre 3 bpm et 4.1 bpm en erreur absolue. La première conclusion est que les résultats sont très bons. L'estimation du rythme cardiaque est très efficace dans les deux cas. Notre méthode améliore légèrement les résultats mais pas de façon significative, ce qui est naturel compte tenu du très bon niveau d'estimation général.

Le tableau 5.1 permet de résumer l'ensemble des résultats obtenus pour toutes les métriques. De meilleurs résultats sont obtenus avec notre méthode pour alternativement une résolution en superpixels temporels de 200 et 400 TSP.

Cette première expérimentation nous permet de valider la segmentation implicite de la ROI pour la rPPG. Il est nécessaire de tester notre méthode sur un ensemble de données plus conséquent et présentant un challenge supérieur afin de valider l'intérêt de la fusion pondérée des signaux comme introduite par notre méthode.

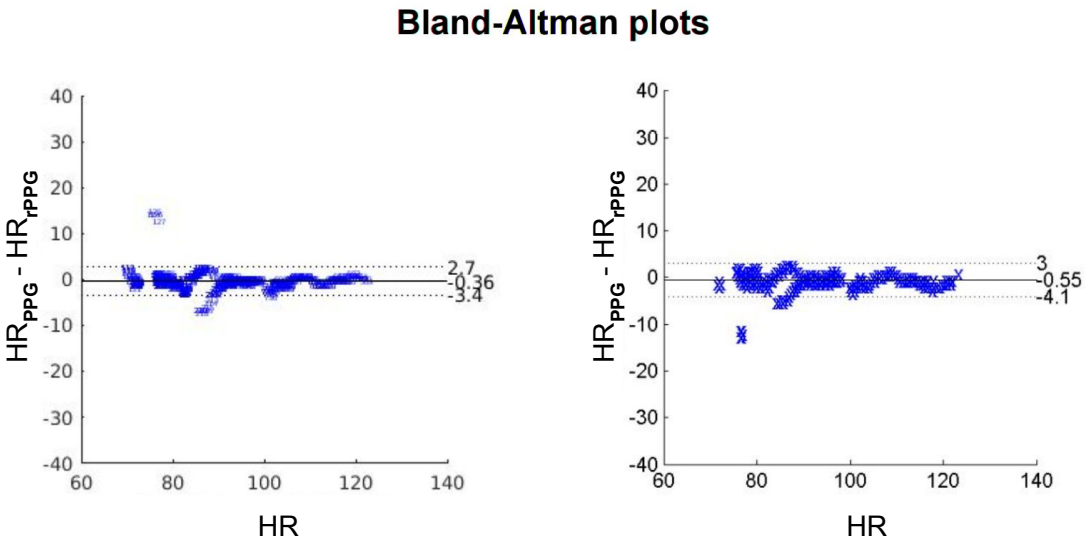


FIGURE 5.7 – Représentation de Bland-Altman pour toutes les estimations issues de toutes les vidéos. Avec à gauche les résultats de notre méthode et à droite les résultats de la méthode de référence. L'abscisse représente la vérité terrain et en ordonnée la différence entre le rythme cardiaque HR depuis le capteur au doigt (HR_{PPG}) et le rythme cardiaque HR estimé par la mesure rPPG (HR_{rPPG}).

TABLE 5.1 – Résumé de l'ensemble des résultats obtenus dans la comparaison de la méthode proposée avec deux résolutions de superpixels temporels N=200 et N=400 et de la méthode de référence notée *skin*.

Métriques	N=200	N=400	*skin*
Précision à 5 BPM	0.986	**0.987**	0.979
Précision à 2.5 BPM	**0.943**	0.938	0.939
RMSE	1.50	**1.43**	2.43
Corrélation *r*	**0.985**	0.984	0.979

5.2.2/ VALIDATION DE LA SEGMENTATION NON SUPERVISÉE : CAS D'USAGE RÉALISTE

Dans cette seconde étude, nous avons comparé les performances de la segmentation non supervisée pour la mesure rPPG en fonction des méthodes de rPPG employées. Également, puisque les résultats ne montraient pas d'avantages significatifs pour notre approche en raison des très bons résultats globaux sur l'ensemble de données *dataset simple*, nous avons cette fois évalué notre méthode sur un ensemble de données plus vaste et plus complexe *dataset réaliste* dont un exemple d'images est proposé Figure 5.2. Dans cette étude, nous avons réalisé deux expérimentations distinctes. La première vise à définir les performances et limitations de notre méthode en fonction de la méthode de rPPG employée. La seconde expérience vise à déterminer l'impact de la résolution en superpixels temporels dans l'image sur la qualité de l'estimation du rythme cardiaque.

Les résultats globaux sont disponibles dans le tableau 5.2. Les meilleurs résultats par algorithme de rPPG (*Green*, *Green – Red*, *PCA*, *CHROM* et *POS*) sont en gras et les meilleurs résultats par métrique (*RMSE*, *SNR* moyen, etc.) sont en rouge. Pour cette étude, nous avons évalué notre méthode avec une résolution de superpixels de 150 TSP. De multiple résolution ont été testées également et les conclusions de cette expérience sont présentées plus loin.

TABLE 5.2 – Comparaison des méthodes de segmentation *FACE*, *CROP* et *SKIN* décrites en section 5.1.2 appliquées aux 5 méthodes de rPPG : *Green*, *Green – Red*, *PCA*, *CHROM* et *POS* décrites en section 3.2. Les métriques (présentées en section 5.1.3) sont calculées pour une résolution en superpixels de 150 TSP.

Métriques	**Méthodes**	**Green**	**Green-Red**	**PCA**	**CHROM**	**POS**
Corrélation	***FACE***	0.522	0.343	**0.442**	0.750	0.729
	CROP	**0.560**	0.422	0.392	0.796	0.759
	SKIN	0.491	0.535	0.435	0.822	**0.795**
	Notre méthode	0.516	**0.595**	0.462	**0.826**	0.782
Estimation à 2.5 BPM	***FACE***	0.614	0.409	0.516	0.766	0.730
	CROP	0.683	0.650	0.589	0.862	0.863
	SKIN	**0.739**	0.782	**0.701**	0.861	0.862
	Notre méthode	0.628	**0.821**	0.509	**0.890**	**0.885**
Estimation à 5 BPM	***FACE***	0.321	0.135	0.110	0.581	0.571
	CROP	0.780	0.748	0.689	0.940	0.952
	SKIN	**0.821**	0.855	**0.794**	0.943	0.941
	Notre méthode	0.669	**0.904**	0.557	**0.961**	**0.958**
RMSE	***FACE***	18.41	24.20	23.36	9.52	10.03
	CROP	**10.04**	12.81	16.05	3.78	3.84
	SKIN	10.13	8.70	**11.08**	3.16	**3.70**
	Notre méthode	16.84	**6.77**	21.22	**2.39**	6.77
SNR moyen	***FACE***	-2.08	-4.04	-2.98	1.68	0.10
	CROP	-0.23	0.27	-0.83	4.08	3.42
	SKIN	**0.60**	3.25	**0.54**	4.32	3.25
	Notre méthode	-0.23	**3.95**	-2.30	**4.97**	**5.18**

Notre première observation est que notre méthode améliore les performances comparées aux autres méthodes de segmentation classiques avec les méthodes *CHROM*, *POS* et *Green – Red*. Par exemple, la métrique de *SNR* moyen qui donne un indice de la qualité moyenne des signaux rPPG calculés sur l'ensemble des vidéos de l'ensemble de données utilisé et pour la méthode de rPPG *POS*, qui est la méthode de rPPG la plus récente et la plus efficace d'après la littérature, notre méthode accroît la qualité des signaux par 59% (à peu près 2 dB) comparé aux méthodes de segmentation classiques. Cette amélioration est clairement significative.

Il est intéressant de noter que notre méthode ne permet pas d'améliorer les performances des méthodes *Green* et *PCA* et que la méthode de segmentation *skin* est la plus efficace avec ces méthodes. Cela peut être attribué aux performances médiocres de ces deux méthodes. En effet, *GREEN* et *PCA* obtiennent en moyenne les pires performances dans notre étude. En conséquences, la valeur du *SNR* est moins bonne pour les zones de peau et il est impossible de les favoriser par rapport aux zones de l'arrière-plan.

Pour illustrer cette observation, la Figure 5.8 montre les valeurs de *SNR* transposées pour chacun des superpixels temporels et pour chaque méthode de rPPG implémentée dans cette étude. Il est clairement visible que le front et les pommettes sont correctement identifiées avec les méthodes de rPPG *CHROM*, *POS* et *Green – Red* tandis que les TSP qui contribuent à l'information rPPG sont

incorrectement répartis à travers l'image pour les méthodes *GREEN* et *PCA*. A cause du fait que la mesure de périodicité / pulsation est effectuée sur une bande de fréquence très étroite [0.7; 3.5] Hz avec une largeur de la fonction porte importante 0.35 Hz, la mesure de *SNR* considère que 27% de l'énergie contenu dans la représentation fréquentielle est du signal utile si le premier harmonique est détecté dans la bande [0.7; 1.75] Hz. Cela peut mener à une mauvaise segmentation des tissus vivant en considérant des signaux bruités comme bon.

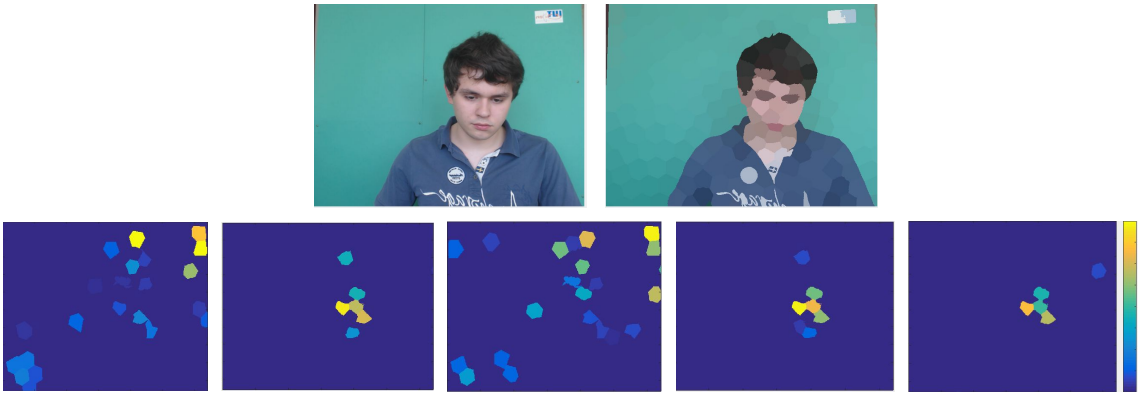


FIGURE 5.8 – Valeur de *SNR* calculé pour chaque TSP, pour des valeurs de *SNR* comprises dans la bande [0; 8] dB. La première ligne montre l'image d'entrée et sa décomposition en superpixels temporels. La seconde ligne montre les valeurs de *SNR* pour les méthodes de rPPG (de gauche à droite) : *Green*, *Green – Red*, *PCA*, *CHROM* et *POS*.

Également, le tableau 5.3 présente les résultats des temps de calcul nécessaires pour l'application des méthodes de rPPG sur une fenêtre de 20 secondes et les étapes de prétraitements : normalisation, *detrending* et filtrage.

TABLE 5.3 – Comparaison des temps de calcul des étapes de prétraitement et du calcul des méthodes de rPPG sur un signal de 20 secondes.

	Green	**Green-Red**	**PCA**	**CHROM**	**POS**
Time (ms)	12.13	12.27	15.5	13.21	13.33

Ces temps de calcul sont fournis à titre indicatif uniquement. En effet, comme ils sont mesurés sur du code non-optimisé dans Matlab, ils ne sont pas complètement fiables. Cependant, ils permettent de mettre en perspective leur complexité relative. Il est intéressant de noter que la différence entre les méthodes est très faible, à peu près 1 ms (à l'exception de *PCA*) et sont négligeables en comparaison d'autres étapes de traitement comme TSP par exemple. De plus il est bon de noter que chaque mesure a été effectuée pour un seul signal $C(t)$. En conséquence, avec notre méthode, les temps de calcul seraient multipliés avec un grand nombre de TSP. Enfin il est possible d'optimiser la méthode proposée ici afin de réduire la complexité imposée par le nombre de superpixels temporels en parallélisant le traitement des signaux $C^i(t)$. En effet, comme chacun des superpixels est indépendant dans l'évaluation de sa contribution, cette méthode est tout à fait adaptée à une implémentation sur une cible matérielle multi-cœurs par exemple.

Dans une seconde expérience, nous avons évalué l'influence de la résolution spatiale de la segmentation défini par le nombre de superpixels temporels comme hyper-paramètre. En accord avec les résultats présentés dans le tableau 5.2, la méthode de rPPG *CHROM* est utilisée pour toutes les études suivantes comme elle fournit les meilleurs résultats sur la majorité des métriques. Les

résultats sont présentés dans la Figure 5.9 et la Figure 5.10. Le nombre de superpixels temporels varie de 100 à 600 afin de mesurer les variations de performance dans l'estimation du rythme cardiaque en fonction de la résolution en TSP. Pour pouvoir comparer, nous avons également ajouté les valeurs obtenues avec les méthodes de segmentation classiques *skin* et *crop*.

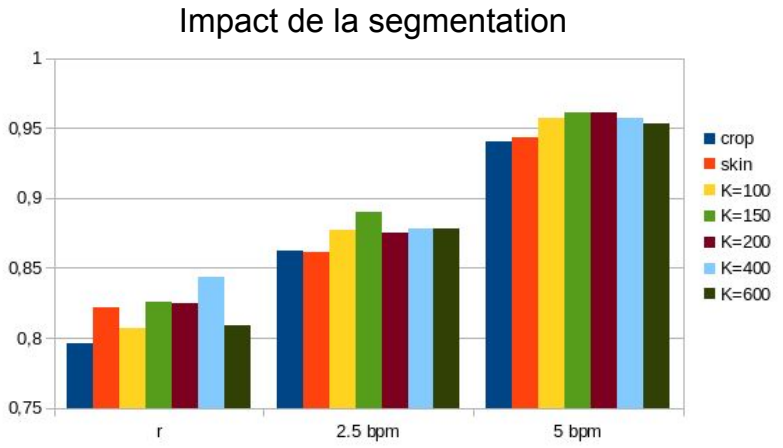


FIGURE 5.9 – Résultats des métriques : corrélation et précisions pour une erreur maximale de 2.5 bpm et 5 bpm pour un nombre de TSP variant de 100 à 600.

Les performances de notre méthode égalent celles obtenues à partir des méthodes *crop* et *skin*, et ceci quelle que soit la résolution spatiale choisie et quelque soit les métriques considérées. L'analyse plus fine des résultats permet même de montrer des améliorations vis à vis de ces méthodes de référence. La corrélation reste assez stable pour les différentes résolutions de superpixels temporels avec des résultats variant de 80% à 85%. Les métriques de précision sont constamment très hautes avec au minimum 95% de bonne estimation pour toutes les résolutions de superpixels temporels avec une erreur maximale de 5 bpm. Notre méthode de segmentation améliore les performances de la mesure avec la métrique de précision pour une erreur maximale à 2.5 bpm avec, dans le cas le plus défavorable, 87% d'estimations correctes et au mieux 89% d'estimations correctes. Ces trois métriques, normalisées dans l'intervalle $[0; 1]$, montrent que notre méthode améliore l'estimation du rythme cardiaque pour toutes les résolutions. De plus, avec une résolution en superpixels temporels égale à 150 TSP, notre méthode obtient les meilleurs résultats en termes de précision, indiquant une résolution optimale pour notre méthode et considérant la résolution spatiale de 640x480 de l'ensemble de données évaluées. Enfin, il est important de noter que, quelle que soit la résolution en TSP appliquée, l'erreur quadratique *RMSE* reste inférieure à 3.5 et qu'avec une résolution de 150 superpixels, le *SNR* moyen sur l'ensemble de test *dataset réaliste* est très bon avec une valeur de 5 dB.

Enfin, les résultats obtenus avec toutes ces métriques montrent la même tendance. En effet, on observe une décroissance des performances continues pour des résolutions s'éloignant de la résolution optimale de 150 TSP. Plusieurs points peuvent être soulevés relativement à cette observation. Premièrement, bien que les performances restent stables avec la résolution en superpixels variant de 100 à 600, comme le montre la Figure 5.11, il y a une perte d'information rPPG comme certaines zones du visage ne sont plus considérées avec le calcul du *SNR* pour des résolutions en superpixels temporels autour de 150. Cela peut être expliqué par la combinaison d'au moins deux phénomènes. D'un côté, la qualité de la mesure se détériore clairement à mesure que l'on réduit la taille des superpixels à cause du bruit de quantification. D'un autre côté, le signal rPPG n'est pas spatialement distribué de façon homogène, et en augmentant la taille des superpixels de plus en plus de zones du visage sont mélangées et les contributions indépendantes ne sont plus distinguées.

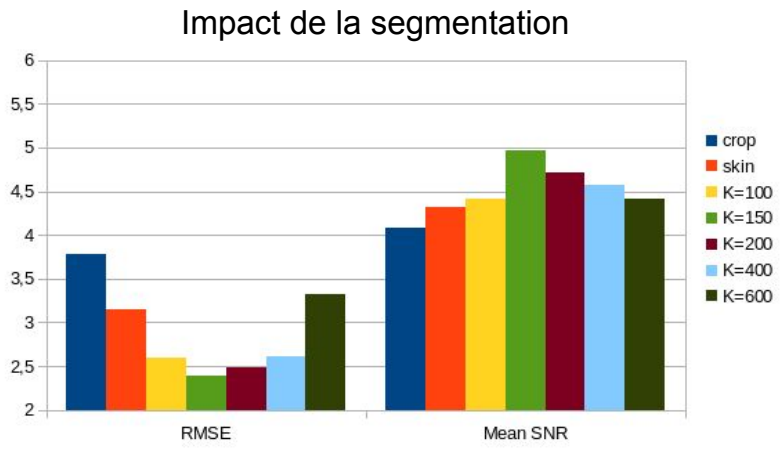


FIGURE 5.10 – Résultats des métriques : erreur quadratique (RMSE) et SNR moyen pour un nombre de TSP variant de 100 à 600.

Aussi, les régions de peau qui contribuent le plus pour la mesure rPPG ne sont plus favorisées.

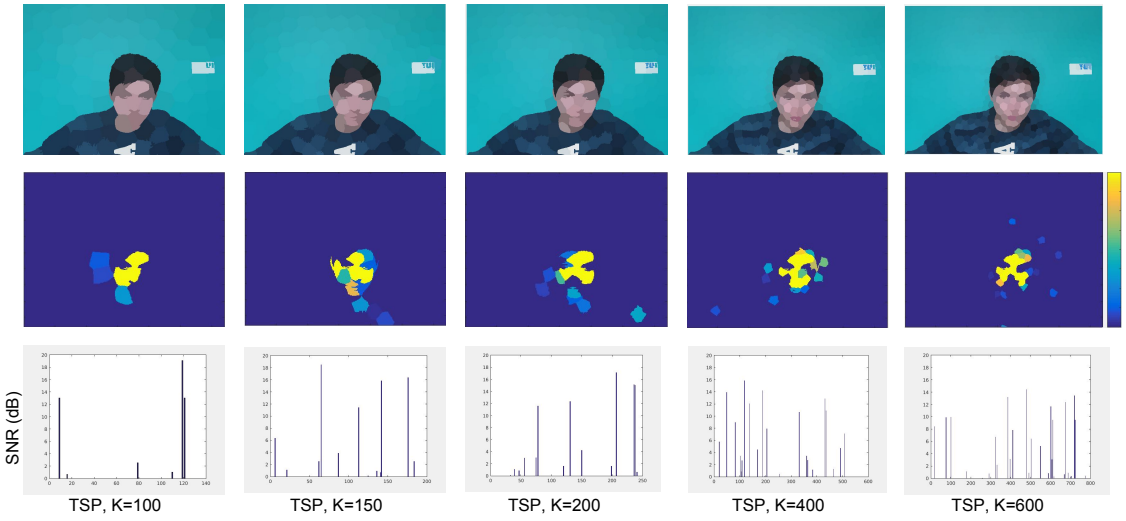


FIGURE 5.11 – Métrique *SNR* appliquée à la segmentation en TSP. La première ligne montre la segmentation en TSP tandis que la ligne 2 montre la correspondance avec la valeur de *SNR* et la ligne 3 montre les valeurs de *SNR* en fonction des différents TSP. Pour toutes les lignes, de gauche à droite, les résolutions en superpixels varient avec 100, 150, 200, 400 et 600 TSP.

Cette étude a permis de mettre en avant les bonnes performances obtenues avec une segmentation non supervisée de la région d'intérêt. Nous utilisons la méthode TSP proposée par Chang et al. [70] que nous avons jugé être un bon compromis entre précision de la segmentation et rapidité. Cependant, nous avons expérimentalement observé que cette méthode de segmentation a plusieurs limitations. Les perturbations dans la scène entrainent une remise à zéro de la valeur du superpixel temporel. Cela peut arriver en cas de rotation ou d'occultation du sujet dans la scène et ce même pour une très courte période de temps. Également, le calcul du flot optique entre les images peut remettre à zéro les superpixels temporels en cas de mouvement des pixels trop importants dans une zone. Comme nous travaillons avec une fréquence d'acquisition de 30 images par seconde sur cet ensemble de données, le flot optique nécessaire au calcul de TSP est sensible au mouvement trop rapide. La Figure 5.12 montre les erreurs de suivi des superpixels dues à ces raisons. Une

fois le superpixel remis à zéro, il faut de nouveau attendre 20 secondes de signal avant de pouvoir mesurer le signal rPPG.

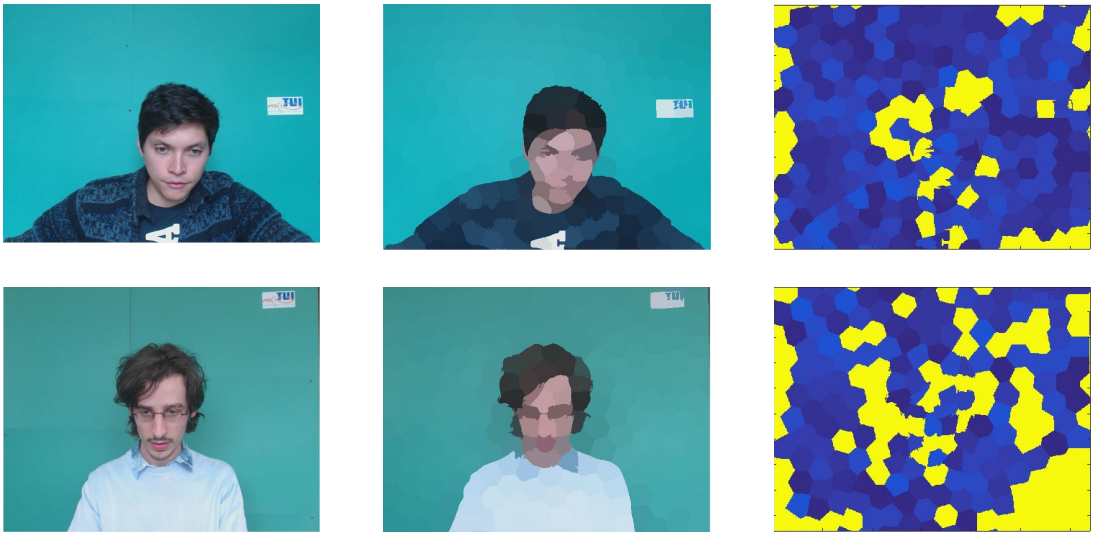


FIGURE 5.12 – Exemples d'erreurs dans le suivi temporel des superpixels de la méthode TSP. La première colonne montre les images d'entrée, la seconde la segmentation en TSP et la troisième les superpixels qui ont été remis à zéro (en jaune) dans la vidéo.

Les frontières des superpixels peuvent également varier légèrement entre les images. Ces variations génèrent un bruit haute fréquence et dans certains cas peuvent interférer avec la mesure du signal rPPG. Nous avons observé avec cet ensemble de données *dataset réaliste* que la méthode TSP est suffisamment efficace pour généralement éviter ce genre de phénomène et permet une très bonne estimation du rythme cardiaque.

En conclusion de cette étude, nous avons validé expérimentalement notre méthode de segmentation avec des résultats bien meilleurs pour notre approche que pour les méthodes de segmentation classiques implémentées : *face*, *crop* et *skin*. Nous avons implémenté et exploité cinq différentes méthodes de rPPG de l'état de l'art : *Green*, *Green – Red*, *PCA*, *CHROM* et *POS* afin de vérifier le bon fonctionnement de notre méthode en fonction des méthodes utilisées. Il s'avère que le choix des méthodes a son importance comme la méthode doit permettre une estimation de bonne qualité avec une quantité de pixels inférieure afin de bénéficier des avantages de cette méthode. Notre méthode a toujours montré des résultats supérieurs en terme de précision et permet d'améliorer la qualité du signal entre 15.1% et 59.4%, en se basant sur la métrique de qualité SNR, ce qui se traduit par une augmentation de la qualité de l'estimation du rythme cardiaque. Cependant, la méthode TSP est très intensive en calcul et elle ne permet pas de réaliser la mesure rPPG en temps réel. Nous présenterons donc dans le chapitre suivant une nouvelle méthode de segmentation en superpixels temporels nettement plus rapide.

5.2.3/ Validation de la métrique de qualité du signal basée sur les modèles de Markov cachés

La métrique de qualité *SNR* a permis de montrer que la segmentation implicite de peau par l'information rPPG était une méthode totalement valide et capable d'une grande précision dans la mesure du rythme cardiaque. Cependant, le calcul de la métrique *SNR* nécessite une transformation du signal $p^i(t)$ dans l'espace fréquentielle. Cette opération est bien optimisée dans tous les systèmes mais dont la multiplication, dans le cas de l'usage de superpixels, peut imposer un cout en calcul important. Afin de fournir une alternative à cette mesure, nous avons proposé d'utiliser une mesure de vraisemblance dans le domaine temporel du signal rPPG. Cette méthode permet de s'affranchir de l'analyse fréquentielle et permet d'identifier les zones de peau qui contribuent le plus pour le signal rPPG.

Afin d'évaluer cette méthode, il nous faut dans un premier temps déterminer le seuil que nous souhaitons utiliser afin de déterminer les superpixels dont les contributions sont à éliminer ainsi que réaliser l'entrainement du modèle de Markov pour déterminer les matrices d'émission et de transmission respectivement des observations et états cachés. Les matrices d'émission, notée Bm, et de transmission, notée Am, sont initialisées avec les valeurs suivantes :

$$\begin{aligned} Am &= \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, \\ Bm &= \begin{pmatrix} 0.6 & 0.1 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.6 & 0.1 \end{pmatrix}. \end{aligned} \tag{5.1}$$

Nous initialisons les matrices d'émissions et de transmissions en fonction de la représentation temporelle idéale du signal rPPG pour un modèle à deux états (*c.f.* Figure 4.7). Cela permet que le modèle converge plus rapidement pendant l'entrainement. Ainsi, on fixe la probabilité de passage d'un état à l'autre de 20% tandis que les probabilités d'émission des symboles sont fonction de la forme du signal mesuré. Comme les symboles les plus représentés avec un signal rPPG sont la croissance et la décroissance continues, on fixe les probabilités d'émissions à 60% pour une observation d'une croissance continue dans l'état 1 et 30% pour une décroissance continue dans l'état 1. Ces probabilités sont inversées dans l'état 2. Les probabilités de mesurer une variation concave ou convexe de la courbe sont nécessairement plus faibles et sont donc fixées à 10%.

Afin de dissocier les données utilisées pour l'entrainement des matrices de notre modèle de Markov, nous avons décidé d'entrainer notre modèle sur l'ensemble de données *dataset simple* et d'évaluer notre méthode sur l'ensemble de données plus complet et plus complexe *dataset réaliste*. Ainsi, après entrainement les matrices d'émission et de transmission ont les valeurs suivantes :

$$\begin{aligned} Am &= \begin{pmatrix} 0.8694 & 0.1306 \\ 0.1298 & 0.8702 \end{pmatrix}, \\ Bm &= \begin{pmatrix} 0.9192 & 0.0808 & 0 & 3.486e^{-5} \\ 0 & 0.0480 & 0.8176 & 0.1345 \end{pmatrix}. \end{aligned} \tag{5.2}$$

A l'issu de l'entrainement, on observe que les valeurs de nos matrices de transmissions et d'émissions ont évolué de façon significative. On constate une probabilité nulle pour l'observation d'une décroissance continue dans l'état 1 et pour l'observation d'une croissance continue dans l'état 2. En raison du filtrage appliqué sur les signaux $C(t)$ et de la très bonne qualité des

signaux pour l'ensemble de données *dataset simple*, aucun exemple de ce type d'émission n'a été observé dans l'entrainement du modèle.

Afin de déterminer la valeur du seuil x, nous avons appliqué notre méthode de segmentation sur l'ensemble de données *dataset simple*. Nous avons évalué la qualité des signaux rPPG *potentiels* avec la mesure de la métrique *SNR* et dans le même temps avec la mesure de vraisemblance BIC_λ pour le modèle de Markov dédié au signal rPPG. Le seuil x est défini comme le pourcentage de la valeur de l'amplitude de vraisemblance des valeurs de BIC_λ pour les signaux dont la mesure de *SNR* est positive, défini comme l'ensemble BIC_{snr}, par rapport à la valeur de vraisemblance maximale, telle que $\forall i \in K(t)$:

$$\begin{aligned} BIC_{snr} &= \{BIC_\lambda^i\}, SNR_i > 0, \\ x &= \frac{max(BIC_{snr}) - min(BIC_{snr})}{max(BIC_{snr})}. \end{aligned} \tag{5.3}$$

Avec toutes les vidéos de l'ensemble de données *dataset simple* moins une, car 25% plus courte que les autres, les valeurs varient entre 0 et 9% avec une valeur moyenne oscillant autour de 5% comme le montre la Figure 5.13.

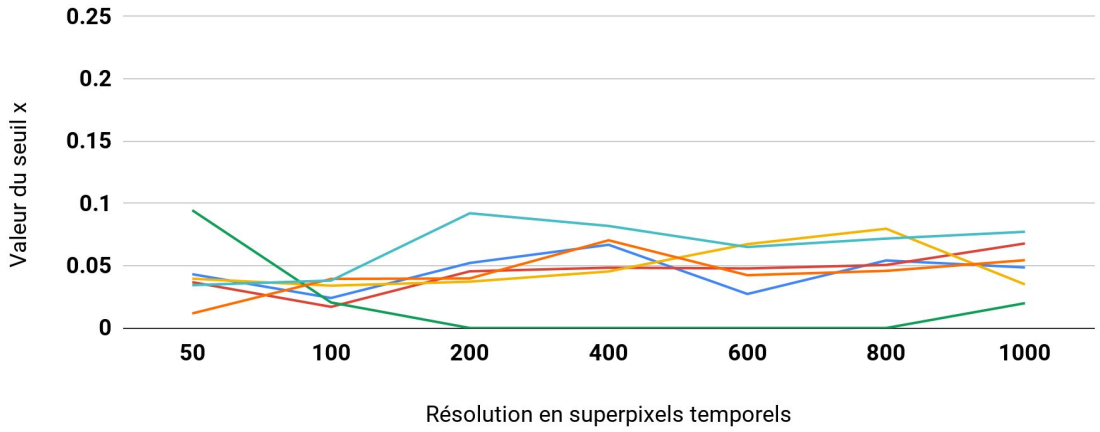


FIGURE 5.13 – Pour chacune des vidéos de l'ensemble de données *dataset simple* : variations des valeurs du seuil x en fonction de la résolution en superpixels temporels.

L'estimation du seuil a été faite pour les résolutions de superpixels temporels de 50, 100, 200, 400, 600, 800 et 1000 afin de vérifier l'impact du bruit de quantification avec la taille des superpixels diminuant. Comme le montre la Figure 5.13, la valeur du seuil reste stable avec les variations de résolutions et le signal ne semble pas suffisamment déformé par le bruit de quantification pour impacter la mesure de vraisemblance. En considérant la phase expérimentale et l'usage de l'ensemble de données *dataset réaliste*, nous avons décidé de considérer l'arrondi à la valeur supérieure de la valeur moyenne du seuil 4.45% soit $x = 5\%$. Cela permet de prendre en compte la complexité plus importante des données utilisées dans la phase de validation de la méthode.

Afin de valider cette nouvelle métrique de qualité du signal rPPG, nous utilisons le même pipeline de fonctions qu'avec la métrique *SNR* (voir Figure 4.3). En effet les signaux $p_i(t)$ sont évalués séparément pour les deux modèles de Markov et les métriques de vraisemblances BIC_{λ_s} et BIC_{λ_n}. Le tableau 5.4 montre les résultats des métriques de précision et d'erreur en comparaison avec la métrique de qualité *SNR*.

TABLE 5.4 – Résultat pour l'estimation du rythme cardiaque avec la métrique de qualité BIC pour la vraisemblance avec les modèles de cachés de Markov en comparaison avec la métrique de qualité SNR.

Métriques	SNR	BIC
Precision à 5 BPM	0.961	**0.971**
Precision à 2.5 BPM	0.890	**0.943**
RMSE	**2.39**	2.77
Correlation	**0.83**	0.82
MAE	1.80	**1.40**

La méthode de rPPG utilisée est la méthode *CHROM* et pour les métriques employées, la valeur MAE représente l'erreur moyenne de l'estimation du rythme cardiaque comme utilisés dans la représentation de Bland-Altman. Les résultats sont comparés pour une résolution en superpixels temporels de 150 TSP. Les valeurs en gras indiquent le meilleur résultat. Ces résultats permettent de valider l'efficacité de la métrique de qualité BIC. On peut noter que l'erreur moyenne est inférieure avec la méthode BIC tandis que l'erreur quadratique est supérieure. Cela signifie que le nombre d'erreurs est plus faible mais que les mauvaises estimations ont des valeurs plus éloignées de la vérité terrain avec la métrique de qualité BIC. Les résultats sont cependant très proches et l'estimation du rythme cardiaque se fait très bien dans les deux cas et sur l'ensemble de l'ensemble de données. Cette nouvelle métrique permet de réduire le coût de l'estimation de la qualité du signal mesuré. De plus, elle améliore l'estimation du rythme cardiaque et est plus robuste aux perturbations. Enfin, comme la mesure est faite dans le domaine temporel, un phénomène périodique dans l'arrière plan de la scène ne constitue pas une source d'information correspondante au modèle utilisé, alors que ce serait le cas avec l'utilisation de la métrique SNR.

La Figure 5.14 permet d'observer la résilience au mouvement des deux méthodes. Tandis que l'arrière-plan est généralement très bien éliminé par la méthode avec la métrique de qualité *SNR* et l'écart entre les poids w_i, Les mouvements et déformation des superpixels ont tendance à faire émerger une fausse information de signal induite par les variations des frontières des superpixels visibles dans cette figure. Au contraire, la métrique de qualité BIC permet d'être plus résilient à ces perturbations mais la distinction entre les superpixels de peau et ceux de l'arrière-plan est moins significative.

5.3/ DISCUSSIONS

Nous avons démontré dans la sous-section précédente l'efficacité de la méthode proposée en termes de mesure par rapport aux autres approches de l'état de l'art. Cependant, dans l'objectif de proposer un capteur intégrant notre approche rPPG, la capacité à estimer le rythme cardiaque en respectant les contraintes de temps des applications visées est fondamentale. Dans notre approche, la segmentation représente l'étape du traitement rPPG la plus critique car la plus consommatrice en temps de calcul. En effet, par rapport aux autres étapes, la complexité algorithmique est plus importante mais de plus le nombre de points traités est beaucoup plus grand. Concernant le nombre de points traités pour les phases de traitement du signal, le filtrage ainsi que la transformée de Fourrier ont des complexités algorithmiques respectives en $O(N)$ et $O(N \cdot \log(N))$. En comparaison, avec la méthode TSP, la segmentation de l'image basée sur la méthode *SLIC* à une complexité en $O(N)$, avec ici N le nombre de pixels dans l'image, tandis que la propagation temporelle qui nécessite le calcul du flot optique entre les images a une complexité en $O(N^2)$ sans considérer d'optimisa-

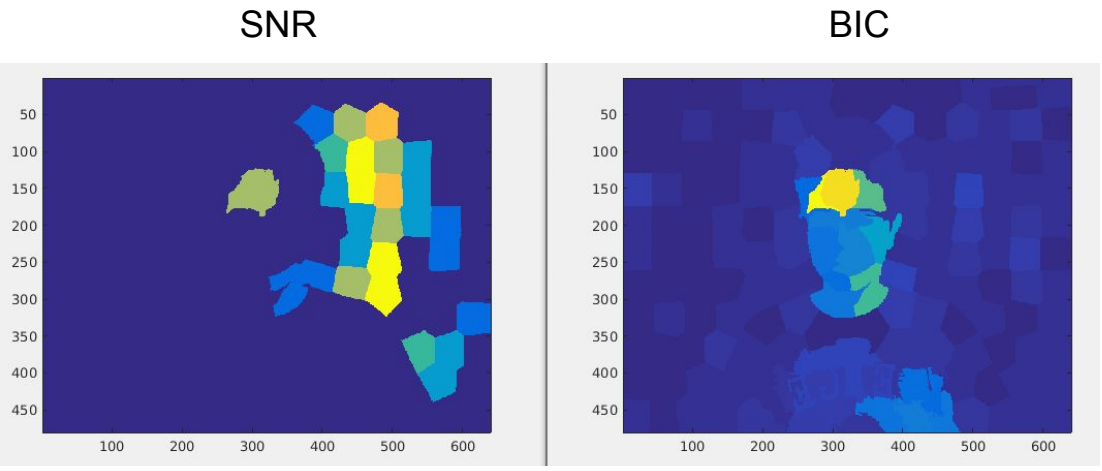


FIGURE 5.14 – Segmentation en TSP pour la métrique de qualité *SNR* à gauche et la métrique de qualité BIC à droite.

tion. Le nombre d'opérations réalisées durant la segmentation est significativement plus important que dans les autres étapes algorithmiques. La mesure de rPPG est couramment réalisée sur une fenêtre glissante de 20 secondes. Dans cette configuration et pour une cadence de 30 images par seconde, le calcul du filtrage et de la transformée de Fourrier est effectuée sur 600 points ($n = 600$) pour chaque mesure. A titre de comparaison, lors de la phase de segmentation, le nombre N de points traités qui correspond donc au nombre de pixels dans l'image, est bien supérieur. En effet, pour une simple résolution VGA (640x480 pixels), le nombre de points à traiter est égal à 307200 pixels. Ainsi, l'ordre de grandeur des temps de calcul, sous l'environnement Matlab, est complètement différents avec en moyenne un facteur 1000 entre la partie traitement du signal, pour un signal unique, et la partie traitement de l'image qui peut prendre plusieurs dizaines de secondes. Bien évidemment les implémentations sous Matlab ne sont pas optimisées et les temps de calcul sont purement indicatifs, les opérations étant beaucoup plus lentes qu'avec une implémentation en C++.

Cette étude démontre que la phase de segmentation de la zone d'intérêt doit être la plus précise possible. En effet, une dégradation de la qualité de la segmentation impacte de manière significative la qualité de la mesure et ceci quelque soit l'approche rPPG utilisée. La segmentation en superpixels est certes précise mais contrairement à des approches supervisées classiques, ne peut être réalisée en temps réel sur une plateforme matérielle standard et pour un flux vidéo d'une caméra standard (i.e. de résolution supérieure au 1Mpixels et 30 fps), nous proposerons donc dans le chapitre suivant une méthode de segmentation non supervisée beaucoup plus rapide que celle utilisée dans ce chapitre afin de répondre à cette problématique.

III

Optimisations et implémentation pour une mesure rPPG temps réel

6

Segmentation en superpixels à complexité réduite pour la rPPG

Sommaire

6.1 Segmentation par superpixels 78
6.2 Présentation de IBIS 80
6.2.1 Identification implicite des frontières 81
6.2.2 Analyse de la mise à jour des seeds en fonction des itérations 83
6.2.3 Optimisation 84
6.2.4 Évaluation de la complexité algorithmique 87
6.3 Propagation temporelle de la segmentation et post-traitement pour la mesure rPPG 92
6.4 Validation 93
6.4.1 Plateforme PC à processeur multi-cœurs 94
6.4.2 Evaluation de la qualité de segmentation de IBIS 96
6.4.3 Mesure rPPG 99
6.5 Discussions 101

Dans ce chapitre, nous proposons une nouvelle méthode pour la mesure photoplethysmographique sans contact en utilisant une segmentation en superpixels plus efficace. La méthode de segmentation proposée identifie de manière implicite les frontières entre les superpixels. Ainsi, la segmentation en superpixels est effectuée en utilisant uniquement une fraction des pixels de l'image d'entrée, ce qui réduit grandement la complexité et par conséquent le temps de calcul. Dans un premier temps, une rapide introduction des méthodes de segmentation en superpixels sera réalisée afin d'identifier les critères d'optimisation qui ont été retenus. Dans un second temps, notre méthode de segmentation appelée IBIS sera détaillée. Enfin, la méthode sera adaptée et appliquée au cas de la mesure rPPG.

Les résultats obtenus lors des expérimentations montrent que la qualité dans la phase de segmentation est comparable à celle des méthodes de l'état de l'art tandis que le temps de calcul est divisé par un facteur 8 sans parallélisation du traitement pour des images de faible résolution (e.g. 320x240 pixels) telles que fournies habituellement dans les ensembles de données publics. L'efficacité de la méthode IBIS nous permet alors de mesurer le signal rPPG et de réaliser l'estimation du rythme cardiaque en temps réel avec des images de résolution 640x480 pixels, à 25 ou 30 images par seconde en utilisant un seul des coeurs de calcul de l'unité de traitement utilisé. Nous sommes ainsi en capacité de réaliser un traitement en temps réel avec une précision de 93% avec une erreur maximale de 2.5 battements par minute.

6.1/ Segmentation par superpixels

La segmentation de l'image par superpixels est une méthode de pré-traitement très répandue dans le domaine du traitement de l'image et de la vidéo. Cependant, la plupart des méthodes souffrent d'un cout en calcul très important qui les rend inutilisables pour des applications à fortes contraintes de temps.

Les superpixels sont construits de façon à regrouper les pixels de l'image d'une façon cohérente. Ainsi, ils permettent de résumer l'image en régions définies par des frontières qui adhèrent aux contours présents dans l'image. Ils sont largement utilisés pour une grande quantité d'applications : suivi temporel d'un objet [115], segmentation de l'image [61], décomposition d'un flux vidéo [70], ou enfin dans des applications de mesure biologique [109]. La complexité des méthodes de segmentation par superpixels est devenue l'un des principaux critères de développement [65, 113]. Dans les faits, un algorithme de superpixels simple et rapide sera souvent privilégié car il sera utile dans un grand nombre d'applications en permettant son utilisation à un grand nombre de nouveaux domaines d'applications [50, 130].

Les superpixels doivent être visuellement cohérents, cela se traduit par un regroupement des pixels qui sont proches dans le domaine spatial ainsi que similaires dans le domaine chromatique. De plus, les méthodes doivent être aussi efficaces que possible comme la segmentation est souvent une étape de prétraitement. Également, les méthodes sont de préférence algorithmiquement simples afin de ne pas avoir de multiples et complexes paramètres à définir par l'utilisateur. Les méthodes de superpixels populaires et récentes comme SLIC [58] ou SEEDS [65] ont une complexité algorithmique en $O(N)$, avec N la résolution spatiale de l'image en pixels. Afin de diminuer les temps de calcul nécessaires, la parallélisation des tâches est devenue une alternative réaliste avec le développement des capacités de calculs concurrents sur les plateformes matérielles telles que les processeurs graphiques GPU ou les processeurs CPU à cœurs multiples. Le processus de segmentation en superpixels a ainsi été adapté dans une version complétement parallèle de la méthode SLIC, nommée gSLIC [26] ou encore en séparant le traitement des opérations de quantification des énergies spatiales et chromatiques dans la méthode USEQ [113] sur une plateforme CPU et en exploitant deux cœurs de façon simultanée.

Plusieurs méthodes de segmentation ont été proposées au cours des dernières années. Ainsi, on peut dissocier deux grandes catégories de méthodes. Celles à base de graphes modélisent les relations entre pixels voisins. Dans ce cas, les arrêtes des graphes quantifient les similarités. La méthode *normalized cuts* proposée par Shi et al. en 2000 [17] modélise ainsi les pixels par des nœuds et les liens entre pixels voisins comme des arrêtes. La segmentation est effectuée en minimisant une fonction de cout dans un processus itératif sur le graphe. La complexité de cette méthode est réduite en 2005 par Mori et al. [30] avec leur méthode *guiding model* et permet ainsi une segmentation de qualité égale tout en améliorant les performances de la méthode en termes de temps de calcul. Ces méthodes ne sont pas considérées dans notre approche. En effet, leurs complexités sont généralement trop importantes et sont difficiles à optimiser. De plus, la seconde grande catégorie des méthodes de segmentation en superpixels permettent de fournir une segmentation plus régulière de l'image en sur-segmentant les objets dans la scène. Ainsi, toute l'image est décomposée de façon homogène en zones contiguës. Ces méthodes utilisent une descente de gradient afin de minimiser une fonction de l'énergie dans l'image (*c.f.* équation4.1). L'image est initialement découpée avec une grille fixe qui est mise à jour dans un processus itératif jusqu'à convergence. Les méthodes comme *mean shift* [21], *SLIC* [58] ou *USEQ* [113] segmentent l'image en zones spatialement régulières, (*c.f.* Figure 6.1), en optimisant la distance euclidienne entre les pixels de l'image et les superpixels initialisés avec la grille fixe.

FIGURE 6.1 – Exemple de segmentation par superpixels avec la méthode IBIS avec à gauche la segmentation initiale pour initialiser la méthode et à droite la segmentation finale.

On peut voir que les frontières entre les superpixels suivent les contours présents dans l'image comme les mains ou le visage par exemple. Bien évidemment, avec une quantité de superpixels très faible comme dans cet exemple, la segmentation est en générale de mauvaise qualité et il faut accroitre la quantité de superpixels afin d'affiner le niveau de détails de la segmentation. Un autre aspect de la complexité algorithmique des méthodes est directement lié au nombre de superpixels. Ainsi, plus le nombre de superpixels fixé par l'utilisateur est élevé plus le nombre d'opérations de calcul des distances entre les pixels de l'image et l'ensemble des superpixels augmentent. Afin de s'affranchir de cette contrainte qui peut rendre la méthode lente avec un nombre élevé de superpixels (typiquement plusieurs centaines ou milliers), les méthodes comme *SLIC* ou *USEQ* définissent un voisinage dans lequel les pixels et superpixels sont évalués. Ainsi, une région de pixels de l'image n'est jamais évaluée par l'ensemble des superpixels possibles mais uniquement par ceux dont la distance spatiale est inférieure à un seuil. Cela permet de réduire considérablement l'impact de la quantité de superpixels spécifiés par l'utilisateur tout en maintenant la qualité de la segmentation. En effet, comme la fonction de l'énergie dans l'image est définie comme la somme pondérée des énergies spatiale E_s et chromatique E_c, l'accroissement de la distance entre les pixels et les superpixels accroit l'énergie entre eux et rend impossible leurs correspondances.

La méthode *SLIC* représente une des contributions majeures dans son domaine durant ces dernières années [58, 62]. Cette méthode a introduit le processus de segmentation itératif en se basant sur l'amélioration de la segmentation à partir d'une grille fixe. Cependant, même si la complexité peut être limitée en travaillant sur une région de l'image, le temps de calcul nécessaire afin de réaliser la segmentation en superpixels est toujours trop important et cela même en considérant les avancées technologiques et la puissance de calcul disponible. En effet, même en considérant des images de très faible résolution (typiquement 240x320) comme celles de l'ensemble de données public de Berkeley [82], il n'est pas possible d'effectuer la segmentation en moins de 70 à 80 ms avec un processeur récent (dans notre cas : i7 4790@3.6Ghz). Aussi cette méthode est toujours trop coûteuse et inadaptée pour des étapes de prétraitement dans le cadre d'applications temps réel. La méthode *SEEDS* réalise la segmentation en superpixels plus rapidement que *SLIC* en

utilisant un processus d'affinage des frontières qui se concentrent sur les modifications potentielles spécifiquement au niveau des frontières et permet donc de réduire le coût en opération. Cependant, son temps de calcul augmente avec le nombre de superpixels, indépendemment des optimisations algorithmiques, et limite donc son utilisation pour notre usage. Qui plus est, il a été montré que les frontières des superpixels suivent moins bien les contours dans l'image que les autres méthodes [62], spécifiquement pour les niveaux de détails fins.

6.2/ Présentation de IBIS

Cette section introduit une nouvelle méthode de segmentation par superpixels nommée *IBIS* pour *Iterative Boundaries implicit Identification for superpixels Segmentation* ou segmentation en superpixels par identification implicite itérative des frontières. Cette méthode permet de limiter significativement le cout en calcul du processus de segmentation en identifiant de manière implicite les frontières des superpixels, ce qui permet de n'utiliser qu'une portion des pixels de l'image. La Figure 6.2 permet de visualiser la proportion de l'image utilisée (en jaune) afin de réaliser la segmentation en superpixels.

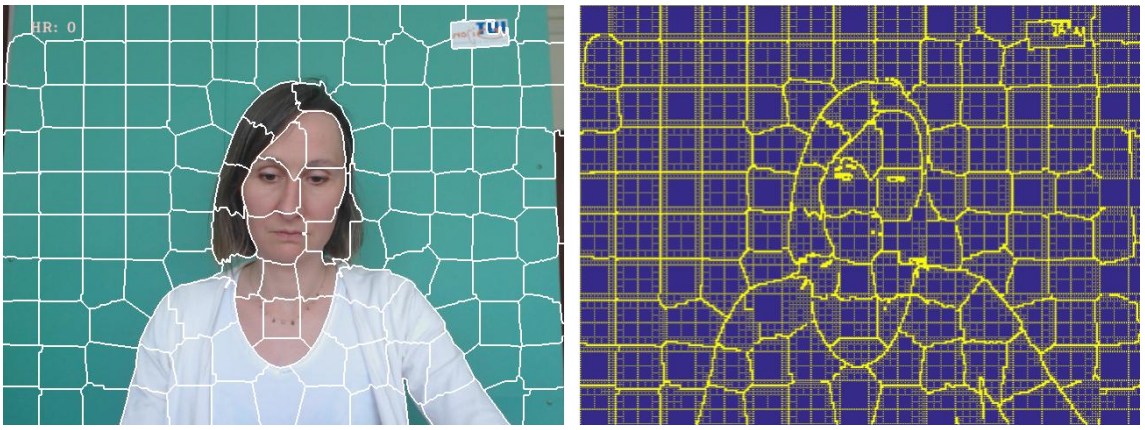


Figure 6.2 – Exemple de segmentation par superpixels avec la méthode IBIS avec à gauche la segmentation finale superposée à l'image d'entrée et à droite l'identification en jaune des pixels de l'image utilisés pour calculer la segmentation.

Afin de mieux visualiser la diminution de complexité, on représente ainsi l'ensemble des pixels considérées : ceux dont la distance est calculée avec les différents superpixels. En raison de la diminution significative du nombre de pixels considérés et du nombre d'opérations réalisées, cette segmentation peut être effectuée, sur une plateforme PC standard, jusqu'à huit fois plus rapidement que les méthodes de l'état de l'art avec des images en entrée de résolution (320x240). De plus, il est possible grâce à cette figure, de visualiser la localisation des opérations (zones avec une forte concentration de pixels en jaune) vers les frontières. Cette convergence est rendue possible par un processus itératif visant à identifier les frontières entre les superpixels.

Introduite par Achanta et al. [58] en 2012, la méthode *SLIC* est une des contributions majeures parmi les méthodes de segmentation par superpixels. C'est encore aujourd'hui l'une des méthodes les plus employées et citées [131,132] compte tenu de ses performances. Le groupement des pixels est réalisé en agglomérant les pixels à des superpixels définis suivant une grille fixe durant l'initialisation. Les pixels sont attribués au superpixel pour lequel la distance chromatique et spatiale est minimale. Les superpixels sont représentés par une notation vectorielle nommée *seeds* dans la littérature. Ce vecteur comporte les informations relatives à la position du barycentre du superpixel

suivant les deux coordonnées x et y ainsi que la valeur chromatique moyenne. L'espace chromatique employée par Achanta et al. est le *CIE l*a*b** car les auteurs montrent que les variations de couleur dans cette espace sont perceptuellement plus homogènes pour de faibles écarts de couleur. Le processus est ainsi répété 10 fois afin de garantir une bonne qualité de la segmentation tandis que les *seeds* sont mises à jour entre chaque itération en tant que vecteurs à 5 dimensions $\{x, y, l^*, a^*, b^*\}$, avec x et y les coordonnées spatiales dans l'image et l^*, a^* et b^* les composantes chromatiques. Chaque pixel est représenté de la même façon par un vecteur à 5 dimensions. Nous nous sommes inspirés de cette méthode en raison de la qualité de la segmentation obtenue ainsi que la simplicité de son approche à l'origine de la popularité de cette méthode. Ainsi, la procédure d'assignation des pixels aux différents superpixels est identique entre les méthodes *IBIS* et *SLIC*. Elle est réalisée par le calcul de la distance euclidienne chromatique D_{lab} dans l'espace de couleur *CIE l*a*b** et par le calcul de la distance euclidienne spatiale D_{xy}. Les pixels sont associés avec le superpixel le plus proche dont la distance est définie comme la somme des distances chromatique et spatiale. La distance totale D entre le i-ième pixel et la k-ième *seeds* est définie comme $D = D_{lab} + \theta \cdot D_{xy}$ avec :

$$D_{lab} = \|(l, a, b)_i - (l, a, b)_k\|, \tag{6.1}$$

$$D_{xy} = \|(x, y)_i - (x, y)_k\|. \tag{6.2}$$

Avec $\|\cdot\|$ la norme euclidienne et θ le facteur de compacité défini par $\theta = 1/c^2$ avec c un paramètre fixé par l'utilisateur. Cette opération correspond à la minimisation des énergies E_c et E_s définies en section 4.1.2 et est la plus couteuse en calculs. En effet, en considérant chacune des itérations, tous les pixels doivent être évalués en considérant leurs affectations possibles à tous les superpixels de l'image ou dans une localité en fonction du niveau d'optimisation de l'implémentation. Afin de limiter le nombre d'opérations nécessaires au calcul des distances entre les pixels et les superpixels, on limite les superpixels potentiels à ceux compris dans une norme autour d'un même pixel. Cette distance est définie en fonction des paramètres d'entrée (nombre de superpixels et compacité). En complément de cette limitation de distance, nous proposons de limiter, comme énoncé précédemment, le nombre de pixels considérés.

6.2.1/ Identification implicite des frontières

En se basant sur les résultats générés par les méthodes de l'état de l'art telles que *SLIC* [58], *SEEDS* [65] ou *USEQ* [113], les superpixels et le procéssus de segmentation associés doivent respecter un certain nombre de règles. Ainsi, les superpixels doivent être réguliers, contigus et doivent avoir une taille minimale. Le résultat du processus de segmentation est une assignation pour chaque pixel à un superpixel représenté par son numéro identifiant (*label*). Nous proposons de définir un bloc carré afin d'identifier les frontières entre les superpixels.

L'image est dans un premier temps divisée en blocs au moyen d'une grille régulière définie manuellement et représentée par un bloc appliqué sur toute l'image. Un bloc est ainsi défini comme cohérent si et seulement si une sélection de pixels à l'intérieur de ce bloc, nommée *pixels utiles*, sont tous associées au même superpixel. Dans ce cas, tous les pixels du bloc sont associés à ce superpixels. Autrement, dès que les *pixels utiles* sont assignés à des blocs différents, le processus est interrompu et le bloc est subdivisé en 4 blocs de taille inférieure. Une fois tous les blocs de la première itération calculées, le processus est répété comme représenté dans la Figure 6.3. Les itérations sont représentées de gauche à droite. Les zones qui sont cohérentes, et ne seront donc pas subdivisées, sont colorées en gris. Ainsi, on peut observer qu'au fur et à mesure des itérations, les frontières sont plus finement identifiées.

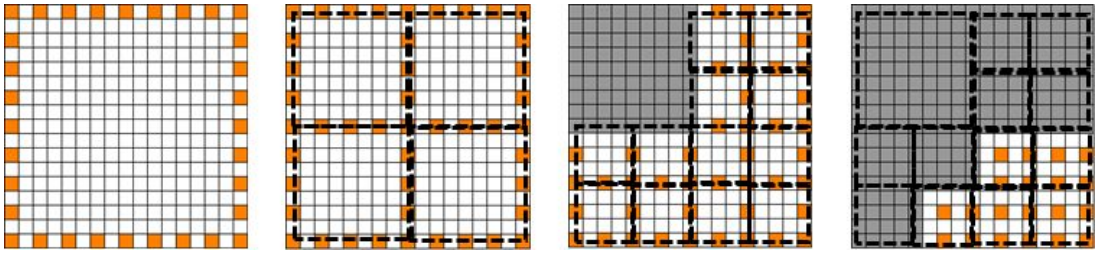


FIGURE 6.3 – Exemple d'itérations dans le déroulement de la méthode IBIS avec un bloc unique. Les zones cohérentes sont éliminées du processus afin de réduire la quantité de pixels utiles à l'identification des frontières entre les superpixels.

Avec cette stratégie, il est possible de limiter drastiquement le nombre de pixels qui sont évalués dans le processus d'agglomération générant les superpixels. En fait, seuls les pixels qui sont situés sur la bordure d'un bloc sont utilisés afin de déterminer si un bloc est cohérent ou non. En faisant cela, nous déterminons de façon itérative les blocs qui contiennent les contours de l'image et les frontières entre les superpixels. La procédure est de plus optimisée en ne considérant qu'un pixel sur deux sur la bordure du bloc. Ces pixels forment la sélection des *pixels utiles* et représentent seulement une fraction des pixels de l'image. La Figure 6.4 permet de visualiser le processus d'identification des frontières en faisant converger les blocs incohérents au niveau des frontières entre les superpixels.

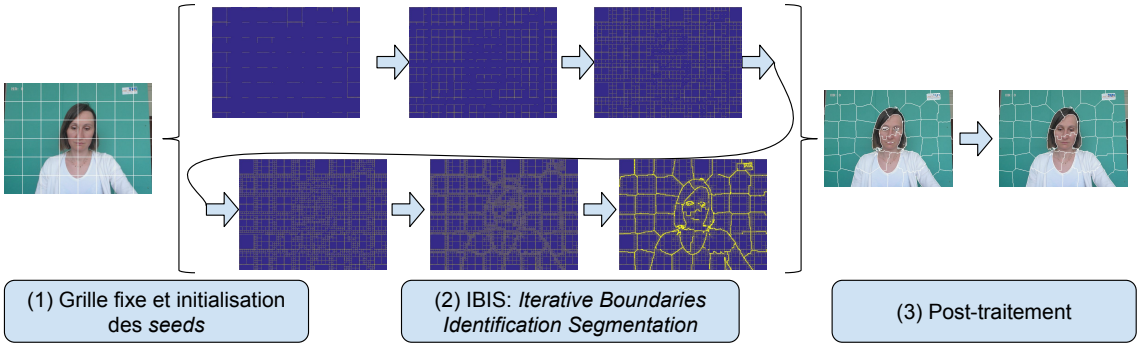


FIGURE 6.4 – Schéma fonctionnel de la méthode IBIS. (1) Les *seeds* sont initialisées avec une décomposition en grille fixe. (2) La méthode IBIS est appliquée et les itérations permettent d'affiner l'identification des frontières et (3) le post-traitement permet de s'assurer de la contiguïté et de la taille minimale des superpixels.

Les blocs sont définis en fonction du pas que l'on souhaite utiliser. Ainsi, avec un pas de 2 (*i.e.* 1 pixel sur 2 utilisés au niveau des frontières du bloc), le plus petit bloc est de taille 3x3 puis le bloc de taille supérieure sera de taille 5x5, etc. Dans ce chapitre, le pas de la méthode sera fixé à 2. Plus généralement, la taille des blocs sur une seule dimension spatiale est définie par τ_i pour la i-ième itération et est définie par :

$$\tau_i = 2^{\alpha - i + 1} + 1, \tag{6.3}$$

avec $\alpha + 1$ le nombre d'itération et :

$$\alpha = \lfloor \log_2(\sqrt{N/K}) - 1 \rfloor, \tag{6.4}$$

avec $\lfloor \cdot \rfloor$ l'opérateur entier inférieure, $i \in [0, \alpha]$, $\sqrt{N/K}$ la taille moyenne d'un superpixel et N le nombre de pixels dans l'image et K le nombre de superpixels. Ainsi, la taille des blocs τ_0 sera toujours inférieure à la taille moyenne des superpixels afin de s'affranchir du cas où un superpixel entier serait compris dans le bloc. Dans cette situation, avec la définition des blocs appliquée, le superpixel sera trop petit et ne doit donc pas être considéré. Les blocs de plus haut niveau, de taille τ_0, sont appliqués de façon à ne pas partager de pixels au niveau de leurs frontières. Ainsi les blocs sont définis les uns contre les autres comme montrés dans la Figure 6.5. Cela permet de séparer les opérations par bloc de niveau 0 et ainsi permettre la parallélisation des opérations sur une architecture comportant plusieurs unités de calcul.

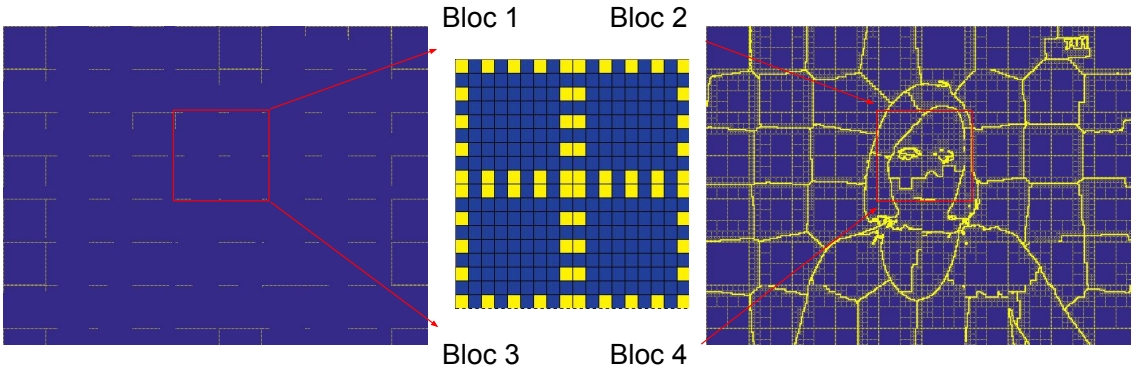


FIGURE 6.5 – Les blocs de plus haut niveau de taille τ_0 ne se chevauchent pas, ainsi ils ne partagent pas de pixels en commun, cela permet une meilleure optimisation de la répartition des opérations en prévision d'une implémentation sur une plateforme avec plusieurs cœurs de calculs.

Afin de conserver le meilleur niveau de détails possible dans l'identification des frontières entre superpixels, une fois la dernière itération avec la taille de bloc $\tau_\alpha = 3$ atteinte, les blocs qui ne sont pas homogènes ne sont plus subdivisées, aussi les pixels restants sont attribués individuellement au superpixel le plus proche défini dans une localité. De cette façon, la qualité de la segmentation est préservée et le nombre de pixels évalués individuellement est faible au regard du nombre total de pixels de l'image.

6.2.2/ ANALYSE DE LA MISE À JOUR DES SEEDS EN FONCTION DES ITÉRATIONS

Dans la méthode *SLIC* initiale, les *seeds* sont mises à jour afin de permettre un affinement de la valeur moyenne des superpixels et ainsi accroitre la qualité de la segmentation avec une augmentation de l'adhérence des frontières des superpixels aux contours dans l'image. Au fur et à mesure que les superpixels agglomèrent de plus en plus de pixels chromatiquement et spatialement homogènes dans le plan image, la segmentation s'améliore. Dans notre approche, les *seeds* sont aussi mises à jour entre chaque itération, comme la valeur moyenne de tous les pixels contenus dans les blocs homogènes en fonction de leurs associations avec les différents superpixels. Ainsi, si un superpixel n'a été associé avec aucun des pixels de l'image pour l'itération courante, sa valeur moyenne, notée *seed*, reste inchangée.

La Figure 6.6 montre l'évolution de l'erreur absolue entre les valeurs moyennes des superpixels à mesure des itérations et leurs valeurs une fois le processus de segmentation terminé. A mesure que le nombre de superpixels augmente, la quantité de pixels par superpixels décroit et par conséquent, les variations de valeurs de la *seed* associée seront plus faibles. Dans le cas de la méthode *SLIC*, le nombre d'itérations est défini à 10 indépendamment du nombre de superpixels spécifié par l'utilisateur. Par défaut, et pour garantir une qualité suffisante de la segmentation dans l'image,

le processus de segmentation sera effectué 10 fois alors même que les valeurs de *seed* n'évoluent quasiment pas avec une faible densité de pixels par superpixel.

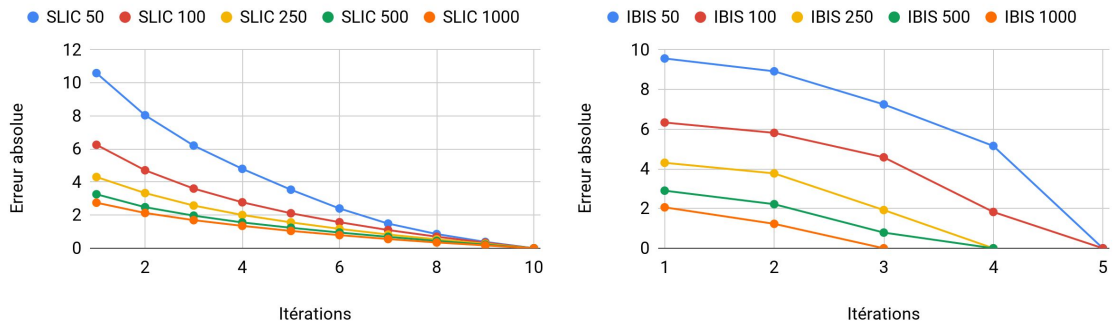


FIGURE 6.6 – Évolution de la valeur de la différence absolue entre les valeurs des *seeds* aux différentes itérations et la valeur finale en fonction du nombre de superpixels défini par l'utilisateur pour les méthodes *SLIC* à gauche et *IBIS* à droite.

Pour la méthode *IBIS*, le nombre d'itérations est défini par : $\alpha + 1$, qui est fonction du nombre de superpixels et de la taille de l'image (voir équation 6.3). Ainsi, notre méthode permet d'introduire une définition automatique du nombre d'itérations en fonction de la densité de pixels pour chaque superpixel. La différence de forme pour les courbes des deux méthodes est due à la façon dont les *seeds* sont mises à jour dans les deux méthodes. En effet, dans le cas de la méthode *SLIC*, tous les pixels de l'image sont associés au superpixel le plus proche. Au contraire de *IBIS*, où à mesure des itérations et avec la décroissance de la taille des blocs, un nombre croissant de blocs seront cohérents comme la probabilité de rencontrer une frontière sera plus faible et ainsi un plus grand nombre de pixels seront associés aux différents superpixels.

L'étape de post-traitement est identique à celle de *SLIC*. Un bloc est appliqué sur toute l'image pour assurer que la contiguïté et que la taille minimale des superpixels soient respectées.

6.2.3/ OPTIMISATION

Nous présenterons dans les sections 6.4.2 et 6.4.3 une évaluation de *IBIS* au regard de la qualité de la segmentation et au regard de l'estimation du signal rPPG. Nous présentons ici une seconde version de IBIS réduisant encore le nombre de pixels traités dans l'image. En effet, le développement de la première version de notre algorithme a été réalisé en considérant le degré de parallélisation potentielle d'une plateforme CPU à plusieurs cœurs de calcul. Par conséquent, dans cette version optimisée nous avons modifié la logique de notre algorithme afin de réduire encore le nombre de pixels traités. Cette nouvelle version est par conséquent plus rapide quelque soit le type de plateforme matérielle envisagée.

Notre objectif étant de réaliser une intégration sur une plateforme embarquée, nous avons proposé une version plus optimisée compatible avec une intégration GPU. La première optimisation fut de laisser les blocs se chevaucher. En effet, la logique de parallélisation apportée par les plateformes massivement parallèles GPU propose une philosophie différente dans l'assignation des tâches et cœurs de calcul. Si gérer des blocs indépendants permettait de simplifier le travail d'intégration dans le cas d'une implémentation sur plateforme CPU et également de laisser la capacité d'interrompre les traitements en cours de route, il est nécessaire de considérer les tâches et traitements au niveau pixel pour permettre une intégration sur des plateformes GPU. Pour ce faire, nous avons défini des vecteurs de coordonnées où appliquer les blocs de différentes tailles. Ainsi, pour cha-

cune des itérations, un bloc représente une tâche indépendante. De plus, les blocs sont alignés de sorte que le premier est appliqué dans l'image (dans le coin supérieur gauche) et a son origine en dehors de l'image, cela permet d'éliminer les pixels qui sont aux frontières de l'image.

La Figure 6.7 permet de visualiser la nouvelle logique appliquée dans la répartition des blocs sur l'image. Les blocs ne sont plus indépendants, ce qui permet d'éliminer de nombreux pixels puisqu'il n'est plus nécessaire de calculer deux colonnes et lignes de pixels adjacents pour deux blocs différents de taille τ_0. De plus, les pixels calculés par chacun des blocs se limitent ainsi aux bordures droite et inférieure moins un pixel pour chacune. Comme illustré par l'image droite de la Figure 6.7, la répartition des pixels par bloc est ainsi plus limitée, avec par exemple uniquement 7 pixels qui seront traités par la tâche chargée de traiter le bloc 4 comme montré par l'image à droite dans la Figure 6.7.

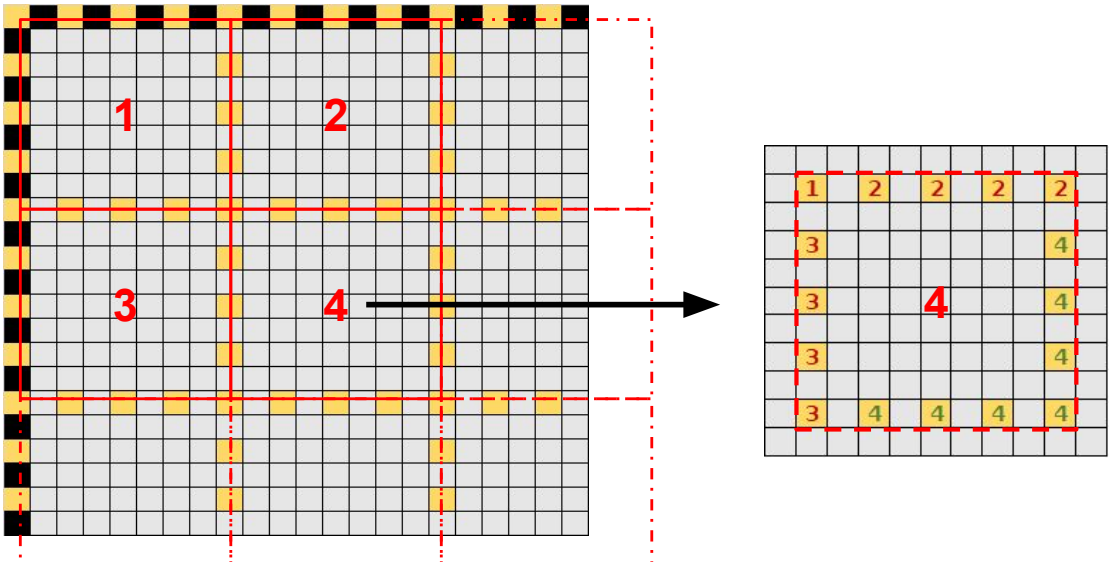


FIGURE 6.7 – Exemple dans l'application des blocs de plus haut niveau avec $\alpha = 2$, les blocs se chevauchent et les pixels calculés par chacun se limite à la bordure droite inférieure. Ainsi le premier bloc a son origine (en haut à gauche) en coordonnées $(-1; -1)$ dans l'image.

Les étapes suivantes de la méthode optimisée restent inchangées par rapport à celle d'origine. Les pixels d'un même bloc sont attribués au même superpixel si le bloc est cohérent. Sinon, le bloc est subdivisé en 4 et le processus est répété jusqu'au bloc de taille τ_α.

Les gains en performance entre les versions de IBIS sont présentés dans le chapitre 7. Cependant, l'économie en nombre de pixels traités dans l'image est résumé Figure 6.8. L'estimation du nombre de pixels utilisés dans l'image est moyennée sur l'ensemble de données de test *BDS500* présenté en 7.2.1 constitué de 500 images très variées afin de fournir un échantillon de test adapté. On peut voir que le nombre de pixels utilisés dans l'image augmente de 45% à 80% dans le cas de l'implémentation initiale tandis qu'on observe une augmentation de 35% à 65% dans le cas de la version optimisée. Nous avons également représenté le gain en nombre de pixels utiles dans l'image défini par la différence entre les deux versions. Ainsi, on peut observer un gain croissant de 11% à 15% en fonction de la résolution en superpixels dans l'image. Ces variations sont facilement expliquées puisque le nombre de pixels utilisés afin d'identifier les frontières entre superpixels augmente à mesure que les superpixels sont de plus en plus petit et que la densité en frontières dans l'image augmente en conséquence.

Cette modification se répercute sur la taille des blocs définies par $\tau_i = 2^{\alpha-i+1} + 1$. En effet comme

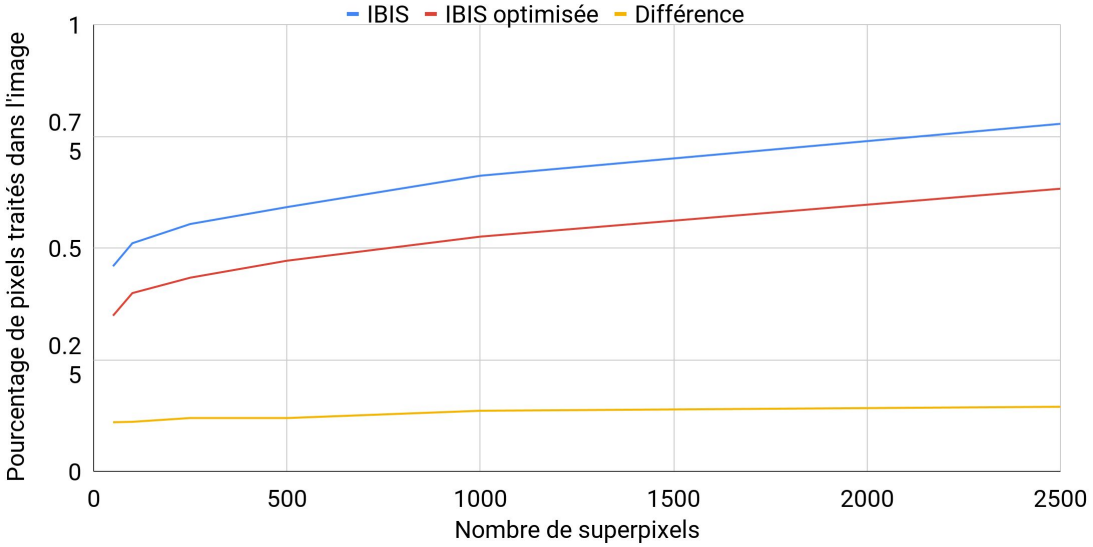


FIGURE 6.8 – Pourcentage des pixels qui sont nécessaires à la segmentation en superpixels avec la méthode *IBIS* en fonction de la résolution en superpixels pour l'implémentation initiale (en bleu), l'implémentation optimisée (en rouge) et la différence des deux (jaune).

les blocs ont la possibilité de se chevaucher, il est nécessaire de redéfinir la valeur de α avec la contrainte d'avoir la plus grande taille τ_0 de bloc inférieure à la taille moyenne des superpixels dans l'image :

$$\begin{aligned} 2^{\alpha+1} + 1 &> \sqrt{N/K} \\ 2^{\alpha+1} &> \sqrt{N/K} - 1 \\ \alpha + 1 &> \log_2(\sqrt{N/K} - 1) \\ \alpha &> \log_2(\sqrt{N/K} - 1) - 1. \end{aligned} \tag{6.5}$$

Comme α doit être un entier, on définit :

$$\alpha = \lfloor \log_2(\sqrt{N/K} - 1) - 1 \rfloor. \tag{6.6}$$

Afin que le $\log_2(\sqrt{N/K} - 1)$ soit valide, on doit respecter la condition $K < N$. Cependant la limitation sur K est plus forte si on considère que α, qui définit le nombre d'itérations du processus, doit être positif ou nul. On obtient alors :

$$\begin{aligned} \alpha &\geq 0 \\ \log_2(\sqrt{N/K} - 1) - 1 &\geq 0 \\ \log_2(\sqrt{N/K} - 1) &\geq 1 \\ \sqrt{N/K} - 1 &\geq 2 \\ \sqrt{N} &\geq 3 \cdot \sqrt{K} \\ K &\leq \frac{N}{9}. \end{aligned} \tag{6.7}$$

Ainsi, le nombre de superpixels définit par l'utilisateur peut être au maximum de $N/9$, cela est cohérent avec le fait que les blocs de plus petite taille 3x3 contiennent 9 pixels. Cette limitation ne pose pas de problème en ce qui concerne le processus de segmentation puisque, à titre d'exemple, avec une image VGA de résolution 640x480=307200 pixels, le nombre de superpixels maximum peut être de 34133, ce qui est largement suffisant, à notre connaissance, pour tous les types de traitement qui utilisent des superpixels.

6.2.4/ Évaluation de la complexité algorithmique

Afin d'évaluer l'efficacité de la méthode, nous avons déterminé la complexité algorithmique en fonction de la quantité de pixels dans l'image. La méthode *SLIC*, qui sert de référence à notre méthode, a une complexité algorithmique linéaire en $O(N)$, dans sa version optimisée ou les superpixels évoluent dans une localité restreinte. Bien que notre méthode permette de gagner du temps dans la construction de la segmentation en superpixels, nous n'avons pas encore évalué son comportement en augmentant le nombre de pixels. Cette question est intéressante car l'association des pixels aux différents superpixels dans notre méthode se fait grâce à une hypothèse qui permet de limiter le nombre d'opérations et de pixels utilisés par un facteur non linéaire pour la phase itérative du processus. En effet, en évaluant la cohérence d'un bloc unique de niveau i et de taille $\tau_i = 2^{\alpha-i+1} + 1$, la quantité de pixels dont la distance aux superpixels potentiels est calculée pour ce bloc est $\gamma_i = 2 \cdot (\tau_i - 1)$. Par exemple pour un bloc de taille 9x9, 16 pixels seront évalués (*c.f.* Figure 6.5). Nous avons donc :

$$\gamma_i = 2^{\alpha-i+2}. \tag{6.8}$$

Cela permet de mettre en perspective l'intérêt de la méthode. Avec un nombre d'association égale à γ_i, le nombre de pixels assignés à un superpixel dans le cas d'un bloc cohérent est égale à τ_i^2. Ainsi on obtient la courbe Figure 6.9 où on peut observer le nombre de pixels attribués en fonction de la quantité d'opérations d'association effectuées pour un bloc de taille maximale τ_0 défini pour un α variant de 0 à 10.

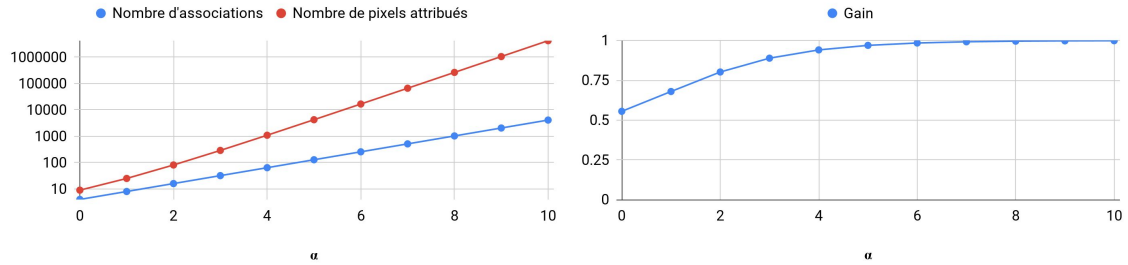


Figure 6.9 – Evaluation du nombre de pixels attribués en fonction de la taille du bloc définit par l'itération notée α. À gauche les courbes, avec une échelle logarithmique pour mieux visualiser la progression, et à droite le gain en pourcentage défini par le nombre de pixels économisés dans l'attribution d'un bloc de taille maximale τ_0 en fonction de α.

Dans la pratique, α n'atteint que très rarement la valeur de 6 (ce qui correspond à un bloc de 256 pixels de côté) dans des images de très grande résolution (UHD 4K) et pour un faible nombre de superpixels et est généralement égal à 3 ou à 4 comme cela suppose des blocs de taille 32 ou 64 pixels de côté offrant une économie de 90% des pixels du bloc.

Il est possible de déterminer le nombre d'opérations d'association de pixels pour chaque niveau d'itération et pour chaque bloc appliqué dans l'image. Bien évidemment, γ_i représente le nombre d'opérations d'association en considérant que le processus n'est pas interrompu dès que le bloc est détecté hétérogène et sans considérer que les blocs se chevauchent. Il est cependant nécessaire d'introduire une mesure du nombre de blocs à traiter pour chaque niveau d'itération, en effet comme les blocs qui sont cohérents ne sont pas subdivisés, il faut définir le nombre de blocs pour chaque niveau d'itération considérant le nombre de blocs de l'itération précédente qui sont incohérents. Ainsi on obtient la définition de U_i, le nombre de blocs à traiter pour l'itération i :

$$U_i = 4 \cdot (U_{i-1} - \epsilon_{i-1}), \tag{6.9}$$

avec $U_0 = \Phi_0$ le nombre de blocs de plus haut niveau appliqué dans l'image et défini par $\Phi_0 = \lceil \frac{W}{\tau_0} \rceil \cdot \lceil \frac{H}{\tau_0} \rceil$, avec W la largeur en pixels de l'image, H la hauteur en pixels de l'image et $\lceil \cdot \rceil$ l'opérateur entier supérieur. ϵ_{i-1} définit le nombre de blocs de l'itération précédente qui sont cohérents et ne sont pas considérés par la subdivision. Le nombre total d'opérations Γ est donc égal à la somme de toutes les opérations d'association de pixels au superpixel le plus proche en considérant chaque niveau de bloc. L'implémentation optimisée de notre méthode considère le chevauchement des blocs au niveau 0 et permet de simplifier l'expression du nombre de pixels à associer par bloc avec :

$$\gamma_i = \tau_i - 2. \tag{6.10}$$

Par exemple, pour un bloc de taille 9x9, 7 pixels seront évalués (*c.f.* Figure 6.7). Ainsi on a un nombre d'opérations total dans le processus itératif égal à :

$$\Gamma = \sum_{i=0}^{\alpha} U_i \cdot \gamma_i = \begin{cases} U_0 \cdot \gamma_0, & \text{si } \alpha = 0, \\ U_0 \cdot \gamma_0 + \sum_{i=1}^{\alpha} (4 \cdot (U_{i-1} - \epsilon_{i-1}) \cdot (\tau_i - 2)), & \text{si } \alpha > 0. \end{cases} \tag{6.11}$$

La dernière étape de la méthode *IBIS* consiste à associer les pixels individuellement si les blocs de taille supérieure n'ont jamais été cohérents. Dans le cas de blocs de taille 3x3, les pixels associés aux différents superpixels le sont même si le bloc n'est pas cohérent. Ainsi, les *seeds* sont mises à jour avant la dernière étape qui consiste à assigner les pixels restants individuellement au superpixel le plus proche. Cette étape ajoute un nombre d'opérations égale à :

$$\rho = 3 \cdot (U_\alpha - \epsilon_\alpha), \tag{6.12}$$

avec 3 le nombre de pixels restant à associer dans un bloc de taille 3x3 dans notre implémentation. Finalement le nombre d'opérations total nécessaire au calcul de la méthode *IBIS* dans notre version la plus optimisée est égale à :

$$\Gamma_{total} = \Gamma + \rho = \sum_{i=0}^{\alpha} U_i \cdot \gamma_i + 3 \cdot (U_\alpha - \epsilon_\alpha), \tag{6.13}$$

qu'il est possible d'écrire en fonction de α :

$$\Gamma_{total} = \begin{cases} \Phi_0 \cdot (2^{\alpha+1} - 1) + 3 \cdot (U_\alpha - \epsilon_\alpha), & \text{si } \alpha = 0, \\ \Phi_0 \cdot (2^{\alpha+1} - 1) + \sum_{i=1}^{\alpha} (4 \cdot (U_{i-1} - \epsilon_{i-1}) \cdot (2^{\alpha-i+1} - 1)) + 3 \cdot (U_\alpha - \epsilon_\alpha), & \text{si } \alpha > 0. \end{cases} \tag{6.14}$$

On cherche à exprimer le nombre d'opérations total en fonction de N la quantité de pixels dans l'image. Cependant, dans notre méthode, l'atténuation du nombre d'opérations est fonction de ϵ qui définit le nombre de blocs cohérents par itération. Cette valeur est fonction de la structure de l'image et est donc indéterminée.

Par construction, on sait que le nombre d'opérations d'assignation par itération est strictement inférieur à N puisque les blocs sont définis pour ne pas prendre en compte tous les pixels qu'ils contiennent. Ainsi, on effectue systématiquement moins d'opérations avec la méthode *IBIS* qu'avec la méthode *SLIC*. Comme le nombre de pixels qui ne sont pas concernés par les opérations d'associations aux différents superpixels est indéterminé, on peut déduire que dans le pire des cas, au regard de la complexité algorithmique de *SLIC* en $O(N)$, notre méthode a une complexité algorithmique également en $O(N)$.

Nous avons déterminé expérimentalement les valeurs moyennes du pourcentage de pixels à associer individuellement dans notre méthode. Nous avons appliqué notre méthode sur l'ensemble de données *BSD500*, fourni par l'université de Berkeley [82] dont un échantillon est donnée Figure 6.10, avec plusieurs résolutions en superpixels.

FIGURE 6.10 – Exemple d'images de l'ensemble de données *BSD500*.

Cet ensemble de données est constitué de 500 images de résolution 240x320 pixels au format portrait ou paysage et est l'un des plus présent dans la littérature pour l'évaluation des méthodes de superpixels. A chacune des images de l'ensemble de données correspond une segmentation vérité terrain qui identifie les objets dans la scène. Cette segmentation de référence nous permet de vérifier la bonne adhérence des frontières définies par la segmentation en superpixels avec les contours des objets dans l'image.

Nous pouvons réaliser l'hypothèse que la moyenne des variations devrait montrer une croissance relative au nombre de superpixels dans l'image afin de correspondre à l'augmentation de la densité des frontières entre superpixels dans l'image. La Figure 6.11 permet de visualiser ces variations.

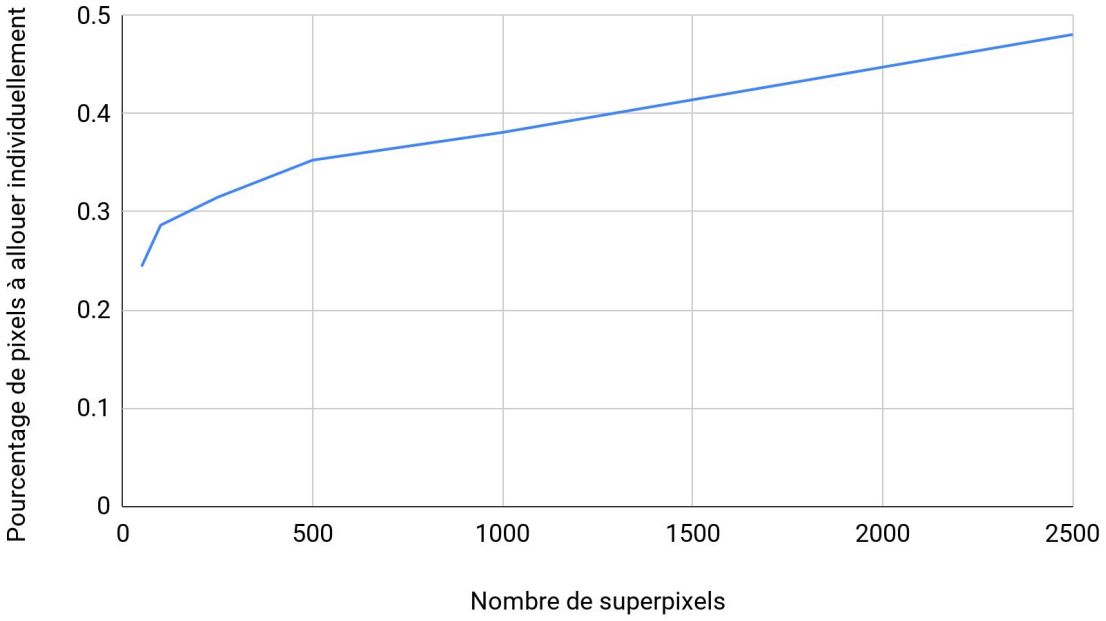


FIGURE 6.11 – Variation du pourcentage de pixels à allouer individuellement en fonction du nombre de superpixels dans l'image sur l'ensemble de données *BSD500* pour une résolution en entrée de 240x320 pixels.

On peut observer que la progression moyenne du nombre de pixels qui sont à évaluer individuellement n'est pas linéaire en fonction de la quantité de superpixels à construire dans l'image. Aussi, c'est le coefficient α qui est le plus déterminant dans la complexité de la méthode. En effet, plus le nombre d'itérations sera important plus le nombre de pixels utilisés dans la dernière étape de la segmentation sera faible puisque cela suppose qu'il y ait moins de superpixels dans l'image donc moins de frontières. Bien que le nombre d'itérations permette de limiter la quantité de pixels utilisés dans l'image, la quantité de pixels utilisés pour définir les frontières au niveau pixels est également fonction de la structure de l'image. La Figure 6.12 permet de visualiser les évolutions du pourcentage de pixels à allouer individuellement en fonction du nombre d'itérations manuellement limité dans notre expérimentation.

Les variations de la quantité de pixels concernés par la dernière étape de la segmentation sont très limitées pour une même résolution en superpixels. On observe que le nombre de pixels à allouer individuellement augmente légèrement, $\approx 1\%$, avec une seule itération et qu'il diminue légèrement, $\approx 1\%$ également, pour 2 ou 3 itérations. Avec un nombre de superpixels croissant, le nombre d'itérations défini par l'équation 6.4 diminue et il n'y a ainsi pas de variations pour 3 ou 4 itérations et 2, 3 et 4 itérations respectivement pour les résolutions en superpixels de 1000 et 2500. Ainsi, on conclut que le nombre d'itérations a une influence négligeable sur la quantité de pixels traités dans la dernière étape de segmentation.

En conclusion de cette section, la méthode *IBIS* permet la segmentation en superpixels avec une complexité comparable à celle des autres méthodes de l'état de l'art. Cependant, le nombre de pixels utiles dans l'image est toujours moindre. Ainsi, pour des paramètres d'entrée identiques, notre méthode est comparativement plus intéressante à mesure que la résolution en pixels de

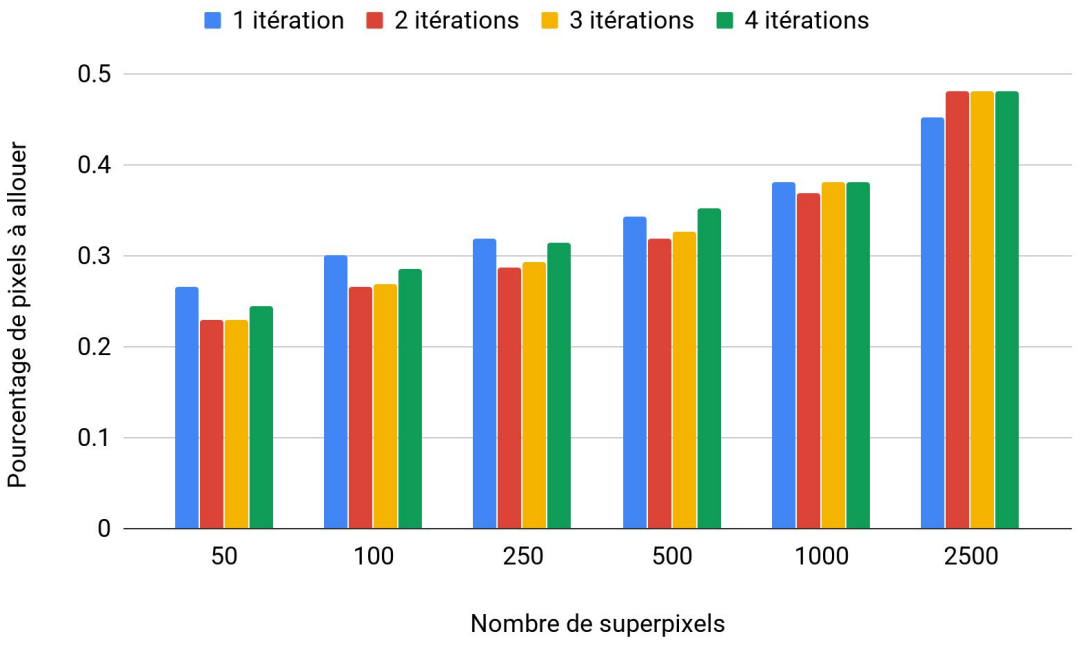


FIGURE 6.12 – Variations du pourcentage de pixels à allouer individuellement en fonction du nombre de superpixels dans l'image et du nombre d'itérations autorisées.

l'image augmente. Afin de valider cette hypothèse nous avons réalisé la segmentation en superpixels sur un nouvel ensemble de données fournissant des images en très haute résolution (*e.g.* 8K). Ces images ont été sous-échantillonnés aux formats 4000x4000, 3000x3000, 2000x2000, 1000x1000, 500x500 et 250x250 afin d'observer les variations du temps de calcul en fonction de la quantité de pixels dans l'image. La Figure 6.13 montre les temps de calcul obtenus par rapport à l'évolution du nombre de pixels dans l'image, qui définit la complexité linéaire.

On peut observer que notre méthode montre une forte linéarité et évolue moins vite que le nombre de pixels dans l'image. Les courbes sont établies en calculant le rapport des valeurs en pixels : N_i/N_0 et pour IBIS : t_i/t_0, avec i variant pour chacune des résolutions appliquées et N_0 et t_0 respectivement la résolution de l'image la plus faible, 250x250 pixels, et le temps de calcul de *IBIS* associé. De plus, la Figure 6.14 permet de visualiser le rapport des deux courbes afin d'observer les variations du coefficient de proportionnalité de notre méthode comparé avec une évolution linéaire du nombre d'opérations.

On observe que le coefficient de proportionnalité évolue légèrement pour des faibles résolutions en pixels de l'image (inférieure à 2 Mpixels) et se stabilise avec un nombre de pixels croissant dans l'image. Comme ce coefficient est calculé à partir des temps de calcul de la méthode *IBIS*, il est fonction de paramètres externes à la méthode : temps d'accès en mémoire par exemple, ce qui explique les variations observées. Ceci étant, les évolutions du coefficient sont très faibles.

Afin de comparer les performances de notre méthode avec les performances des méthodes de l'état de l'art, et afin de montrer la faisabilité d'une intégration sur une plateforme embarquée, une comparaison des méthodes *SLIC* et *IBIS* dans leur version sur GPU sera présentée au Chapitre 7 avec deux intégrations matérielles différentes.

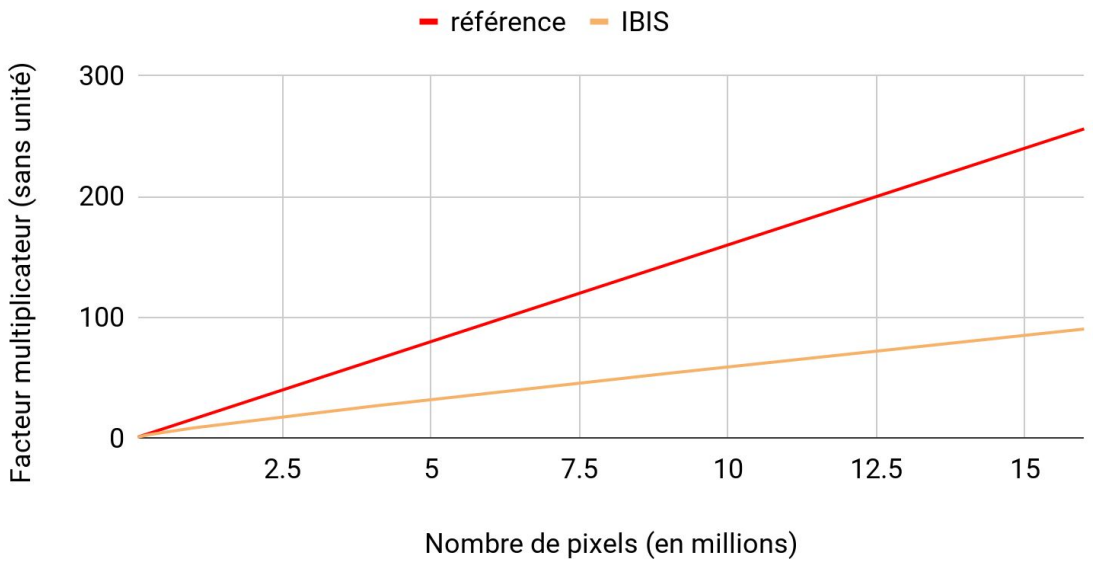


FIGURE 6.13 – Évolution du temps de calcul de la méthode IBIS en fonction de la résolution en pixels de l'image. En rouge, la progression linéaire de la quantité de pixels dans l'image et en jaune la progression du temps de calcul pour la méthode *IBIS* avec notre implémentation optimisée.

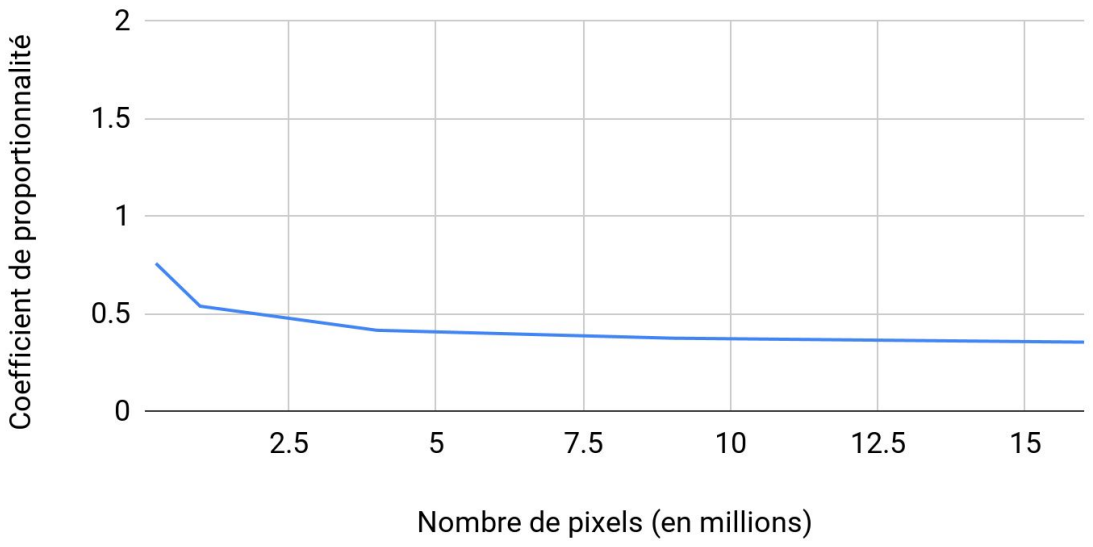


FIGURE 6.14 – Évolution du coefficient de proportionnalité (*i.e.* rapport des deux courbes de la figure 6.13) de la méthode *IBIS* comparé à une évolution linéaire du nombre de pixels.

6.3/ PROPAGATION TEMPORELLE DE LA SEGMENTATION ET POST-TRAITEMENT POUR LA MESURE RPPG

Afin de permettre la mesure rPPG telle qu'elle est proposée dans le chapitre 4 et validée dans le chapitre 5, la segmentation en superpixels doit être propagée temporellement. Pour ce faire, nous avons dans un premier temps propagé les valeurs des *seeds* pour les images consécutives.

Ainsi, chaque superpixel temporel fournit un triplet R, G, B à chaque instant ainsi que le numéro identifiant du superpixel de l'image précédente qui est le parent le plus vraisemblable. Deux superpixels sont considérés adjacents s'ils partagent une frontière commune. Les superpixels parents pour l'image à l'instant $t-1$ sont définis par la minimisation de la distance euclidienne entre les valeurs des *seeds* à l'instant t et celles de tous les superpixels adjacents au superpixel, représenté par la *seeds* courante, à l'instant $t-1$. Comme les pixels peuvent seulement être échangés entre les superpixels adjacents, cette méthode permet de construire des signaux $C^i(t)$ les plus cohérents possible avec une approche simple.

La propagation temporelle des superpixels est réalisée pour les superpixels dont la différence entre les pixels pour deux images consécutives est suffisamment importante. Ces pixels sont réassignés si la différence absolue de la valeur chromatique est supérieure à 10%, ce seuil est défini de façon empirique. Cette approche permet de réduire d'avantage le coût en calcul entre les images en considérant que les scénarios propres à la rPPG impliquent souvent une scène statique ou seul le sujet peut être amené à bouger. Également, les superpixels qui définissent l'arrière-plan sont ainsi stables dans le temps et il n'y a pas de variation au niveau des frontières qui peuvent induire une perturbation dans la mesure du signal et ainsi créer un faux positif dans la détection de tissus vivants.

L'étape de post-traitement est adaptée également. Elle permet de s'assurer que les superpixels sont contigus dans la scène et qu'ils sont tous supérieur en taille à une taille minimale définie par $\frac{\sqrt{N/K}}{4}$ (seuil qui est issu de la méthode *SLIC*), avec N le nombre de pixels de l'image et K le nombre de superpixels spécifié par l'utilisateur. De plus, elle permet de s'assurer que le nombre de superpixels reste stable dans le temps. Pour ce faire, les p plus grands superpixels sont identifiés et coupés en deux nouveaux superpixels, avec p le nombre de superpixels éliminés dans l'étape précédente. Les nouveaux superpixels ainsi créés ont une valeur de *seeds* identique en valeur chromatique et diffèrent uniquement dans leurs coordonnées spatiales. Une fois que les valeurs de *seeds* sont mises à jour, seuls les pixels contenus dans les superpixels éliminés par l'étape de post-traitement et les nouveaux superpixels sont de nouveaux traités afin de mettre à jour la segmentation. La Figure 6.15 permet de visualiser la chaîne algorithmique propre à la propagation temporelle développée pour la mesure rPPG.

Bien évidemment, cette propagation temporelle est très simple et est significativement moins robuste au mouvement que la méthode *TSP* par exemple (utilisée dans les chapitre 4 et 5). Cependant, cette méthode ne requiert que quelques millisecondes afin d'établir la propagation temporelle et permet ainsi de construire les signaux $C^i(t)$ sans avoir à recourir à des méthodes bien plus complexes et couteuses en temps de calcul.

6.4/ Validation

Dans cette section, nous avons évalué notre méthode de segmentation au travers de deux validations différentes. Dans un premier temps, nous avons comparé notre méthode aux méthodes de l'état de l'art afin de valider les performances en termes de qualité de segmentation par superpixels. Cette première validation nous permet de proposer *IBIS* comme une alternative viable aux algorithmes de superpixels existants. En outre, et pour répondre à la problématique de cette thèse, nous avons évalué la version temporellement propagée de notre méthode au travers de l'estimation du rythme cardiaque sur l'ensemble de données *dataset réaliste*. Les temps de calcul sont mesurés afin de vérifier la faisabilité de la mesure rPPG dans un temps acceptable.

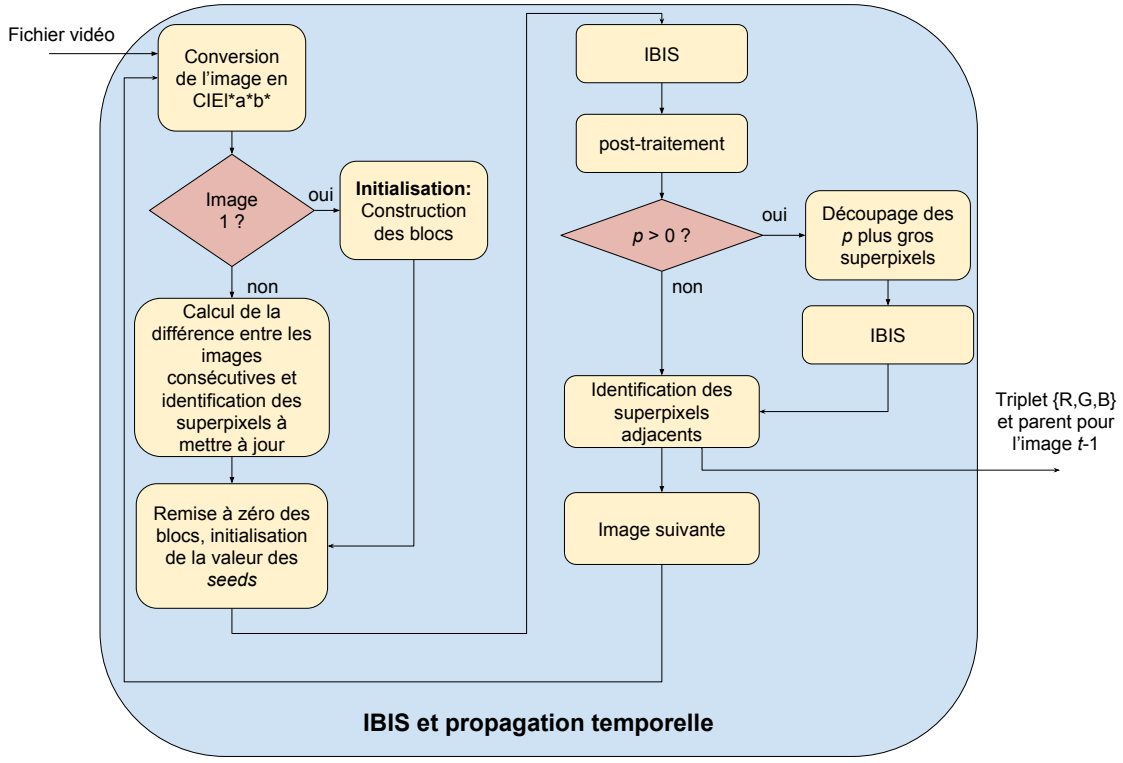


FIGURE 6.15 – Schéma logique de la propagation temporelle pour la segmentation en superpixels temporels *IBIS* telle qu'elle fut développée pour la mesure rPPG dans [139].

6.4.1/ PLATEFORME PC À PROCESSEUR MULTI-CŒURS

Notre première implémentation de la méthode *IBIS* a été développée pour un processeur Intel i7-4790 cadencés à 3.6 GHz. Cette plateforme offre la possibilité de travailler avec 4 cœurs de calcul différents et ainsi de paralléliser l'exécution de notre algorithme. En considérant cet objectif de parallélisation du traitement, nous avons proposé une implémentation permettant de tirer parti d'un nombre variable de cœurs de calculs disponible. Ainsi, notre méthode a donc été développée de sorte à permettre une exécution concurrente dans le calcul des distances entre les pixels et les *seeds*.

Chaque bloc de plus haut niveau et de taille τ_0 est défini par une classe propre. L'objet instancié par cette classe est lui-même composé de 4 sous-objets représentant les 4 sous blocs potentiels qui seront calculés si le bloc de niveau supérieur est incohérent dans l'association de ses pixels aux superpixels potentiels. La hiérarchie des objets est visible Figure 6.16 et permet de visualiser le modèle de parallélisation de notre implémentation de la méthode *IBIS* avec un nombre d'itérations $\alpha = 2$.

La parallélisation s'effectue ici au niveau des blocs de plus haut niveau. Ainsi, l'exécution du traitement pour chaque bloc de niveau 0 est indépendant, comme les blocs ne se chevauchent pas dans cette implémentation, les espaces mémoires sont également indépendants, il est alors possible d'exécuter la chaîne de traitements complète pour chaque bloc de façon indépendante avec les autres. La Figure 6.17 montre l'algorithmique appliqué pour chaque bloc. Pour chaque niveau d'itération, tous les blocs sont exécutés sans distinction entre eux. Ainsi, ce sont les blocs eux-mêmes qui contiennent les informations des pixels dont ils doivent réaliser l'association ainsi

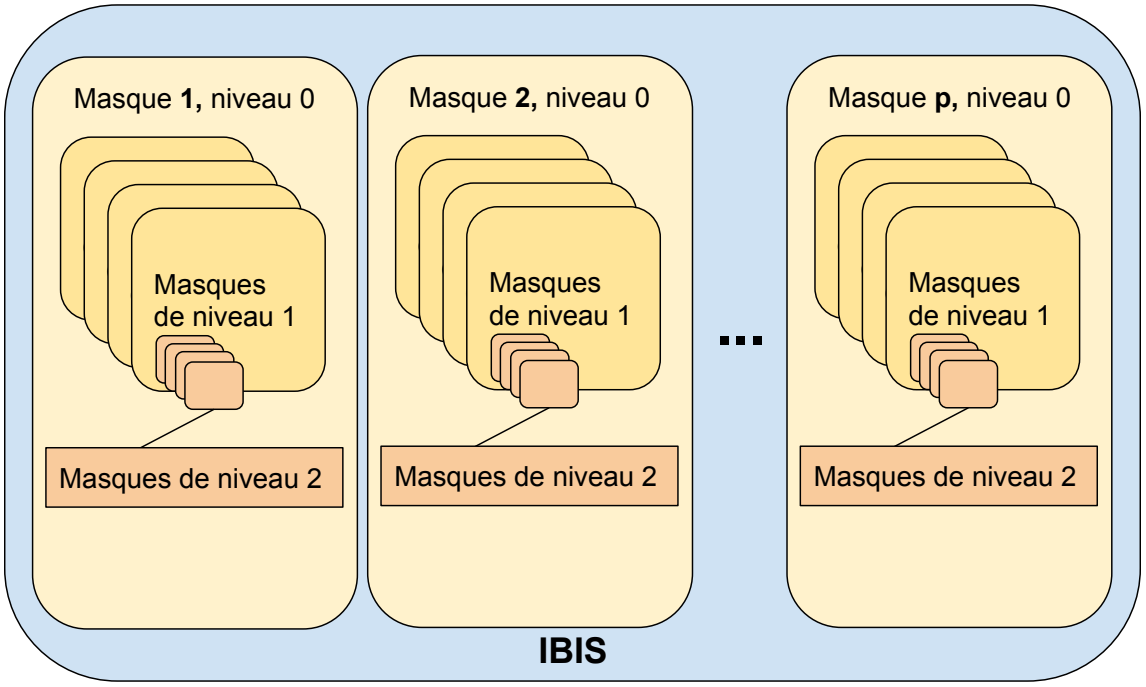


FIGURE 6.16 – Hiérarchie des blocs utilisés dans l'implémentation initiale de la méthode *IBIS* sur CPU. La parallélisation de l'algorithme s'effectue au niveau des blocs de plus haut niveau.

que de leur information de cohérence. Si un bloc de niveau i est cohérent, alors tous les pixels qu'il représente sont déjà associés à un superpixel et il est donc inutile de poursuivre le traitement. En entrée, l'exécution des blocs prend en paramètre le niveau d'itérations courant j, si le niveau demandé est différent de celui du bloc en cours, les sous-blocs sont exécutés à leur tour et le processus de segmentation est ainsi relayé à travers la hiérarchie des blocs.

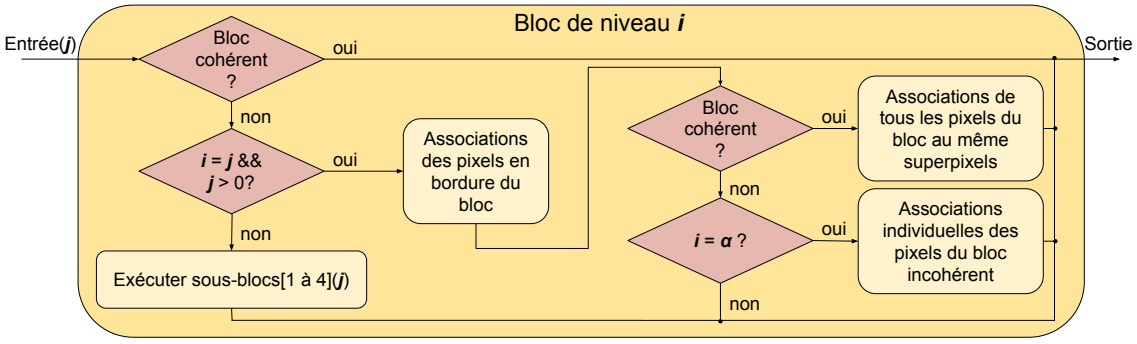


FIGURE 6.17 – Algorithme propre à tous les blocs de niveau i pour une exécution à l'itération j.

Cette implémentation permet d'obtenir les niveaux de performance présentés en section 7.2.1 et 7.2.2. La Figure 6.18 montre l'évolution de la répartition du temps de calcul en fonction de la résolution en superpixels dans l'image. Cela permet de mettre en évidence le coût respectif des fonctions réaliser dans IBIS. De plus cette répartition évolue très peu en fonction du nombre de superpixels comme le montre les courbes. Compte tenu de sa valeur très faible, de 0.02% à 0.2%, la mise à jour des *seeds* est négligé dans la représentation à gauche.

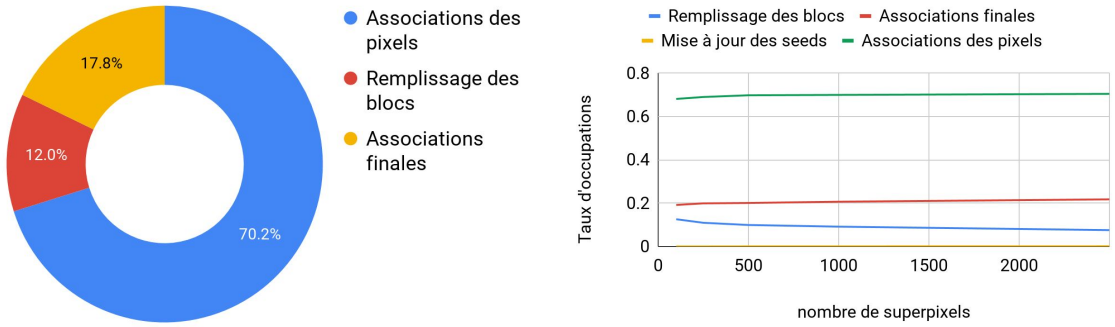


FIGURE 6.18 – Répartition des temps de calcul en fonction de la résolution en superpixels dans l'image. A gauche, la répartition globale moyennée sur toutes les résolutions en superpixels et à droite l'évolution de la répartition des temps de calcul en fonction de la quantité de superpixels pour 50, 100, 250, 500, 1000 et 2500 superpixels.

Comme le montre la Figure, la majorité du temps de calcul est attribué au calcul des distances entre les pixels et superpixels dans le processus itératif avec 70% du temps requis tandis que seulement 18% du temps est nécessaire pour finaliser la segmentation. L'assignation des blocs cohérents représente seulement 12% du temps, ce qui est cohérent puisqu'il ne s'agit que d'assignations en mémoire d'une même valeur, opération très peu coûteuse sur plateforme CPU. Cela permet de mobiliser 88% du temps de traitement pour l'association des pixels aux différents superpixels.

Cette première implémentation a permis de valider les performances de la méthode *IBIS*. Nous avons montré que cette implémentation permet un gain en temps de calcul significatif en comparaison aux méthodes de l'état de l'art sur une même plateforme matérielle et avec des paramètres d'entrée identiques.

6.4.2/ EVALUATION DE LA QUALITÉ DE SEGMENTATION DE IBIS

Dans cette section, la qualité de la segmentation de *IBIS* est évaluée et comparée à 3 autres méthodes de l'état de l'art. Nous comparons ainsi notre approche avec la méthode *SLIC* [58] qui nous sert de référence puisque la méthode d'agrégation des pixels avec le calcul de distance dans l'espace chromatique *CIE l*a*b** et dans le domaine spatial est conservé. Egalement, nous comparons notre approche avec la méthode *SEEDS* [65] qui forme avec *SLIC* une solide base de comparaison dans la littérature pour les nouvelles méthodes de superpixels proposées. Enfin, nous avons également comparé notre approche avec la méthode récente *USEQ* [113] qui exploite les capacités en parallélisation des plateformes modernes pour diviser son temps de calcul.

Chacune des méthodes a été implémentée à partir des codes fournis par leurs auteurs. A ce titre et afin de garantir la reproductibilité de nos expériences, les codes utilisés pour la segmentation en superpixels *IBIS* [1] et la version incluant la propagation temporelle *IBIS_temporal* [2] sont disponibles.

Afin de comparer la qualité de la segmentation en superpixels entre les méthodes, nous avons choisi d'utiliser l'ensemble de données publiquement disponible *BSD500*, présenté en section 6.2.4.

Les métriques utilisées sont également toutes couramment employées dans la littérature afin d'évaluer le niveau de précision de la segmentation en superpixels :

1. https://github.com/xapha/IBIS
2. https://github.com/xapha/IBIS_Temporal

— La métrique **Boundary Recall (BR)** représente le taux d'adhérence entre les frontières qui sont définies par les superpixels et celles fournies par la vérité terrain.
— La métrique **Under segmentation Error (UE)** mesure le taux de chevauchement de chaque superpixel créé par le processus de segmentation sur différents objets identifiés dans la vérité terrain.
— Les **temps de calcul** ont été mesurés à l'aide de la méthode *std : :chrono : :duration* disponible dans la librairie standard du C++. Les temps mesurés comprennent l'étape de segmentation ainsi que le post-traitement. Les tâches d'initialisation nécessaires à la construction de la grille fixe afin d'initialiser les *seeds* ou de construire les blocs dans le cas d'*IBIS* ne sont pas pris en compte du fait que plusieurs images avec les mêmes paramètres d'entrée (taille de l'image et quantité de superpixels désiré) sont traitées sans avoir besoin de réinitialiser les algorithmes. Pour que les comparaisons soient équitables, les algorithmes sont évalués avec une exécution forcée sur un seul cœur de calcul comme *SLIC* et *SEEDS* ne sont pas prévues pour profiter d'une plateforme avec des capacités matérielles de parallélisation avec les implémentations fournies par les auteurs. Également, une exécution sur plusieurs cœurs de calculs a été testée afin d'évaluer le gain en performance sur ce genre de plateforme par les méthodes *USEQ* et *IBIS*.

Pour chacune des métriques, les résultats sont moyennés sur l'ensemble de données pour chacune des résolutions en superpixels employées : 50, 100, 250, 500, 1000 et 2500. Les temps de calcul sont indicatifs des complexités relatives des méthodes. En effet, comme la complexité algorithmique des trois méthodes utilisées pour comparaison est en $O(N)$, avec N la quantité de pixels dans l'image, les temps de calcul sont relatifs au niveau d'optimisation des implémentations utilisées et sont également relatifs à la plateforme matérielle utilisée. Néanmoins, la différence de performance observée entre les méthodes est suffisamment importante pour mettre en évidence les différences de complexité relative.

Dans un premier temps, nous avons comparé nos implémentations afin de vérifier que les optimisations proposées en 6.2.3 n'impactent par significativement la qualité de la segmentation en superpixels. La Figure 6.19 permet de visualiser la qualité de la segmentation pour notre méthode avec les deux implémentations qui sont présentées dans cette thèse. Cette expérience valide la bonne performance de la méthode quelle que soit l'implémentation qui est utilisée. En effet, le principe algorithmique étant inchangé, il ne s'agit là que d'une réduction du nombre de pixels utilisé dans l'image par une répartition plus efficace des blocs dans l'image. Cette version plus performante est utilisée pour le reste des expériences décrites dans ce document.

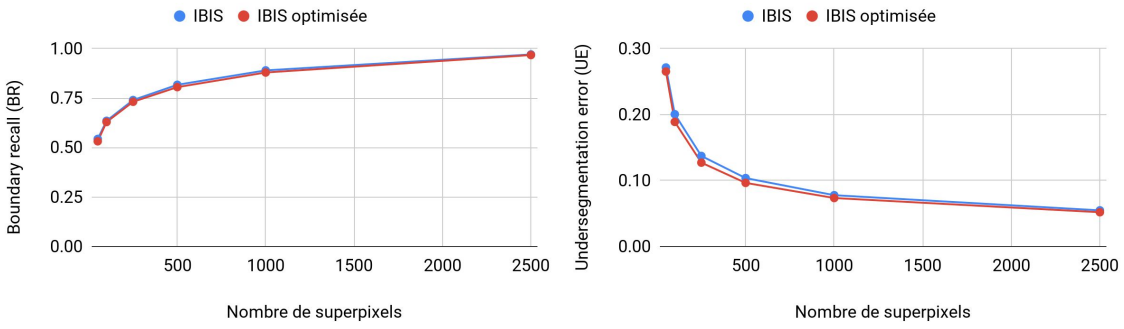


FIGURE 6.19 – Comparaison de *IBIS* et *IBIS* optimisée. À gauche le *Boundary Recall* et à droite la métrique *Undersegmentation Error*. Les métriques sont calculées pour les résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et pour une compacité de 20.

Dans un second temps, nous comparons *IBIS* avec les trois méthodes de l'état de l'art *SLIC*, *SEEDS* et *USEQ*. La Figure 6.20 donne les résultats des métriques pour les différentes résolutions de superpixels. La méthode *SEEDS* obtient de meilleurs résultats en *boundary recall (BR)* et en *undersegmentation error (UE)* et est ainsi la méthode la plus performante en termes de qualité de segmentation dans notre évaluation. La méthode *USEQ* est également très performante en *BR* mais ne se démarque pas sur la métrique *UE*. Enfin les méthodes *SLIC* et *IBIS* sont similaires, ce qui était attendu puisque la méthode est fondamentalement la même en ce qui concerne la construction des superpixels.

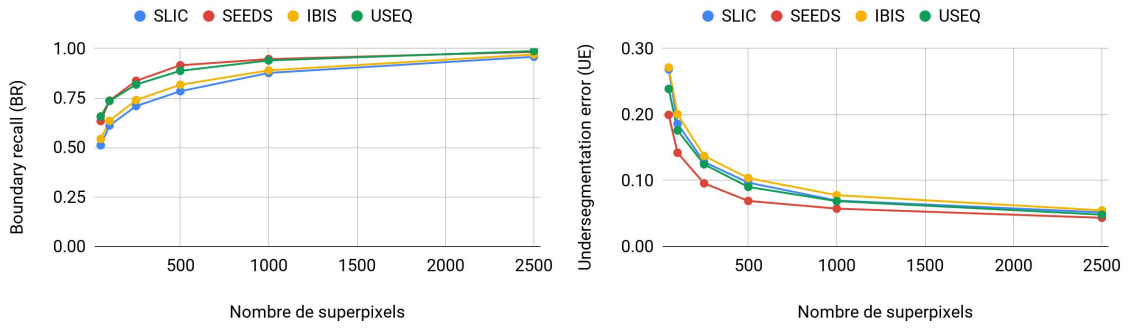


FIGURE 6.20 – Résultats pour la qualité de la segmentation par superpixels pour les méthodes *SLIC*, *SEEDS*, *IBIS* et *USEQ*. Les métriques sont calculées pour les résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et pour une compacité de 20 pour les méthodes qui utilisent ce paramètre.

Comme notre méthode n'utilise qu'une fraction de l'image d'entrée afin de réaliser la segmentation en superpixels, il est intéressant que la qualité de la segmentation soit maintenue avec un temps de calcul significativement plus faible. La Figure 6.21 donne les résultats en temps de calcul obtenus dans notre expérience avec les différentes méthodes implémentées et pour des résolutions en superpixels de 50 à 2500.

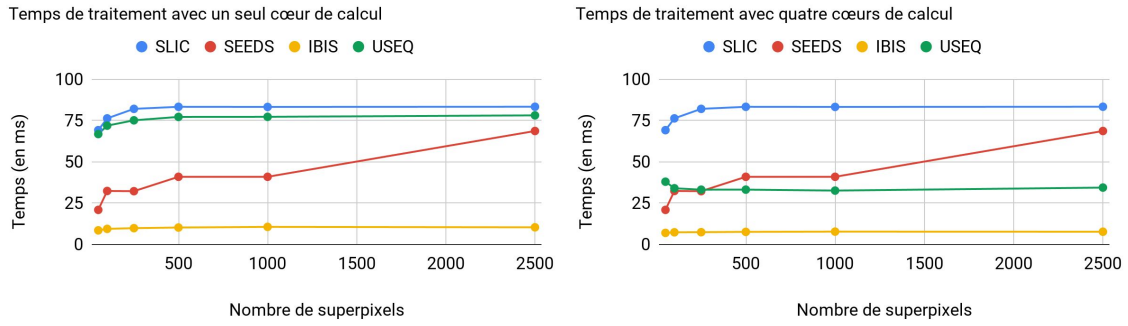


FIGURE 6.21 – Temps de calcul obtenus pour la segmentation en superpixels des méthodes *SLIC*, *SEEDS*, *IBIS* et *USEQ*. L'exécution est réalisée sur 1 cœur de calcul à gauche et à droite sur les 4 disponibles sur le processeur i7-4790 utilisé pour l'expérience.

Du point de vue des performances, les temps de calcul sont sans appel. Avec une exécution forcée sur un seul cœur d'abord, les méthodes *SLIC* et *USEQ* prennent autant de temps pour réaliser la segmentation, environ 80 ms par image indépendamment de la résolution en superpixels. La méthode *SEEDS* est bien plus performante. Cependant, les temps de calcul s'allongent à mesure que la résolution en superpixels augmente, ce qui rend la méthode peu pratique pour du calcul embarqué ou avec des résolutions d'images importantes. Enfin, notre méthode permet d'effectuer

la segmentation avec le même niveau de qualité que la méthode *SLIC* avec un facteur 8 en temps de calcul. La Figure 6.22 met en perspective les temps de calcul relatifs des méthodes évaluées avec la méthode *IBIS* en référence. Ainsi, il est possible de voir que notre méthode est de 2 à 6 fois plus rapide que la méthode *SEEDS* en fonction de la résolution en superpixels. En moyenne, notre méthode maintient sa performance indépendamment de la résolution en superpixels avec un facteur 7.5 en temps de calcul comparé aux autres méthodes.

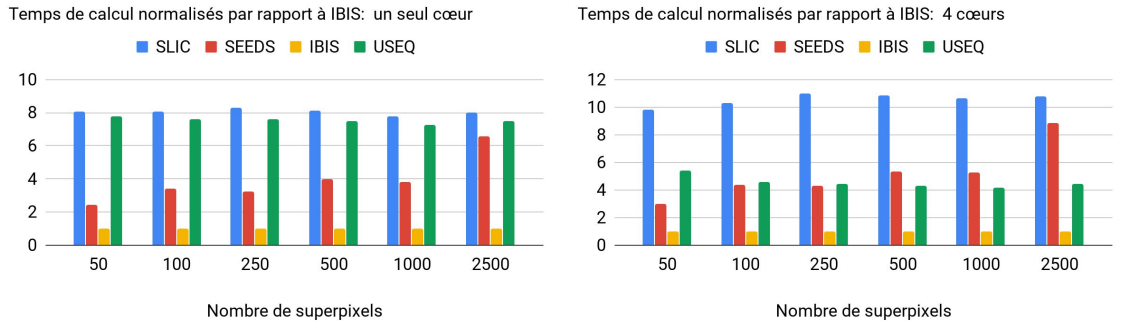


FIGURE 6.22 – Temps de calcul relatifs à la méthode *IBIS* comme référence.

L'accès à la parallélisation, grâce aux multiples cœurs de calcul à disposition sur la plateforme matérielle utilisée, permet à la méthode *USEQ* de rentabiliser son implémentation avec sa répartition du processus de quantification des distances spatiales et chromatiques sur deux *threads* (tâches) séparées. Ainsi, alors que l'écart avec les méthodes *SLIC* et *SEEDS* se creuse au point d'atteindre un facteur de magnitude pour la méthode *SLIC*, la méthode *USEQ* n'est plus que 4 fois plus lente.

En plus des métriques usuelles utilisées afin de quantifier la qualité de la segmentation, il est d'usage de fournir un ensemble d'images segmentées avec deux granularités différentes afin de permettre au lecteur une évaluation qualitative de la qualité de la segmentation. Ainsi, la Figure 6.23 permet de visualiser une mosaïque de la segmentation en superpixels appliquée à un ensemble d'images de l'ensemble de données *BSD500* avec deux résolutions de 250 et 500 superpixels qui sont affichées sur la même image.

Les superpixels générés par notre méthode sont visuellement très similaires avec ceux générés par la méthode *SLIC*, ce qui, une fois encore, était espéré compte tenu des similarités entre les deux méthodes. Dans certains cas cependant, la segmentation avec *IBIS* apparait plus efficace, avec les frontières qui adhèrent mieux aux contours dans l'image, et il est ainsi par exemple possible de voir que les nuages sont mieux délimités avec notre méthode dans la première image. Dans les zones homogènes, les méthodes *IBIS*, *SLIC* et *SEEDS* génèrent des superpixels réguliers tandis que ceux générés par la méthode *SEEDS* sont plus irréguliers ce qui donne un aspect moins qualitatif à la segmentation quand bien même les métriques quantitatives montrent le contraire.

6.4.3/ MESURE RPPG

Notre méthode *IBIS* avec propagation temporelle est également évaluée sur la qualité de la mesure rPPG. Pour ce faire, nous avons remplacé la méthode *TSP* pour la construction des signaux $C^i(t)$ par la méthode *IBIS* dans notre algorithmique qui est présentée au chapitre 4. L'évaluation est faite sur l'ensemble de données *dataset réaliste*. Le rythme cardiaque est estimé sur une fenêtre glissante de 20 secondes avec la méthode *CHROM* [71].

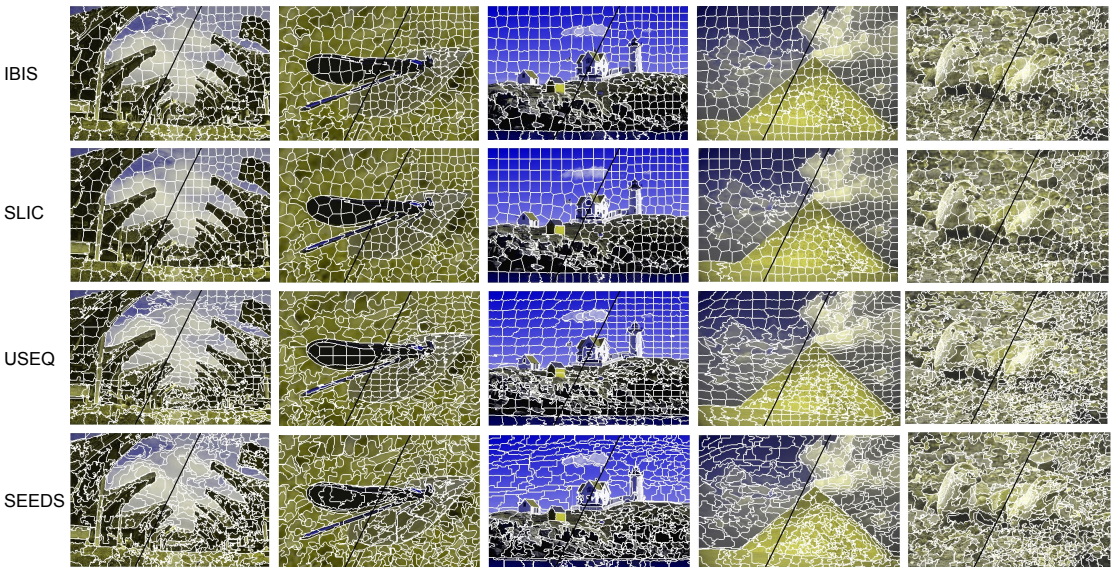


FIGURE 6.23 – Comparaison qualitative pour les méthodes de segmentation par superpixels *IBIS*, *SLIC*, *USEQ* et *SEEDS*. Chaque image est séparée en deux régions qui sont segmentées avec à gauche 250 superpixels et à droite 500 superpixels.

Dans cette expérience, nous avons également évalué l'impact d'une nouvelle approche visant à concilier un grand nombre de superpixels, pour une plus grande précision de la mesure, et une segmentation indépendante de la peau faiblement impactée par le bruit. En effet, pour fournir une segmentation avec la meilleure qualité possible, il est nécessaire de définir une résolution importante en superpixels, mais cela dégrade la qualité du signal en augmentant l'impact du bruit de quantification dans la mesure. Pour prendre en compte ces deux arguments, les superpixels *adjacents*, identifiés par le processus de propagation temporelle, et qui partagent une information chromatique similaire (évaluée avec le calcul de la distance euclidienne chromatique) sont aussi intégrés dans le calcul de la moyenne pour générer le triplet $\{R, G, B\}$ à chaque instant afin de construire les signaux $C^i(t)$. En permettant l'utilisation d'une résolution en superpixels supérieure afin de maintenir la qualité de la segmentation dans l'image et en fusionnant l'information chromatique de plusieurs superpixels temporels adjacents entre eux, il est possible de réduire l'impact du bruit de quantification en considérant une plus grande quantité de pixels. On définit ainsi le nombre de superpixels comme $K = \nu \cdot \beta$, avec β le nombre moyen de superpixels adjacents (en général $\beta = 6$) et ν la résolution en superpixels optimale fixée à 150 dans cette expérience, en référence aux résultats obtenus en section 5.2.2.

La méthode proposée est comparée avec la méthode *TSP*, appelée *référence* dans nos résultats, qui est évaluée dans le chapitre 7. La résolution en superpixels est définie à 150 pour les deux méthodes et à 900 pour notre nouvelle approche en considérant la moyenne des superpixels temporels adjacents, nommée *IBIS adjacent* dans nos résultats. Les métriques employées pour valider la méthode sont : le facteur de corrélation, l'erreur moyenne quadratique (RMSE) et l'erreur moyenne (MAE) ainsi que la précision de la mesure pour des erreurs maximales de 2.5 battements par minute et 5 battements pas minute, ces métriques sont présentées plus en détail dans la section 5.1.3. La précision ainsi que la corrélation sont exprimées entre 0 et 1 et doivent être les plus élevées possible tandis que la valeur de l'erreur RMSE est sans unité et doit être la plus faible possible. Le tableau 6.1 permet de voir les résultats obtenus dans notre expérience.

Avec la métrique de qualité SNR, les performances sont globalement maintenues avec notre nouvelle méthode de segmentation tandis que le temps de calcul pour une image est en moyenne de

Table 6.1 – Résultats pour l'estimation du rythme cardiaque avec la métrique de qualité SNR pour les méthodes *IBIS* et *IBIS adjacent* en comparaison avec la méthode de référence évaluée au chapitre 5.

	Référence SNR	**IBIS SNR**	**IBIS SNR adjacent**
Corrélation	**0.826**	0.751	0.812
Estimation à 2.5 BPM	0.890	0.899	**0.909**
Estimation à 5 BPM	**0.961**	0.940	0.948
RMSE	**2.38**	4.00	3.44
MAE	**1.80**	2.05	2.06

25 ms. Ces résultats permettent d'attester de la qualité de la construction des signaux $C^i(t)$ en utilisant notre méthode de segmentation *IBIS* et sa propagation temporelle. Également, même si elle est mineure, on constate une augmentation de la qualité de la mesure cardiaque sur toutes les métriques avec notre approche *IBIS adjacent* par rapport à *IBIS* simple. Cependant, la méthode de référence est toujours légèrement plus précise dans l'évaluation du rythme cardiaque.

Nous avons également évalué notre méthode avec la métrique de qualité BIC comme les résultats étaient meilleurs dans notre précédente évaluation section 5.2.3. Le tableau 6.2 montre les résultats obtenus avec la métrique de qualité BIC dans les mêmes conditions de tests. Cette fois ci, on constate que notre nouvelle approche visant à offrir un compromis entre bruit de quantification et précision de la segmentation porte ses fruits avec notre méthode *IBIS adjacent* offrant des résultats meilleurs pour toutes les métriques.

Table 6.2 – Résultats pour l'estimation du rythme cardiaque avec la métrique de qualité BIC pour les méthodes *IBIS* et *IBIS adjacent* en comparaison avec la méthode de référence évaluée au chapitre 5.

	Référence BIC	**IBIS BIC**	**IBIS BIC adjacent**
Corrélation	0.819	0.824	**0.859**
Estimation à 2.5 BPM	0.943	0.951	**0.958**
Estimation à 5 BPM	0.971	0.975	**0.984**
RMSE	2.77	2.41	**1.49**
MAE	1.40	1.21	**0.88**

Bien que les écarts soient faibles, en considérant l'excellence des résultats pour l'ensemble des métriques, il est possible de noter que notre méthode de segmentation en superpixels temporels *IBIS* couplée à notre approche de moyenne spatiale avec les superpixels temporels adjacents permet d'obtenir de meilleurs résultats que notre précédente implémentation dans l'estimation du rythme cardiaque. Bien évidemment, le plus grand impact de notre méthode de segmentation *IBIS* est son temps de calcul qui permet la mesure en temps réel du signal rPPG sur l'ensemble de données évalué *dataset réaliste*.

6.5/ Discussions

A l'issue de cette étude, nous avons réalisé une implémentation temps réel de la mesure rPPG sans contact. Avec un temps de calcul moyen de 25 ms, nous avons la capacité de faire la mesure rPPG avec une fréquence image moyenne de 40 images par seconde sur une plateforme PC standard. Cette fréquence image est suffisante pour avoir une définition temporelle du signal rPPG

permettant l'estimation du rythme cardiaque avec la précision obtenue dans les tableaux 6.1 et 6.2. L'étape suivante de notre développement est donc de proposer cette solution au sein d'une plateforme embarquée dédiée. Ainsi, et compte tenu des capacités de parallélisation de notre méthode, nous avons opté pour une plateforme GPU. En effet, ce type de processeur est spécialisé dans le traitement de l'image et permet une parallélisation bien plus poussée des opérations. De plus, nous avons retenu cette plateforme pour les temps de développement courts sur ce type de plateforme en comparaison avec l'utilisation de FPGA par exemple. La plateforme retenue est une carte embarquée de la société Nvidia référencée Jetson TX1. Cette plateforme a été retenue pour son faible coût et ses bonnes performances. En raison des limites de temps disponible pour réaliser notre implémentation, seule la segmentation en superpixels, qui représente l'étape la plus complexe de notre approche, a été développée et évaluée. En effet, les méthodes de traitements du signal utilisées au sein de notre algorithme ne sont pas complexes et sont largement répandues dans les librairies logicielles disponibles.

7

Implémentation sur plateforme embarquée GPU

Sommaire

7.1 Intégration et optimisation : IBIScuda 104
7.2 Comparaison des méthodes sur GPU 108
7.2.1 gSLICr et IBIScuda 108
7.2.2 Evolution des temps de calcul sur Jetson TX1 et Geforce GTX 1080 ti . 109
7.3 Gain par limitation du nombre d'itérations 112
7.3.1 Impact du nombre d'itérations 112
7.3.2 Performances maximales d'IBIScuda sur Jetson TX1 et Geforce GTX 1080 ti 113

Les solutions embarquées de caméras *intelligentes* (*smartcam* dans la littérature) associent un capteur vidéo à une unité de traitement basée sur différents types d'accélérateurs matériels. Ces derniers peuvent être dédiés à une application dans le cas des circuits spécialisés de type ASIC (*Application-specific integrated circuit* : circuit intégré pour une application spécifique), soit plus flexibles tels que les cibles programmables (processeurs multi-cœurs, processeurs spécifiques tels que les DSP ou graphiques GPU) ou encore configurable comme les grands réseaux logiques configurables (FPGA). Ce choix s'effectue en fonction de l'application et plus précisément des caractéristiques (parallélisme potentiel, régularité des traitements, etc.) des algorithmes utilisés. Pour des besoins de traitement à hautes cadences, on peut utiliser des accélérateurs matériels FPGA [122] comme co-processeur vidéo [25]. On peut également utiliser des plateformes hybrides (FPGA + processeur ARM par exemple) [41] afin de s'adapter plus facilement aux besoins de l'application (comme le déploiement rapide d'interfaces de communications standards). Dernièrement, l'utilisation de GPU comme accélérateur matériel s'est multipliée pour un nombre important d'applications en traitement d'images en raison du faible temps de développement nécessaire ainsi que des puissances de calcul croissantes et de la consomation énergetique de plus en plus maîtrisée. On trouve ainsi des exemples d'applications dans la détection du visage [93]. L'usage de ces caméras se multiplie. Des réseaux de caméras *intelligentes* peuvent même être constitués. Les mesures réalisées par chacune des caméras peuvent alors être regroupées, et par inférences logiques, permettre des prises de décisions, par exemple à l'échelle d'un bâtiment [141].

Un capteur rPPG basé sur l'architecture d'une caméra *intelligente* apparaît envisageable au regard des capacités des cibles embarquées actuelles. L'intégration de notre algorithme a permis de mettre en évidence des besoins d'optimisation dans la structuration du code afin de permettre les meilleures performances possibles. Initialement développée en C++ et pour une exécution sur CPU, notre méthode a ensuite été portée sur une plateforme embarquée *Jetson TX1* du construc-

teur *NVIDIA*. Cette plateforme embarque un processeur graphique GPU à faible consommation et puissance. Elle est destinée au marché des systèmes embarqués. Afin d'adapter notre code à cette nouvelle plateforme matérielle, nous avons utilisé les fonctions et outils proposées par la librairie *CUDA* développée par *NVIDIA*.

Dans ce chapitre, nous allons détailler l'implémentation réalisée pour les plateformes GPU. L'objectif du chapitre est de montrer la faisabilité et les avantages d'une solution de mesure rPPG non supervisée au sein d'une plateforme embarquée dédiée : la *Jetson TX1*. Afin d'évaluer les performances de notre méthode avec les méthodes de l'état de l'art, nous comparons l'implémentation sur GPU de notre méthode avec celle de *SLIC* nommée *gSLICr* [103], dont le code est fourni publiquement. Également, les deux méthodes ont été implémentées sur un GPU de référence beaucoup plus performant : *NVIDIA GTX 1080 ti*. Finalement, dans l'objectif d'augmenter encore les performances de calcul, nous avons proposé une optimisation relative au nombre d'itérations de la méthode *IBIS*. Afin de valider cette optimisation, son impact sur la qualité de la segmentation en superpixels est étudiée et enfin les performances maximales de nos implémentations sont détaillées.

7.1/ Intégration et optimisation : IBIScuda

Le tableau 7.1 résume les spécifications [1] architecturales des plateformes embarquées *Jetson TX1* et *Jetson TX2*. La version utilisée dans notre implémentation dispose donc de 256 cœurs de calculs qui sont répartis en 2 unités de traitement et avec 4 Go de mémoire vive qui est partagée avec le processeur *ARM A57*.

Table 7.1 – Spécifications pour les cartes de développement *Jetson TX1* et *Jetson TX2*. L'intégration sur GPU a été réalisée sur la plateforme *Jetson TX1*.

	Jetson TX2	**Jetson TX1**
GPU	NVIDIA PASCAL, 256 coeurs CUDA	NVIDIA Maxwell, 256 coeurs CUDA
CPU	HMP Dual Denver 2/2 MB L2 + Quad ARM A57/2 MB L2	Quad ARM A57/2 MB L2
Mémoire	8 GB 128 bit LPDDR4 59.7 BG/s	4 GB 64 bit LPDDR4 25.6 GB/s

Avec la multiplication des cœurs de calcul, les GPU sont particulièrement adaptés à certains traitements d'images possédant un parallélisme de données et également pour un grand nombre d'applications telles que l'intelligence artificielle, le minage de cryptomonnaie et bien d'autres. Avec la mise à disposition de librairies hauts niveaux comme CUDA et son utilisation dans des librairies de traitement d'image telles qu'OpenCV, les GPU sont aujourd'hui extrêmement versatiles et il est possible de réaliser quasiment n'importe quel type de traitement sur ces plateformes. Bien évidemment, c'est à l'utilisateur de s'assurer de l'intérêt de la plateforme en fonction du calcul à réaliser.

Ainsi, le GPU, utilisé pour nos tests, est un processeur graphique *NVIDIA Maxwell* doté de 256 cœurs de calcul et de 4 Go de mémoire vive partagée avec le processeur *ARM A57* fonctionnant avec 4 cœurs de calcul à une fréquence de 1.2 GHz. Au sein du processeur graphique, les cœurs de calcul sont répartis et sont accessibles suivant une géométrie bien particulière. Ainsi, l'architecture de la cible matérielle employée dispose de deux processeurs de flux (*streaming multiprocessors (SM)*), chaque processeur de flux permet l'exécution d'ensemble de tâches (*threads*), au sein d'une

1. https://www.nvidia.com/fr-fr/autonomous-machines/embedded-systems-dev-kits-modules/?section=jetsonTX1

unité indivisible nommée *warp*, d'une taille minimale de 32 *threads*. Ainsi, il n'est pas possible d'exécuter une tâche sur moins de 32 cœurs de calcul en simultanés et au minimum 64 cœurs afin d'exploiter la pleine puissance de la carte en utilisant les deux *SM*.L'exécution des tâches s'effectue suivant 3 niveaux de géométrie. Chaque *thread* est exécuté au sein d'un *bloc* qui est lui-même défini dans une *grille*. La *grille* référence les blocs en 3 dimensions qui eux même référencent les *threads* en 3 dimensions. Les différents niveaux d'abstraction permettent de différencier les zones mémoires accessibles. Le nombre de blocs n'est pas limité mais il est impossible de lancer plus de 1024 *threads* par *bloc*. La Figure 7.1 permet d'illustrer la répartition des *threads* au sein de l'architecture du GPU.

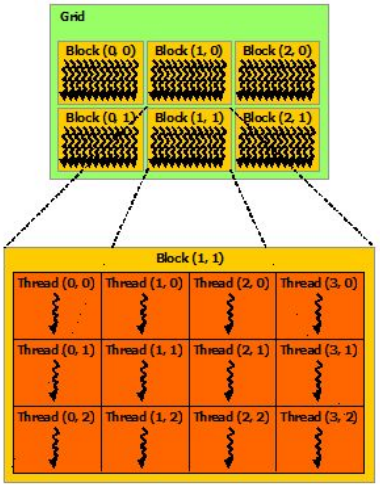


FIGURE 7.1 – Représentation de la répartition des *threads* au sein des *blocs* eux mêmes répartis dans la *grille* [2].

Il existe 4 types de mémoire accessibles depuis les *threads*. La mémoire *constante* permet de stocker 64 Ko d'éléments qui sont accessibles très rapidement depuis tous les *threads*, les éléments qui sont stockés dans cet espace mémoire sont inscrits à l'exécution et sont modifiables en cours de fonctionnement. Cependant les temps nécessaires à la définition des valeurs en mémoire constante ainsi que la taille disponible rendent cet espace mémoire inadapté pour une utilisation en cours de processus. On préfèrera utiliser la mémoire *partagée* qui est un espace mémoire de 64 Ko également accessible en lecture et écriture par tous les *threads* d'un même *bloc*. Cette mémoire à l'avantage d'être bien plus rapide en accès que la mémoire *globale*. La mémoire *globale* est la mémoire complète du GPU la plus vaste (4 Go dans notre cas). Cette dernière est accessible à tous les *threads* de n'importe quel *bloc*. Cependant, les temps d'accès sont beaucoup plus important qu'avec l'utilisation de la mémoire partagée. Ces temps sont d'autant plus long si les accès, pour un même *warp*, ne sont pas réalisés pour des adresses consécutives en mémoire. Il est donc recommander de limiter au maximum les accès en mémoire globale lors de l'exécution. Enfin, chaque *thread* dispose d'un espace de mémoire *local* qui lui est réservé, d'une taille de 64 Ko. Les mémoires *locale* et *partagée* sont libérées à la libération du *bloc* contenant les *threads* en cours d'exécution. Comme les *threads* sont exécutés au sein de *warp* d'une taille minimale de 32 *threads*, ils doivent se terminer en même temps sous peine d'immobiliser les ressources qui sont disponibles pour l'ensemble du *warp*. La Figure 7.2 montre les différents types de mémoires qui sont spécifiques aux éléments de la géométrie *CUDA* (*threads*, *blocs* et *grille*).

En tenant compte de ces éléments, nous avons proposé une implémentation qui permet de tirer par-

2. https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html

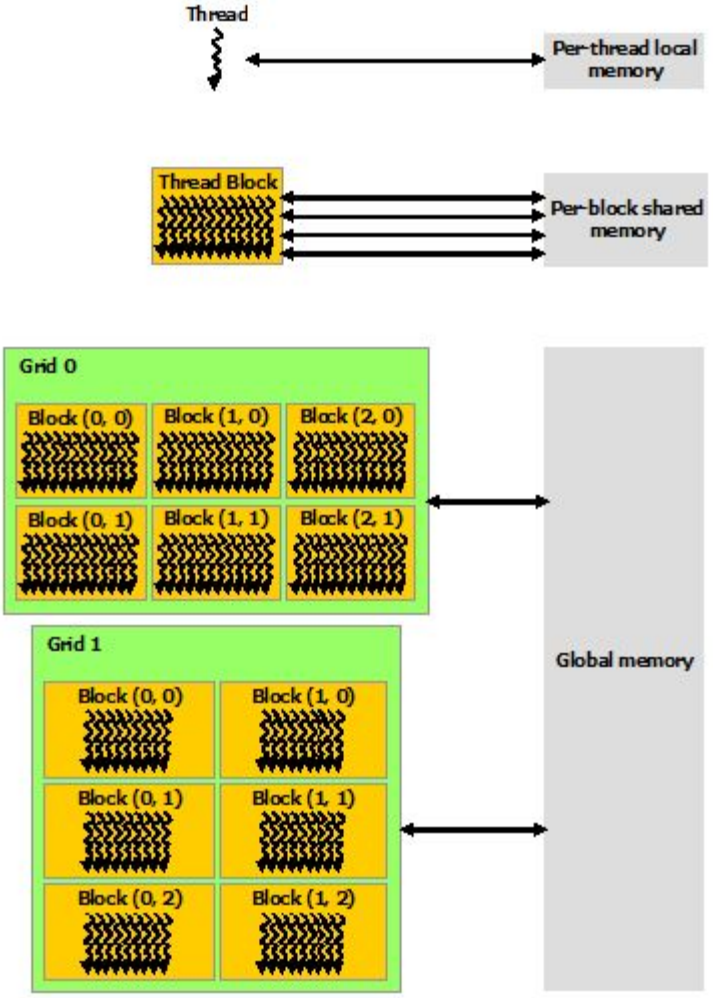


FIGURE 7.2 – Les différents types de mémoire accessibles suivant le niveau d'abstraction dans la géométrie des *threads* CUDA [2].

tie des capacités de la plateforme embarquée. L'exécution des étapes du traitement de la méthode *IBIS* est pilotée de façon séquentielle depuis le processeur CPU de type ARM. Elles sont exécutées en une seule fois sur le processeur graphique GPU. Ainsi, l'architecture logicielle de notre solution diffère dans son modèle de parallélisation des étapes de traitement. Afin de limiter les accès à la mémoire *globale*, la géométrie des blocs ainsi que les coordonnées des pixels qui sont utiles dans l'image sont stockées dans la mémoire *constante* au démarrage de l'application. Ainsi il est possible de définir les coordonnés en mémoire des pixels qui doivent être adressés sans avoir besoin d'accéder à la mémoire *globale*. Comme cet espace mémoire est très limité, les blocs d'une taille τ_i, pour l'ensemble de valeurs $i \in [0; 6]$ sont stockés. Comme il est très peu probable qu'un bloc de dimension supérieure à 256 pixels de coté soit nécessaire dans une utilisation réaliste de la méthode, cela permet de réaliser la segmentation en superpixels quelle que soit les paramètres d'entrée.

Les coordonnées des blocs devant être traités sont stockées dans un vecteur dans la mémoire *globale* du GPU. La répartition de la *grille* et des *threads* au sein des *blocs* se fait par le calcul du nombre de blocs qui doivent être divisés par le nombre de *threads* exécutables en une seule fois

au minimum, dans notre cas 64. Ainsi, on construit un vecteur des coordonnées des blocs à traiter pour chacune des étapes de la méthode *IBIS*. Afin de construire cette liste d'exécution, les coordonnées des blocs à traiter sont ajoutés par différents *threads* répartis dans plusieurs *blocs* à l'aide d'une instruction dite *atomique*. A la différence avec les instructions sur CPU, les instructions *atomiques* sur GPU désignent un accès à une ressource mémoire avec un mécanisme de verrou automatisé afin que plusieurs tâches asynchrones puissent accéder à la même ressource mémoire avec des temporalités obligatoirement différentes. Ce mécanisme est semblable à l'utilisation de *mutex* ou *sémaphore* dans les programmes multitâches sur CPU.

Comme le nombre de blocs par itération est variable en fonction de la structure de l'image dans la méthode *IBIS*, il est nécessaire de retourner vers la mémoire du CPU les informations relatives au nombre de blocs qui doivent être traités pour chaque nouvelle itération. Ces échanges doivent bien évidemment être limités au maximum comme ils ralentissent le traitement. Les différentes fonctions qui sont exécutées par le GPU, nommées *kernels*, sont réparties en deux groupes. Pour la partie mémoire : il s'agit soit de copie depuis la mémoire du CPU vers celle du GPU ou inversement et de copie interne à la mémoire du GPU. Pour la partie calcul : les opérations sont définies par l'utilisateur et doivent être les plus simples possible. Dans le cas de la plateforme *Jetson TX1* les mémoires GPU et CPU sont regroupées au sein d'un même circuit de mémoire, les espaces alloués au CPU et au GPU sont définis en fonction de la quantité de mémoire utilisée par l'un ou l'autre des processeurs. Aussi, les transferts entre la mémoire CPU et GPU reviennent à faire des copies en mémoire par le système, ce qui permet de très bonnes performances en considérant la fréquence de la mémoire et celle du bus de données qui sont faibles en comparaison de plateformes GPU plus puissante. Bien que très performante au regard de la plateforme, cette copie reste pénalisante et peut représenter à elle seule un pourcentage important du temps d'occupation du GPU. Afin de limiter l'impact des copies en mémoire, il est possible de paralléliser les opérations de calcul avec celles de transfert.

En raison des différences architecturales, il est possible de voir que la répartition du temps est complètement différente entre les implémentations sur CPU et GPU. La Figure 7.3 montre la répartition des tâches en fonction du type d'opérations.

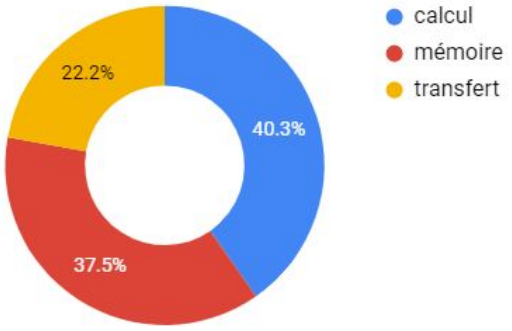


FIGURE 7.3 – Répartition du temps de calcul nécessaire à la segmentation en superpixels avec la méthode *IBIS* en fonction du type d'opérations effectuées.

Ainsi, on distingue trois types d'opérations qui regroupent tous les *kernels* définis dans le code :

— Les opérations de **calcul** qui regroupent les *kernels* : *RGB2LAB* qui permet la conversion de l'image vers l'espace chromatique *CIE l*a*b** depuis l'espace chromatique RGB, *assign_px* qui associe les pixels au superpixel le plus proche, *update_seeds* qui permet de mettre à jour les *seeds* entre chaque itération et *assign_last* qui réalise l'association des pixels restant à la fin du traitement.

— Les opérations sur la **mémoire** qui regroupent les *kernels* : *__c_reset* qui permet d'initialiser les valeurs des *seeds* au début du processus de segmentation, *check_boundaries* qui vérifie pour l'ensemble des blocs de l'itération courante si le bloc est cohérent ou non, *fill_mask* qui met à jour la valeur de la somme de l'ensemble des pixels des différents superpixels afin d'en calculer la moyenne, *fill_mask_assign* qui assigne en mémoire la valeur du superpixel pour l'ensemble des pixels des blocs cohérents et *split_mask* qui met à jour le vecteur de coordonnées où appliquer les blocs pour la prochaine itération.
— Les opérations de **transfert** en mémoire regroupent les appels à la librairie CUDA permettant de réaliser des transferts entre les espaces mémoires avec : *MemCpy (HtoD)* les transferts de données depuis la mémoire du CPU (*Host*) vers la mémoire du GPU (*Device*), *MemCpy (DtoH)* les transferts de données depuis la mémoire du GPU vers la mémoire du CPU et *MemCpy (DtoD)* pour faire des copies internes dans la mémoire du GPU.

On peut voir que les transferts en mémoire occupent 22.2% du temps d'exécution du processus tandis que les opérations de calcul et celles dédiées à l'organisation des données en mémoire et aux transferts entre les espaces mémoires nécessitent sensiblement la même quantité de temps avec 40.3% et 37.5%. Les limitations imposées par la plateforme sont clairement identifiables : on remarque que, par rapport à l'implémentation sur CPU, on exploite beaucoup moins les ressources matérielles afin effectuer les calculs relatifs à la segmentation tandis que l'organisation des données en mémoire représente une partie plus importante du processus.

7.2/ Comparaison des méthodes sur GPU

Nous avons comparé l'implémentation GPU d'*IBIS* avec la version GPU de la méthode *SLIC* : *gSLICr*. Cette implémentation est également développée en s'appuyant sur la librairie CUDA afin d'exploiter les ressources des processeurs graphiques GPU. Afin de maintenir le niveau de qualité de la segmentation, le nombre d'itérations dans l'implémentation *gSLICr* est fixé à 10 et la compacité est fixée à 20, comme dans notre expérience chapitre 5. Les temps de calculs sont mesuré depuis l'acquisition de l'image jusqu'à la fin de la segmentation en superpixels. Ces temps ne comprennent pas les étapes de post-traitement ainsi que la copie de la segmentation depuis la mémoire du GPU vers la mémoire du CPU. En effet, ces étapes ne sont pas obligatoires et sont souvent absentes des traitements.

De plus, les deux méthodes *IBIS* et *gSLICr* sont implémentées sur les deux plateformes GPU *Jetson TX1* et *GTX 1080 ti*. La première plateforme est une solution dédiée de traitement pour les systèmes embarqués tandis que la seconde est une carte de calcul sur station de travail PC qui est largement utilisée dans la littérature.

7.2.1/ gSLICr et IBIScuda

La Figure 7.4 montre les temps de calcul obtenus pour les méthodes *gSLICr* et *IBIS* comparativement à la complexité algorithmique théorique représentée par l'évolution du nombre de pixels dans l'image pour une complexité en $O(N)$ et un coefficient directeur de 1. La courbe est établie en calculant le rapport des valeurs en pixels : N_i/N_0. Pour *IBIS* et *gSLICr*, c'est le rapport des temps : t_i/t_0, avec i variant pour chacune des résolutions appliquées qui permet d'observer l'évolution des temps de calcul pour les deux méthodes. Avec N_0 et t_0 qui sont respectivement la résolution la plus faible 250x250 pixels et les temps de calcul de *IBIS* et *gSLICr* associés.

Comme nous avons montré dans la section 6.2.4, notre méthode est de plus en plus avantageuse à mesure que le nombre de pixels augmente. En effet, les temps de calcul pour notre méthode

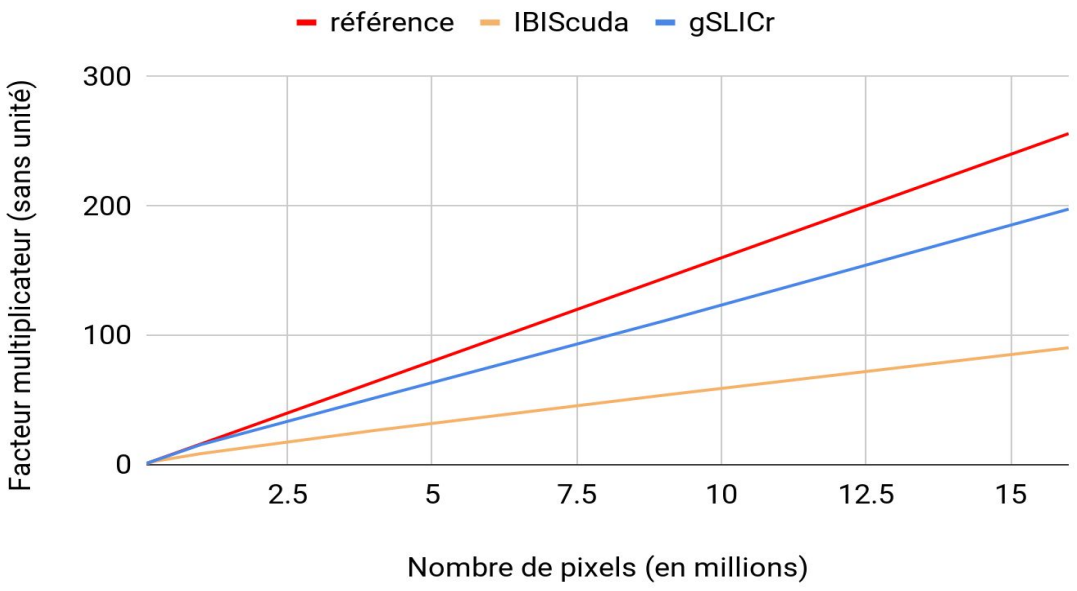


FIGURE 7.4 – Évolution des temps de calcul des méthodes *IBIS* et *gSLICr* en fonction de la résolution en pixels de l'image. En rouge, la progression de la quantité de pixels dans l'image (qui représente la complexité linéaire), en bleu la progression du temps de calcul pour la méthode *gSLICr* et en jaune la progression du temps de calcul pour la méthode *IBIS* avec notre implémentation optimisée.

augmentent moins rapidement qu'avec la méthode *gSLICr*. Notre méthode est plus avantageuse avec les formats vidéo actuels qui tendent à accroitre le nombre de pixels dans l'image comme la qualité de la segmentation est maintenue.

7.2.2/ EVOLUTION DES TEMPS DE CALCUL SUR JETSON TX1 ET GEFORCE GTX 1080 TI

Avec les implémentations des deux méthodes *gSLICr* et *IBIS* sur les deux plateformes GPU *Jetson TX1* et *GTX 1080 ti*, les différences entre les implémentations ainsi que les avantages et limitations des méthodes sont identifiées. En effet, de par leurs différences architecturales, les deux plateformes ne disposent pas de la même puissance de calcul et ont des temps d'accès en mémoire également différents. De plus, la répartition des cœurs de calcul est modifiée, avec un nombre de processeurs de flux multiplié par 12 pour la *GTX 1080 ti* par exemple. Cette comparaison de performances pour des plateformes différentes doit permettre en outre de confirmer la capacité de parallélisation de l'algorithme proposé. La Figure 7.5 permet de visualiser l'évolution des temps de calcul en fonction de la quantité de pixels dans l'image et pour différentes résolutions en superpixels sur la plateforme GPU *Jetson TX1*. On observe que pour la méthode *IBIS*, nommée *IBIScuda* dans la figure, les temps de calcul varient peu en fonction de la résolution en superpixels avec un temps moyen de 355 ms pour une résolution d'image de 4000x4000 pixels et une variation de 18% entre les résolutions en superpixels 50 et 2500. De plus, ces premiers résultats permettent de prédire des capacités de traitements en temps réel jusqu'à des résolutions de 1000x1000 pixels avec une moyenne de 18.3 ms par image.

En ce qui concerne la méthode *gSLICr*, la dispersion est légèrement plus faible avec un temps moyen de 2.56 secondes pour traiter une image de résolution 4000x4000 et une variation de 16%

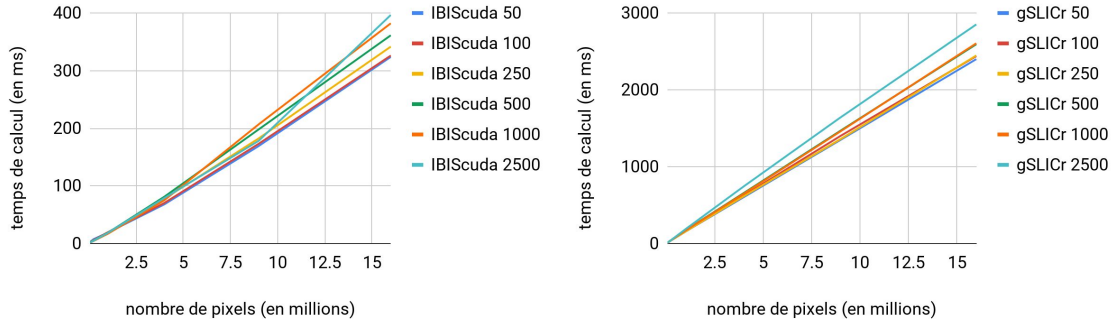


FIGURE 7.5 – Évolution des temps de calcul des méthodes *IBIS* et *gSLICr*, sur la plateforme GPU *Jetson TX1*, en fonction de la résolution en pixels de l'image et pour les résolutions en superpixels 50, 100, 250, 500, 1000 et 2500. A gauche la méthode *IBIScuda* et à droite la méthode *gSLICr*.

entre les résolutions en superpixels 50 et 2500. Les capacités de l'algorithme pour permettre un traitement en temps réel sont bien inférieures avec 42 ms nécessaire pour traiter une image en 500x500 pixels ce qui est inférieur au format standard VGA (640x480 pixels).

Avec la plateforme GPU *GTX 1080 ti*, les choses sont un peu différentes. En effet, la Figure 7.6 montre qu'avec un traitement d'une image de 4000x4000 pixels en un temps moyen de 40.48 ms pour une variation de 8% entre 50 et 2500 superpixels, l'implémentation permet le traitement d'images de résolution 3000x3000 pixels en 24 ms. Ces performances permettent de considérer la segmentation en superpixels pour des applications nouvelles où les temps de traitement étaient jusqu'alors un frein à l'utilisation de ces méthodes.

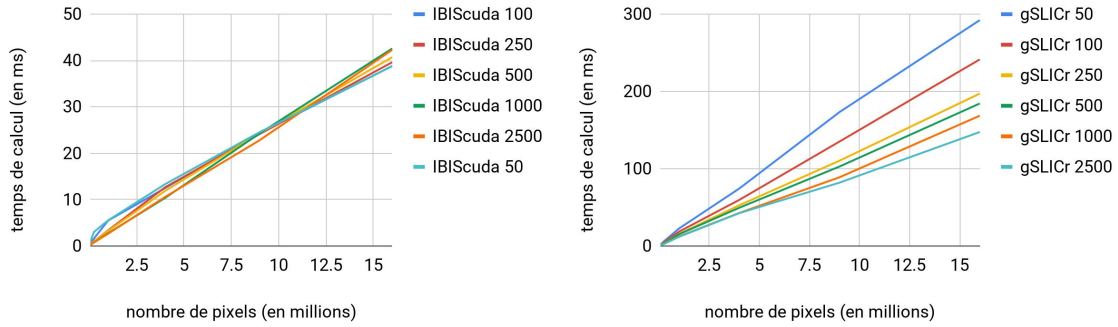


FIGURE 7.6 – Évolution des temps de calcul des méthodes *IBIScuda* et *gSLICr*, sur la plateforme GPU *GTX 1080 ti*, en fonction de la résolution en pixels de l'image et pour les résolutions en superpixels 50, 100, 250, 500, 1000 et 2500. A gauche la méthode *IBIScuda* et à droite la méthode *gSLICr*.

Avec la méthode *gSLICr*, le traitement d'une image de résolution 4000x4000 pixels prend en moyenne 204 ms avec une variation de 53% des temps de calcul entre 50 et 2500 superpixels pour respectivement 300 ms et 140 ms. Ainsi la résolution spatiale de l'image à traiter sera limitée à 1000x1000 afin de garantir des temps de calcul compatible avec une fréquence image de 30 images par seconde par exemple.

Enfin, en comparaison directe sur les deux plateformes utilisées, les méthodes implantées montrent une évolution du temps de calcul très différente en fonction de la quantité de pixels dans l'image. La Figure 7.7 permet de visualiser l'évolution des temps de calcul en fonction de la résolution en

pixels de l'image avec les temps de calcul moyennés pour les différentes résolutions de superpixels utilisées.

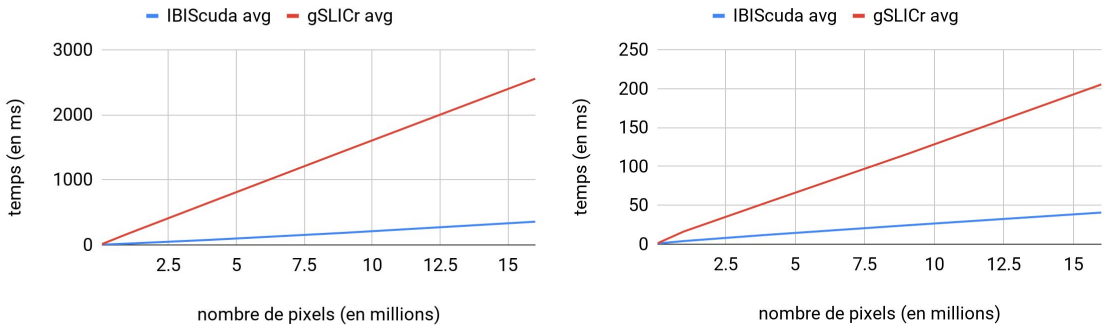


FIGURE 7.7 – Évolution des temps de calcul des méthodes *IBIScuda* et *gSLICr* en fonction de la résolution en pixels de l'image. A gauche sur la plateforme GPU *Jetson TX1* et à droite sur la plateforme GPU *GTX 1080 ti*.

La progression du temps de calcul est ainsi bien plus atténuée pour la méthode *IBIScuda* et on observe un profil similaire à la Figure 7.4 pour les deux plateformes matérielles. On constate également que le gain en performance est plus important avec la plateforme GPU *Jetson TX1* avec une division du temps de calcul de 7.2 fois contre 5 entre les méthodes *gSLICr* et *IBIScuda* pour une résolution en pixels de 4000x4000. En conclusion de cette étude de performance, notre méthode permet de réduire les temps de calcul de manière très significative par rapport à *gSLICr*, considérée actuellement comme une référence dans la littérature. Cette tendance est renforcée lors de l'utilisation de plateformes dédiées à l'embarquée.

Ce gain plus important sur la plateforme *Jetson TX1*, dédiée aux applications embarquées, s'explique en analysant les temps de transfert entre la mémoire CPU et la mémoire GPU. Pour les deux méthodes, l'image est acquise puis transférée vers la mémoire du GPU afin de réaliser la segmentation en superpixels. Cependant, dans le cas de la méthode *IBIS*, en raison du faible nombre d'opérations requises pour effectuer la segmentation, l'opération de transfert peut occuper jusqu'à 60% du temps requis pour le traitement. En conséquence, le gain en puissance de calcul relatif à l'utilisation de la plateforme GPU *GTX 1080 ti* permet une réduction du temps de traitement global plus important qu'avec la méthode *gSLICr*.

En conclusion, l'implémentation *IBIScuda* permet une exploitation optimale des ressources disponibles en raison des limitations en terme de bande passante entre les mémoires CPU et GPU. On peut également noter que la bande passante maximale atteinte est de 10.5 Go/s sur la plateforme *Jetson TX1* alors que l'on atteint un débit maximal de 12.7 Go/s sur la plateforme *GTX 1080 ti*. Le gain en puissance de calcul est beaucoup plus important. Le nombre d'opérations maximal évolue de 511 GigaFlops/s maximal à 11.8 TeraFlops/s maximal pour des opérations sur des flottants en simple précision (32 bits). Cette différence est à imputer aux différences architecturales entre les plateformes matérielles. En effet, sur la plateforme embarquée, la mémoire CPU et GPU est commune et le transfert de l'une vers l'autre se traduit par une copie vers un espace d'adresses différent tandis que pour la plateforme GPU *GTX 1080 ti* à destination des plateformes PC, le transfert s'effectue via le bus PCI-express, généralement avec 16 canaux de communication. Cette différence montre une limitation importante des plateformes PC dans leur interface avec les accélérateurs GPU.

7.3/ Gain par limitation du nombre d'itérations

Afin de limiter le temps de calcul, il est courant dans le cas de la méthode *SLIC* de limiter le nombre d'itérations. Ceci se traduit par une dégradation de la qualité de la segmentation. Il s'agit alors de réaliser un compromis entre temps de calcul désiré et qualité de la segmentation nécessaire pour le traitement que l'on souhaite réaliser. Cette flexibilité est l'une des raisons de la popularité de cette méthode en comparaison à la méthode *SEEDS* par exemple qui permet pourtant une meilleure qualité de segmentation.

Nous avons déjà déterminé que l'influence du nombre d'itérations sur la quantité de pixels à associer individuellement était très limitée avec une variation de moins d'1% (cf. section 6.2.4). Cependant, réduire le nombre d'itérations signifie également une modification de la valeur des *seeds* des superpixels moins importante puisque celles ci sont mises à jours entre chaque itération. Cela se traduit par une chute de la qualité de la segmentation tout comme pour la méthode *SLIC* originale et dans le même temps par une amélioration du temps de calcul. Afin d'évaluer la pertinence d'une telle optimisation et ainsi de conserver une modularité importante, nous avons implémentée une version en fixant la valeur de α à 0. Ainsi, la taille du bloc de plus haut niveau est $\tau_\alpha = 3$ et les *seeds* ne sont mises à jour qu'une seule fois entre la phase d'association des pixels des blocs et l'association individuelle des pixels restants.

7.3.1/ Impact du nombre d'itérations

Comme expliqué en section 6.2.2 le procédé itératif hérité de la méthode *SLIC* permet de mettre à jour les valeurs des *seeds* à la fin de chaque itération. Cette modification améliore la qualité de la segmentation en spécialisant les superpixels au travers de leurs dimensions spatiale et chromatique. Ainsi, les différences entre les superpixels sont accrues et l'adhérence aux contours dans l'image est renforcée. Notre méthode permet de limiter le nombre d'itérations pour des petites tailles de superpixels, où la mise à jour des seeds a peu d'impact, en introduisant une spécification automatique du nombre d'itérations par le calcul de la valeur α, fonction de la résolution en pixels de l'image et de la quantité de superpixels qui sont définis par l'utilisateur. En fonction des paramètres d'entrée, on vérifie expérimentalement que le nombre d'itérations varie entre $\alpha = 1$ et $\alpha = 4$ sur l'ensemble de données *BSD500* pour des résolutions en superpixels de 50 à 2500. Aussi, nous avons évalué l'impact sur la qualité de la segmentation d'une limitation manuelle du nombre d'itérations.

La Figure 7.8 permet de visualiser l'évolution des métriques *boundary recall (BR)* et *undersegmentation error (UE)* utilisées pour quantifier la qualité de la segmentation en fonction de la quantité de superpixels dans l'image et pour un nombre différent d'itération maximal. Ainsi, pour des résolutions en superpixels faibles, la limitation de la valeur du nombre d'itérations entraine une modification des performances tandis que pour des résolutions en superpixels plus importantes les modifications sont faibles ou nulles si le nombre d'itérations défini par le calcul de α est initialement faible et limite donc les optimisations potentielles sur ce paramètre. On explique intuitivement que les pertes de performance sont plus importantes avec des résolutions en superpixels faibles puisque les pixels agglomérés ensembles sont potentiellement très différents puisque répartis sur une zone spatiale plus importante. Ainsi, on observe une perte de performance pour les faibles résolutions en superpixels mais elle reste cependant très modérée avec une modification maximale de 0.526 à 0.539 pour la métrique *BR* et de 0.259 à 0.270 pour la métrique *UE* pour une résolution de 50 superpixels. À mesure que le nombre de superpixels augmente, les pertes de performances diminuent. Ce qui est cohérent avec le fait que le nombre d'itérations défini par le

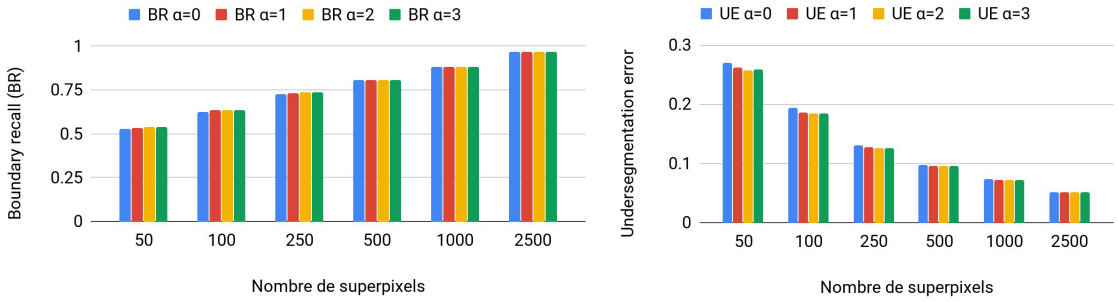


FIGURE 7.8 – Évolution de la qualité de la segmentation pour les métriques *Boundary recall* et *Undersegmentation error* sur l'ensemble de données *BSD500* pour des résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et avec un nombre d'itérations limité à 1, 2, 3 et non limité avec $\alpha = 3$ (soit 4 itérations).

calcul de α est naturellement plus faible et la qualité de la segmentation subit donc moins l'impact de cette réduction.

Du fait de la perte de performance minime dans la qualité de la segmentation, une alternative optimisée de notre méthode est une solution viable. Ainsi, notre méthode offre une adaptabilité importante en fonction des besoins de l'utilisateur entre performances et qualité où le sacrifice en termes de qualité n'est pas significatif avec le calcul actuel des distances entre les pixels et superpixels et la méthode de mise à jour des *seeds*. En effet, on peut prédire qu'avec un procédé plus efficace d'agglomération des pixels en superpixels ainsi qu'une spécialisation de la valeur des *seeds* également plus efficace, le nombre d'itérations aura un impact plus important sur le processus de segmentation en superpixels. Avec une spécialisation des *seeds* plus efficace, la correction apportée par les itérations serait plus importante et la qualité de la segmentation devrait être meilleure.

7.3.2/ PERFORMANCES MAXIMALES D'IBIScuda SUR JETSON TX1 ET GEFORCE GTX 1080 TI

Du fait de la chute de la qualité de la segmentation minimale induite par la limitation de la valeur du nombre d'itérations, nous avons réalisé l'implémentation d'une version optimisée avec α fixé à 0. Cette version encore plus rapide nous permet d'estimer les performances maximales de notre méthode en considérant une perte de qualité pour la segmentation. En considérant les conclusions établies en section 7.2.2, notre méthode montre des variations en performance différentes sur les deux plateformes. En effet, si sur la plateforme embarquée les opérations de calcul et d'organisation des données en mémoire sont les plus consommatrices en temps, ce n'est pas le cas avec la plateforme *GTX 1080 ti* où il s'agit des opérations de transfert entre les mémoires CPU et GPU qui requièrent le plus de temps et limitent les performances de l'implémentation. En conséquence, nous avons réalisé l'implémentation sur les deux plateformes matérielles GPU *Jetson TX1* et *GTX 1080 ti* afin d'identifier l'impact de la plateforme sur cette optimisation. Au vu des résultats décris en section 6.2.2, on peut réaliser l'hypothèse que la diminution du nombre d'itérations devrait avoir un impact beaucoup plus remarquable sur la version embarquée de notre méthode.

La Figure 7.9 montre les temps de calcul pour les méthodes *IBIS* avec α non limité, *gSLICr* et *IBIS* avec α fixé à 0 sur les deux plateformes matérielles utilisées et en fonction du nombre de pixels dans l'image. Les courbes sont issues du calcul de la moyenne sur toutes les résolutions en superpixels utilisés : 50, 100, 250, 500, 1000 et 2500. Comme attendu, les variations en performance sont beaucoup plus remarquables sur la plateforme embarquée *Jetson TX1* avec une diminution

du temps de calcul pour une résolution d'image de 4000x4000 de 355 ms à 86.5 ms soit un facteur d'accélération de 4.1. Pour la plateforme *GTX 1080 ti*, les variations en performance sont moins marquées avec une diminution du temps de calcul maximale de 40.5 ms à 28.6 ms pour une résolution en pixels de 4000x4000, soit un facteur d'accélération de 1.41.

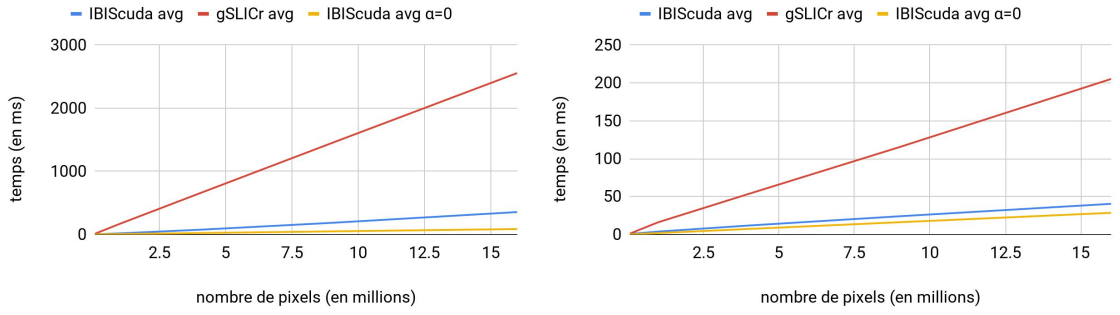


FIGURE 7.9 – Évolution des temps de calcul des méthodes *gSLICr*, *IBIScuda* et *IBIScuda* avec α fixé à 0, en fonction de la résolution en pixels moyennée sur les résolutions en superpixels 50, 100, 250, 500, 1000, 2500. A gauche sur la plateforme GPU *Jetson TX1* et à droite sur la plateforme GPU *GTX 1080 ti*.

La différence entre les deux plateformes entraine une perte d'intérêt dans l'utilisation d'une version limitant le nombre d'itérations avec des plateformes matérielles plus puissantes. En effet, on montre que l'accélération des performances est de moins en moins intéressante avec la plateforme GPU *GTX 1080 ti* avec une résolution en pixels de plus en plus importante. Cependant, l'effet est inversé avec une plateforme moins puissante comme la *Jetson TX1* qui tire davantage parti de la réduction du nombre d'opérations comme le transfert entre les espaces mémoires bride comparativement moins les performances. La Figure 7.10 permet de visualiser le facteur d'accélération entre les deux implémentations de *IBIScuda* en fonction de la résolution en pixels et pour les deux plateformes matérielles.

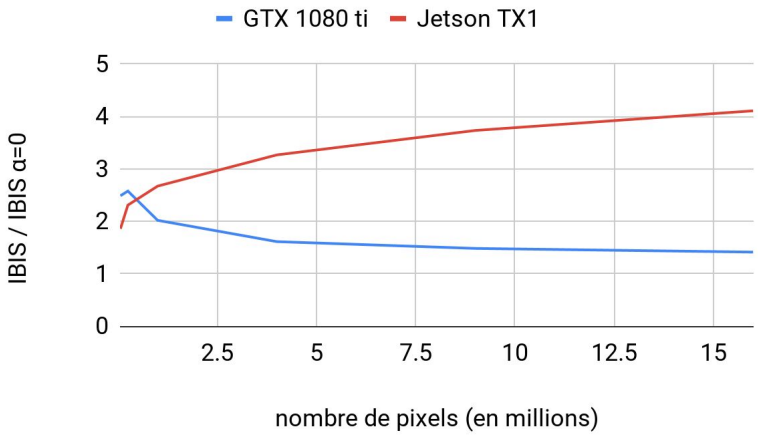


FIGURE 7.10 – Rapport des temps de calcul pour la méthode *IBIScuda* et la méthode *IBIScuda* avec α fixé à 0. En rouge sur la plateforme GPU *Jetson TX1* et en bleu sur la plateforme *GTX 1080 ti*.

Cette représentation permet de mieux visualiser les différences de performance de notre méthode sur les différentes plateformes. À mesure que le nombre de pixels augmente, les accélérations amenées par la limitation du paramètre α s'inverse en fonction des plateformes. Ainsi, pour des

résolutions en pixels très faibles, le cout en transfert mémoire ne représente pas une partie importante du processus de segmentation et l'accélération des performances est plus importante avec la plateforme la plus puissante (*GTX 1080 ti*). Dès lors que l'on augmente le nombre de pixels dans l'image le cout en transfert entre les mémoires CPU et GPU devient, relativement au nombre d'opérations en calcul, beaucoup plus important et limite le gain en performance. Avec une puissance de calcul plus réduite cependant, la plateforme *Jetson TX1* permet de plus tirer avantage de la réduction du nombre d'opérations et l'accélération augmente rapidement à mesure que l'on augmente le nombre de pixels dans l'image.

La Figure 7.11 représente la répartition du temps pour le processus de segmentation en superpixels avec la méthode *IBIScuda* en bridant (à droite) ou non (à gauche) le nombre d'itérations sur la plateforme GPU *Jetson TX1*. Les valeurs sont moyennées sur toutes les résolutions spatiales de 250x250 pixels jusqu'à 4000x4000 pixels et toutes les résolutions en superpixels de 50 à 2500. En prenant en compte que le temps nécessaire au transfert entre les mémoires reste inchangé et représente 22.5% sans limitation du nombre d'itérations et 28.6% avec limitations du nombre d'itération, on observe que la diminution de ce nombre d'itérations entraine une diminution du temps relatif nécessaire aux opérations d'organisations en mémoire de 5.3% et de calcul de 1%. Comme les opérations de calcul conservent une part de $\approx$ 40% dans les deux implémentations et que les opérations sur la mémoire diminuent de 5% en comparaison, on en conclut que la plateforme GPU *Jetson TX1* tire davantage avantage de la limitation du nombre d'itérations. En effet, les opérations en mémoire occupent les cœurs de calcul à 100% alors même que ce sont les temps d'accès en mémoire *globale* qui ralentissent le processus et bloquent les ressources en calcul. Une diminution de la proportion de temps passé à traiter les opérations en mémoire signifie ainsi une utilisation plus efficace de la puissance de calcul à disposition pour les opérations d'associations des pixels aux superpixels.

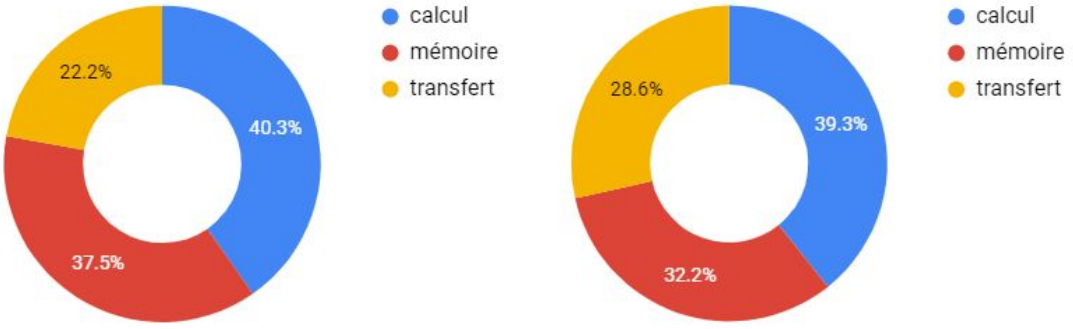


FIGURE 7.11 – Répartition du temps de calcul pour la méthode *IBIScuda* sur la plateforme GPU *Jetson TX1* en fonction du nombre d'itérations et moyenné pour les résolutions en superpixels de 50 à 2500 et pour des résolutions d'images de 250x250 pixels à 4000x4000. A gauche le nombre d'itérations non limité et à droite α est fixé à 0.

L'outil de diagnostic qui est fourni par NVIDIA considère que le taux d'occupation des cœurs de calcul est de 100% lors d'opérations sur la mémoire et ne permet donc pas d'obtenir une mesure fiable du pourcentage d'utilisation des ressources de calcul à disposition sur la plateforme matérielle. Cette approche permet de mettre en perspective les facteurs limitant dans le processus de segmentation en fonction du type d'opérations réalisées.

Avec la plateforme GPU *GTX 1080 ti*, on a montré que le facteur d'accélération était moindre compte tenu de la répartition du temps de calcul et considérant des temps de transfert entre les mémoires limitant. La Figure 7.12 met en perspective la répartition des temps dans le processus de

segmentation avec la méthode *IBIScuda* avec un nombre d'itérations limité (à droite) et non limité (à gauche). Comme précédemment, les valeurs sont moyennées sur l'ensemble des résolutions en pixels des images segmentées et pour toutes les résolutions en superpixels utilisées.

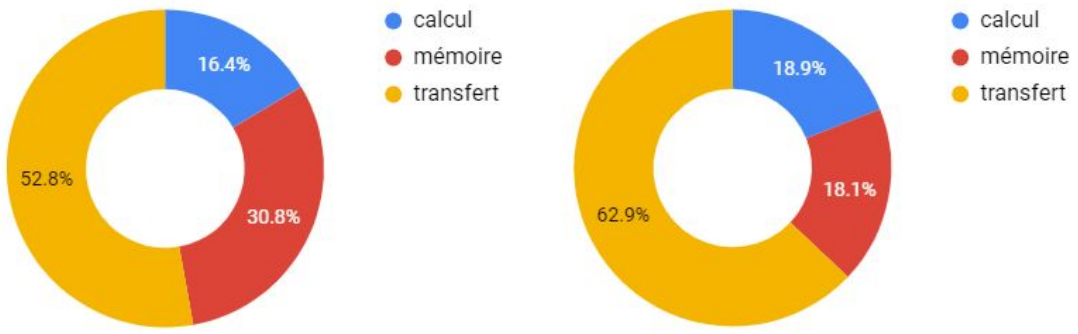


FIGURE 7.12 – Répartition du temps de calcul pour la méthode *IBIScuda* sur la plateforme GPU *GTX 1080 ti* en fonction du nombre d'itérations et moyenné pour les résolutions en superpixels de 50 à 2500 et pour des résolutions d'images de 250x250 pixels à 4000x4000. A gauche le nombre d'itérations non limité et à droite α est fixé à 0.

On observe que dans la répartition des temps pour le processus de segmentation, le transfert occupe plus de la moitié du temps de traitement avec 52.8% sans limitation du nombre d'itérations et 62.9% en limitant le nombre d'itérations et donc d'opérations. Avec la plateforme embarquée, la principale différence entre les répartitions en temps touche les opérations en mémoires qui nécessitent beaucoup moins de temps dans le processus avec une différence de 12.7% entre les implémentations. Le taux d'occupation en opérations de calcul augmente avec un gain de 2.5% entre les implémentations. Ainsi, avec la plateforme GPU *GTX 1080 ti*, le temps de transfert représente une grande majorité du temps nécessaire au processus de segmentation et représente clairement la principale limitation dans la diminution du temps de traitement. La répartition en temps est encore plus avantageuse avec cette plateforme puisque l'on passe d'une répartition avec presque 2 fois plus de temps nécessaire à l'organisation des données en mémoire à une répartition équivalente entre les deux types d'opérations.

La Figure 7.13 donne les fréquences images des différentes méthodes implémentées *IBIScuda*, *gSLICr* et *IBIScuda* avec α fixé à 0 pour des résolutions en pixels en entrée de 250x250 à 4000x4000 pixels et moyennées sur les résolutions en superpixels de 50 à 2500. Ces fréquences images ne considèrent pas les étapes de post-traitement ainsi que le transfert mémoire de la segmentation résultat depuis la mémoire du GPU vers la mémoire du CPU puisque ces étapes sont dépendantes de l'application et ne sont pas toujours nécessaire. De plus ces implémentations ne tirent pas parti de la gestion multi-flux qui permet l'exécution de *kernels* en simultanée pour du calcul et des transferts en mémoire. Cela permet d'optimiser la gestion des ressources dans le cas où l'on traite un flux vidéo par exemple. Avec la plateforme GPU *GTX 1080 ti*, pour les résolutions spatiales où les temps de transfert sont plus longs que les temps de traitement pour la segmentation, il est possible de réaliser l'intégralité du traitement d'une image d'indice i tandis que l'image suivante d'indice $i + 1$ sera copiée depuis la mémoire du CPU vers la mémoire du GPU. Cette optimisation n'est pas implémentée ici mais permettrait une accélération encore plus importante du processus de traitement.

L'étape de segmentation étant l'élément le plus couteux dans la chaîne de traitement de la méthode de rPPG non supervisée tel que développé dans ces travaux de thèse, ces estimations sont compatibles avec la mesure photoplethysmographique. En effet, les différentes étapes de traitement du

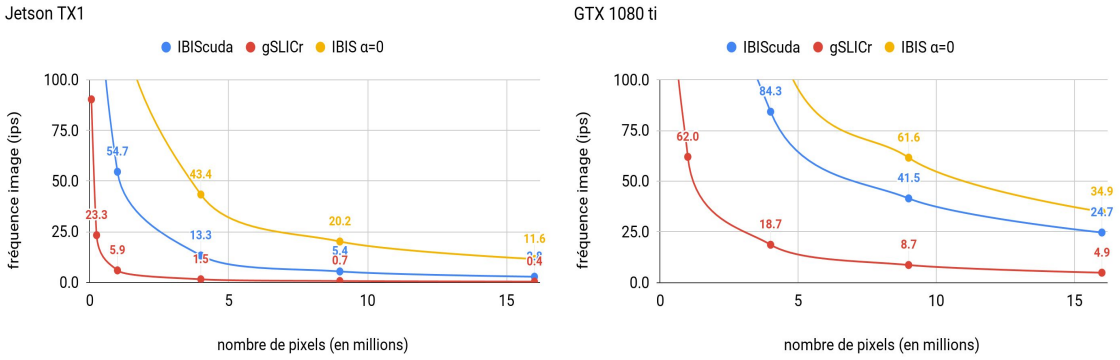


FIGURE 7.13 – Fréquence image pour la segmentation en superpixels pour les méthodes *IBIS*, *gSLICr* et *IBIS* avec α fixé à 0 en fonction du nombre de pixels dans l'image. A gauche pour la plateforme GPU *Jetson TX1* et à droite pour la plateforme GPU *GTX 1080 ti*.

signal sont aujourd'hui très bien optimisées dans les librairies *NEON* pour processeur *ARM*, *IPP* pour les processeurs à architecture *x86* et *CUDA* pour les implémentations GPU de la firme *NVIDIA* ou encore dans la librairie de traitement d'images et d'algèbre linéaire *OpenCV*. De plus l'algorithme de propagation temporelle présenté en section 6.3 est compatible avec une parallélisation importante telle que proposée par les plateformes GPU. La solution de segmentation *IBIS* et sa version optimisée avec α fixé à 0 sont à ce jour, à notre connaissance, les seules solutions permettant une mesure non supervisée avec des résolutions d'image importantes et embarquables dans une plateforme de mesure dédiée.

8

Conclusions et perspectives

Sommaire

8.1 Conclusions 120
8.2 Valorisation 123
8.3 Ouvertures 124

Avec la conclusion de ces travaux de thèse vient la fin de 3 ans de travaux qui ont permis de mettre en évidence des limitations dans les méthodes de photopléthysmographie sans contact et d'y apporter des réponses innovantes. Avec le développement de la mesure non supervisée et l'optimisation de la qualité du signal en favorisant les zones fournissant une information de meilleure qualité, nous avons proposé une nouvelle méthode de mesure qui ne dépend pas d'outils de segmentation supervisée telle que la détection du visage ou de la peau. Ce faisant, nous avons proposé une nouvelle approche de segmentation par superpixels en réduisant grandement le nombre de pixels nécessaires au processus de segmentation. Ces méthodes ont été présentées à 1 conférence internationale, une revue scientifique internationale et 1 workshop international. Également, les efforts de valorisation des travaux de l'équipe de recherche ont permis la publication d'un brevet sur les usages de la technologie dans le cadre de la surveillance de la qualité du sommeil chez les jeunes enfants. Si ces travaux de thèse prennent fin ici, il reste cependant un grand nombre de pistes et validations supplémentaires qu'il est nécessaire de réaliser afin de complètement répondre à la problématique identifiée ici. En effet, il faut encore valider une implémentation complète sur une plateforme dédiée de la méthode afin de valider la conception d'un capteur intégré pour la mesure rPPG. Si les méthodes présentées ici et leurs performances nous permettent de valider la faisabilité d'une telle implémentation, il reste encore des améliorations potentielles à identifier et développer.

Ce chapitre sera décomposé suivant trois sections qui présenteront respectivement les conclusions relatives aux travaux réalisés durant cette thèse de doctorat, les valorisations qui ont été réalisées au cours des travaux de thèse sur les différents aspects des recherches et enfin les perspectives et ouvertures sur les développements et recherches qui sont envisagés dans la poursuite de ces travaux.

8.1/ Conclusions

La photopléthysmographie sans contact est un champ de recherche relativement jeune. Les premiers travaux qui ont réellement mis en avant ce champ de recherche et montrés les possibilités d'une telle mesure sont parus en 2010. Ainsi, jusqu'à aujourd'hui, les principales contributions dans le domaine ont été de proposer des améliorations dans la méthode de combinaison des canaux RGB afin de permettre une mesure de meilleure qualité. On retrouve des méthodes qui permettent de s'affranchir, avec plus ou moins de succès, des différentes sources de perturbation potentielle dans la mesure, qui sont : les variations dans les conditions de l'éclairage, les mouvements du sujet et enfin le bruit de quantification propre au capteur utilisé.

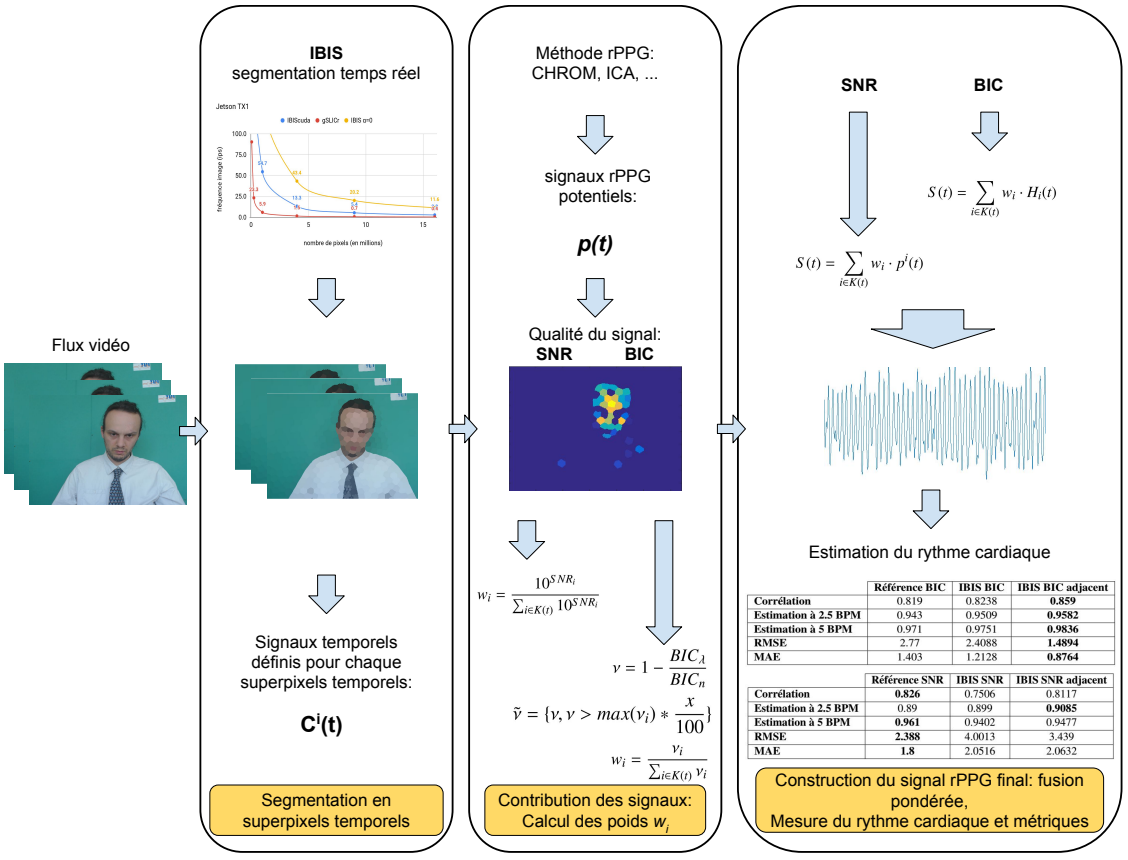
Les procédés de segmentation des zones d'intérêt ont toujours été présentés comme critique mais n'ont pratiquement jamais été modifiés depuis l'utilisation de la détection du visage couplée à la détection de peau. L'utilisation de la technologie rPPG afin d'identifier dans une scène les zones de peau vivante qui contribuent à l'information rPPG est une approche également très récente. Compte tenu du cout en calcul et des plateformes matérielles disponibles, l'intégration de la technologie rPPG au sein d'un capteur dédié n'était possible qu'avec l'utilisation de méthodes de détection supervisée et dont l'implémentation avait été optimisée pour rendre la mesure temps réel, par exemple avec une limitation de la taille des blocs de convolution pour la détection du visage ou encore en limitant la zone de recherche entre plusieurs images successives.

Ces éléments mis en évidence nous ont conduit à définir nos objectifs afin de répondre aux limitations de la technologie rPPG. Ainsi, le développement d'une méthode de segmentation non supervisée pour la mesure rPPG s'est imposé comme une piste de travail intéressante de notre perspective. En effet, comme la littérature est principalement concentrée sur le développement de nouvelles méthodes de rPPG afin d'extraire le signal utile du triplet temporel $\{R(t), G(t), B(t)\}$, il nous est apparu que les choix opérés en matière de segmentation de la zone utile ne permettaient pas de tirer complètement profit des capacités de la mesure rPPG en limitant sa zone d'application ainsi que la qualité de la mesure et par conséquent l'estimation du rythme des pulsations cardiaques.

Notre première contribution fut donc de proposer une nouvelle chaîne algorithmique afin de permettre l'identification implicite des zones de peau au travers de la mesure de la qualité du signal rPPG pour des zones spatiales différentes segmentées à l'aide d'une méthode de segmentation en superpixels temporels. La Figure 8.1 permet de visualiser les différentes contributions qui ont été réalisées dans la thèse. Avec l'identification implicite des zones de peau en fonction de la qualité du signal mesuré, il est possible de favoriser les zones de peau qui contribuent le plus à l'information rPPG. Afin de quantifier la qualité de l'information rPPG, nous avons également proposé une nouvelle approche en utilisant une mesure de vraisemblance (BIC) en lieu et place du rapport signal sur bruit (SNR). Cette contribution nous permet de faire baisser la complexité algorithmique de cette étape d'une complexité en $O(n \cdot \log(n))$ de la transformée de Fourrier en complexité linéaire $O(N)$. De plus, la qualité de l'estimation du rythme cardiaque est meilleure en reconstruisant la série d'états cachés les plus probables qui ont mené aux observations, et en les combinant avec une fusion pondérée. L'analyse du signal résultant permet d'avoir une estimation plus précise du rythme cardiaque. Ces premiers travaux nous ont permis de valider notre approche et nous ont encouragé à poursuivre sur cette voie. Cependant, bien que la méthode soit efficace et permette une mesure de bonne qualité, les contraintes temporelles fortes associées à l'utilisation des superpixels temporels rendaient impossible une intégration temps réel de notre méthode et ce même sur une station de travail aux performances importantes.

Notre deuxième contribution a donc été de proposer une alternative aux méthodes de segmenta-

tion en superpixels actuelles. En effet, si l'on considère l'évolution du format des images ainsi que l'avancée technologique des capteurs optiques matriciels, il devient impossible de réaliser la segmentation en superpixels avec des temps compatibles avec une intégration temps réel, typiquement moins de 30ms. Alors même que la plupart des caméras modernes type *webcam* fournissant un flux vidéo non compressé délivrent des images au format VGA (640x480 pixels), les méthodes de la littérature tel que *SLIC* ou *SEEDS* ne sont pas capables d'opérer avec des fréquences images suffisantes.



	Référence BIC	IBIS BIC	IBIS BIC adjacent
Corrélation	0.819	0.8238	**0.859**
Estimation à 2.5 BPM	0.943	0.9509	**0.9582**
Estimation à 5 BPM	0.971	0.9751	**0.9836**
RMSE	2.77	2.4088	**1.4894**
MAE	1.403	1.2128	**0.8764**

	Référence SNR	IBIS SNR	IBIS SNR adjacent
Corrélation	**0.826**	0.7506	0.8117
Estimation à 2.5 BPM	0.89	0.899	**0.9085**
Estimation à 5 BPM	**0.961**	0.9402	0.9477
RMSE	**2.388**	4.0013	3.439
MAE	**1.8**	2.0516	2.0632

FIGURE 8.1 – Résumé des travaux de thèse en fonction des étapes de la mesure rPPG. De gauche à droite : segmentation en superpixels temporels en temps réel, métriques de qualité pour les signaux issus des méthodes de rPPG et fusion pondérée des signaux rPPG afin de maximiser la qualité de l'estimation du rythme cardiaque.

La méthode proposée *IBIS* permet de répondre à cette problématique en réalisant la segmentation en superpixels avec une fraction des pixels de l'image en entrée. La qualité de la segmentation est maintenue grâce à une attribution individuelle des pixels qui sont les plus proches des frontières entre les superpixels tandis que le temps de calcul est réduit. Également, en augmentant le nombre de pixels avec la résolution des images en entrée grandissante, le temps de calcul augmente avec une forte atténuation en comparaison aux méthodes de l'état de l'art. Cette seconde contribution nous a permis de réaliser la segmentation en superpixels temporels et de reconstruire le signal rPPG avec une moyenne de 25 ms par image sur une plateforme PC équipé d'un processeur *Intel i7-4790k* cadencé à 3.6 GHz. Avec ces résultats, nous avons validé la faisabilité d'un capteur photopléthysmographique sans contact capable de mesurer le signal rPPG sur toutes les surfaces de peau qui fournissent suffisamment d'information de pulsation.

Notre dernière contribution, qui a dirigée la conduite des travaux jusqu'à la fin de la thèse, fut de

réaliser une implémentation sur une plateforme embarquée de notre solution. La plateforme GPU embarquée *Jetson TX1* du constructeur *NVIDIA* est une plateforme avec une puissance de calcul suffisante pour nos besoins. Elle est équipée d'un processeur graphique doté de 256 cœurs de calcul. Nous avons alors implémenté notre solution en portant notre algorithme sur GPU à l'aide de la librairie *CUDA*. Notre implémentation permet de traiter des images avec une résolution spatiale d'un million de pixels en moins de 20 ms. Dans le but d'observer les performances de notre implémentation sur une plateforme bien plus puissante, nous avons réalisé notre intégration sur une plateforme GPU *GTX 1080 ti* dotée d'une architecture matérielle bien plus performante et notamment équipée d'un processeur graphique avec 3584 cœurs de calcul. Notre implémentation de la méthode *IBIS* a alors permis de tirer pleinement parti de la plateforme GPU en étant limité dans sa vitesse de traitement par le taux de transfert entre la mémoire du CPU et celle du GPU. Cela nous a permis d'obtenir des performances très intéressante avec des images de 16 millions de pixels segmentées en $\approx$ 40 ms et une répartition du temps de calcul avec plus de 50% dédié au transfert entre les mémoires du CPU et du GPU de l'image à segmenter. Également, et pour proposer une optimisation supplémentaire si nécessaire, nous avons proposé une implémentation en limitant le nombre d'itérations à 0, ainsi le temps de calcul est encore divisé. Nous avons validé que l'optimisation entrainait une perte de qualité pour la segmentation en superpixels mais que celle-ci restait minime. Cette nouvelle implémentation nous a permis d'obtenir une accélération du procédé de segmentation jusqu'à 4 sur la plateforme embarquée *Jetson TX1*.

Tous les travaux réalisés durant cette thèse permettent de conclure de la validité d'une solution embarquée pour le développement d'un capteur photoplethysmographique sans contact et les méthodes proposées tout au long de la thèse ont été validés sur des ensembles de données publiques. Également, afin de répondre aux exigences de reproductibilité des expériences et pour permettre la libre comparaison des méthodes, nous avons laissé les codes sources en libre accès de la méthode *IBIS* et de sa version sur GPU : *IBIScuda*.

8.2/ Valorisation

Revue internationale :

— **Serge Bobbia**, Richard Macwan, Yannick Benezeth, Alamin Mansouri, Julien Dubois, Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognition Letters, 2017

Conférences et workshops internationaux :

— Richard Macwan, **Serge Bobbia**, Yannick Benezeth, Julien Dubois, Alamin Mansouri, Periodic Variance Maximization using Generalized Eigenvalue Decomposition applied to Remote Photoplethysmography estimation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018

— **Serge Bobbia**, Duncan Luguern, Yannick Benezeth, Keisuke Nakamura, Randy Gomez, Julien Dubois, Real-Time Temporal Superpixels for Unsupervised Remote Photoplethysmography, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018

— **Serge Bobbia**, Yannick Benezeth, Julien Dubois, Remote photoplethysmography based on implicit living skin tissue segmentation, Pattern Recognition (ICPR), 23rd International Conference on, 361-365, 2016

— Julien Dubois, Axel Moinet, **Serge Bobbia**, Roberto Marroquin, Barthélémy Heyrman, Pierre Bonazza, Benoit Darties, Christophe Nicolle, Yannick Benezeth, Johel Mitéran, Dominique Ginhac, WiseEye : A Platform to Manage and Experiment on Smart Camera Networks, 5th Workshop on Architecture of Smart Cameras (WASC), 2016

— **Serge Bobbia**, Julien Dubois, Christophe Nicolle, Axel Moinet, Benoit Darties, A new development framework for multi-core processor based smart-camera implementations, Workshop on Architecture of Smart Cameras (WASC), 2015

Brevet :

— Yannick Benezeth, **Serge Bobbia**, Julien Dubois, Richard Macwan, Alamin Mansouri, “Surveillance de la qualité du sommeil d’un bébé par combinaison de données multimodales incluant la photopléthysmographie“, publié le 14/12/17 sous le numéro WO 2017/212174 A1.

Développements publics :

— Méthode IBIS sur CPU : https://github.com/xapha/IBIS

— Méthode IBIS + propagation temporelle sur CPU : https://github.com/xapha/IBIS_Temporal

— Méthode IBIS sur GPU : https://github.com/xapha/IBIS_cuda

8.3/ Ouvertures

Les travaux de thèse présentés dans ce document ont une grande marge de manœuvre afin d'améliorer les différentes contributions faites ici et pour les différents aspects de la mesure photoplethysmographique sans contact. Tout d'abord, la fusion de données pour la mesure non supervisée n'est pas optimal. En effet, comme montré récemment dans la littérature [133], la phase à l'origine des différents signaux rPPG mesurés à différents endroits du visage peut varier. Ainsi, effectuer la moyenne pondérée de ces signaux peut mener à une perte d'informations en raison des différences de phases entre les signaux. Afin de résoudre ce problème, l'une des solutions présente dans la littérature consiste à utiliser une matrice de retard afin de construire une représentation indépendante de la phase à l'origine des signaux rPPG [142]. Ensuite, la propagation temporelle doit être modifiée afin de proposer une version plus robuste aux perturbations. En effet, l'algorithme proposée dans cette thèse est très simple afin de limiter au maximum le coût en calcul de cette opération. Cependant, cela le rend inadapté à des scénarios plus complexes avec des mouvements importants, occultations et changements d'échelle. Avec notre méthode actuelle, le suivi des superpixels serait grandement perturbés et cela mènerait à une estimation du rythme cardiaque médiocre. De plus, en considérant les performances sur la plateforme GPU *Jetson TX1* atteintes avec notre méthode, il y a la possibilité d'implémenter des algorithmes plus complexes et coûteux en ressources comme la segmentation de l'image se fait beaucoup plus vite. Les algorithmes employés avec la méthode *TSP* par exemple offre une bien meilleure robustesse aux perturbations dans le suivi des superpixels. Bien que l'utilisation des méthodes de calcul de flot optique soient toujours inenvisageable pour une implémentation temps réel, il est possible de complexifier le traitement jusqu'à saturation de la plateforme de calcul.

Enfin, par faute de temps, il nous est impossible de réaliser une intégration complète de la solution de mesure rPPG sur la plateforme GPU. Aussi, cette intégration doit être effectuée afin d'identifier les leviers d'optimisation futurs dans l'utilisation de ces méthodes sur des plateformes embarquées aux capacités limités. Cependant, l'évolution des plateformes matérielles en puissance de calcul est très importante et les limitations qui sont présentes aujourd'hui dans l'usage de certaines méthodes ne le seront plus dans un futur proche. La société NVIDIA vient par exemple d'annoncer une nouvelle plateforme embarquée avec des vitesses d'accès en mémoire supérieures et le double du nombre de cœurs de calcul en comparaison avec la plateforme utilisée dans cette thèse. Ces nouvelles capacités de calcul vont permettre de complexifier les approches afin d'améliorer la robustesse des systèmes et algorithmes existants ou bien de travailler avec des images de résolutions plus importantes.

En marge de la mesure rPPG, la méthode de superpixels IBIS peut également faire l'objet d'améliorations. Dans nos travaux, nous avons effectué une simplification algorithmique de la méthode SLIC. Cependant, les méthodes de superpixels ont depuis bien évolué et les opérations destinées à l'association des pixels aux différents superpixels ainsi que les modalités de mise à jour des *seeds* peut se faire de plusieurs façons [58, 65, 113]. Une étude plus approfondie de l'état de l'art suivie d'une étude expérimentale afin de comparer les performances est une piste envisagée dans la poursuite de ces travaux. Les méthodes de superpixels sont des outils utilisés dans plusieurs thématiques de recherche, il est donc important de pouvoir rivaliser avec les méthodes récentes en termes de qualité de la segmentation tout en proposant une réduction de la complexité algorithmique.

Bibliographie

[1] Hertzman, A. B., and Spealman, C. **Observations on the finger volume pulse recorded photoelectrically**. *American Journal of Physiology-Legacy Content 119* (1937), 334–335. American Physiological Society.

[2] Hertzman, A. B. **The blood supply of various skin areas as estimated by the photoelectric plethysmograph**. *American Journal of Physiology-Legacy Content 124*, 2 (1938), 328–340. American Physiological Society.

[3] Anderson, R. R., and Parrish, J. A. **The optics of human skin**. *Journal of investigative dermatology 77*, 1 (1981), 13–19. Elsevier.

[4] Tur, E., Tur, M., Maibach, H. I., and Guy, R. H. **Basal perfusion of the cutaneous microcirculation : measurements as a function of anatomic position.** *Journal of investigative dermatology 81*, 5 (1983), 442–446. Elsevier.

[5] Karvonen, J., and Vuorimaa, T. **Heart rate and exercise intensity during sports activities**. *Sports Medicine 5*, 5 (1988), 303–311. Springer.

[6] Kamal, A., Harness, J., Irving, G., and Mearns, A. **Skin photoplethysmography—a review**. *Computer methods and programs in biomedicine 28*, 4 (1989), 257–269. Elsevier.

[7] Tremper, K. K. **Pulse oximetry**. *Chest 95*, 4 (1989), 713–715. Elsevier.

[8] Cui, W., Ostrander, L. E., and Lee, B. Y. **In vivo reflectance of blood and tissue as a function of light wavelength**. *IEEE transactions on biomedical engineering 37*, 6 (1990), 632–639. IEEE.

[9] Tomasi, C., and Detection, T. K. **Tracking of point features**. Tech. rep., Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.

[10] Turk, M. A., and Pentland, A. P. **Face recognition using eigenfaces**. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on* (1991), pp. 586–591. IEEE.

[11] Ugnell, H., and Öberg, P. **The time-variable photoplethysmographic signal ; dependence of the heart synchronous signal on wavelength and sample volume**. *Medical engineering & physics 17*, 8 (1995), 571–578. Elsevier.

[12] Dassel, A. C. M., Graaff, R., Meijer, A., Zijlstra, W. G., and Aarnoudse, J. G. **Reflectance pulse oximetry at the forehead of newborns : the influence of varying pressure on the probe**. *Journal of clinical monitoring 12*, 6 (1996), 421–428. Springer.

[13] Such, O. **Near infrared ccd imaging of hemodynamics**. In *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE* (1996), vol. 5, pp. 2105–2106. IEEE.

[14] Such, O., Acker, S., and Blazek, V. **Mapped hemodynamic data acquisition by near infrared ccd imaging**. In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE* (1997), vol. 2, pp. 637–639. IEEE.

[15] Tamura, T., Ogawa, M., Yoda, M., and Togawa, T. **Fully automated health monitoring system in the home**. *IEEJ Transactions on Electronics, Information and Systems 118*, 7-8 (1998), 993–998. The Institute of Electrical Engineers of Japan.

[16] Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E., and Lauer, M. S. **Heart-rate recovery immediately after exercise as a predictor of mortality**. *New England journal of medicine 341*, 18 (1999), 1351–1357. Mass Medical Soc.

[17] Shi, J., and Malik, J. **Normalized cuts and image segmentation**. *IEEE Transactions on pattern analysis and machine intelligence 22*, 8 (2000), 888–905. IEEE.

[18] Wu, T., Blazek, V., and Schmitt, H. J. **Photoplethysmography imaging : a new noninvasive and noncontact method for mapping of the dermal perfusion changes**. In *Optical Techniques and Instrumentation for the Measurement of Blood Composition, Structure, and Dynamics* (2000), vol. 4163, pp. 62–71. International Society for Optics and Photonics.

[19] Kuller, J. M. **Skin breakdown : risk factors, prevention, and treatment**. *Newborn and Infant Nursing Reviews 1*, 1 (2001), 35–42. Elsevier.

[20] Viola, P., and Jones, M. **Rapid object detection using a boosted cascade of simple features**. In *null* (2001), p. 511. IEEE.

[21] Comaniciu, D., and Meer, P. **Mean shift : A robust approach toward feature space analysis**. *IEEE Transactions on pattern analysis and machine intelligence 24*, 5 (2002), 603–619. IEEE.

[22] Huelsbusch, M., and Blazek, V. **Contactless mapping of rhythmical phenomena in tissue perfusion using ppgi**. In *Medical Imaging 2002 : Physiology and Function from Multidimensional Images* (2002), vol. 4683, pp. 110–118. International Society for Optics and Photonics.

[23] Saxena, S., Kumar, V., and Hamde, S. **Feature extraction from ecg signals using wavelet transforms for disease diagnostics**. *International Journal of Systems Science 33*, 13 (2002), 1073–1085. Taylor & Francis.

[24] Tarvainen, M. P., Ranta-Aho, P. O., and Karjalainen, P. A. **An advanced detrending method with application to hrv analysis**. *IEEE Transactions on Biomedical Engineering 49*, 2 (2002), 172–175. IEEE.

[25] Dubois, J., and Mattavelli, M. **Embedded co-processor architecture for cmos based image acquisition**. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on* (2003), vol. 2, pp. II–591. IEEE.

[26] Ren, X., and Malik, J. **Learning a classification model for segmentation**. *International Geoscience and Remote Sensing Symposium* (2003), 10. IEEE.

[27] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. **Face recognition : A literature survey**. *ACM computing surveys (CSUR) 35*, 4 (2003), 399–458. ACM.

[28] Teng, X., and Zhang, Y.-T. **The effect of contacting force on photoplethysmographic signals**. *Physiological measurement 25*, 5 (2004), 1323. IOP Publishing.

[29] Bušek, P., Vaňková, J., Opavskỳ, J., Salinger, J., and Nevšímalová, S. **Spectral analysis of heart rate variability in sleep**. *Physiol res 54*, 4 (2005), 369–376.

[30] Mori, G. **Guiding model search using segmentation**. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (2005), vol. 2, pp. 1417–1423. IEEE.

[31] Byeon, M.-K., Han, S.-W., Min, H.-K., Wo, Y.-S., Park, Y.-B., and Huh, W. **A study of hrv analysis to detect drowsiness states of drivers**. vol. 5, p. 6.

[32] Kim, B. S., and Yoo, S. K. **Motion artifact reduction in photoplethysmography using independent component analysis**. *IEEE transactions on biomedical engineering 53*, 3 (2006), 566–568. IEEE.

[33] Allen, J. **Photoplethysmography and its application in clinical physiological measurement**. *Physiological measurement 28*, 3 (2007), R1. IoP Publishing.

[34] Conaire, C. Ó., O'Connor, N. E., and Smeaton, A. F. **Detector adaptation by maximising agreement between independent data sources**. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–6. IEEE.

[35] Mannheimer, P. D. **The light–tissue interaction of pulse oximetry**. *Anesthesia & Analgesia 105*, 6 (2007), S10–S17. LWW.

[36] Takano, C., and Ohta, Y. **Heart rate measurement based on a time-lapse image**. *Medical engineering & physics 29*, 8 (2007), 853–857. Elsevier.

[37] Trivedi, M. M., Gandhi, T., and McCall, J. **Looking-in and looking-out of a vehicle : Computer-vision-based enhanced vehicle safety**. *IEEE Transactions on Intelligent Transportation Systems 8*, 1 (2007), 108–120. IEEE.

[38] Zheng, J., and Hu, S. **The preliminary investigation of imaging photoplethysmographic system**. In *Journal of Physics : Conference Series* (2007), vol. 85, pp. 12–31. IoP Publishing.

[39] Hülsbusch, M. **An image-based functional method for opto-electronic detection of skin perfusion**. *Ph.D. dissertation* (2008).

[40] Maeda, Y., Sekine, M., Tamura, T., Moriya, A., Suzuki, T., and Kameyama, K. **Comparison of reflected green light and infrared photoplethysmography**. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* (2008), pp. 2270–2272. IEEE.

[41] Mosqueron, R., Dubois, J., Mattavelli, M., and Mauvilet, D. **Smart camera based on embedded hw/sw coprocessor**. *EURASIP Journal on Embedded Systems 2008* (2008), 3. Hindawi Publishing Corp.

[42] Verkruysse, W., Svaasand, L. O., and Nelson, J. S. **Remote plethysmographic imaging using ambient light.** *Optics express 16*, 26 (2008), 21434–21445. Optical Society of America.

[43] Zheng, J., Hu, S., Azorin-Peris, V., Echiadis, A., Chouliaras, V., and Summers, R. **Remote simultaneous dual wavelength imaging photoplethysmography : a further step towards 3-d mapping of skin blood microcirculation**. In *Multimodal Biomedical Imaging III* (2008), vol. 6850, p. 68500S. International Society for Optics and Photonics.

[44] Zheng, J., Hu, S., Chouliaras, V., and Summers, R. **Feasibility of imaging photoplethysmography**. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on* (2008), vol. 2, pp. 72–75. IEEE.

[45] Hu, S., Peris, V. A., Echiadis, A., Zheng, J., and Shi, P. **Development of effective photoplethysmographic measurement techniques : from contact to non-contact and from point to imaging**. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE* (2009), pp. 6550–6553. IEEE.

[46] Vogel, S., Hulsbusch, M., Hennig, T., Blazek, V., and Leonhardt, S. **In-ear vital signs monitoring using a novel microoptic reflective sensor**. *IEEE Transactions on Information Technology inBiomedicine 13*, 6 (2009), 882.

[47] Poh, M.-Z., McDuff, D. J., and Picard, R. W. **Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.** *Optics express 18*, 10 (2010), 10762–10774. Optical Society of America.

[48] Verkruijsse, W., and Bodlaender, M. **A novel biometric signature : multi-site, remote** $(100m)photo - plethysmographyusingambientlight.$

[49] Lewandowska, M., Rumiński, J., Kocejko, T., and Nowak, J. **Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity**. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on* (2011), pp. 405–410. IEEE.

[50] Lim, Y. G., Hong, K. H., Kim, K. K., Shin, J. H., Lee, S. M., Chung, G. S., Baek, H. J., Jeong, D.-U., and Park, K. S. **Monitoring physiological signals using nonintrusive sensors installed in daily life equipment**. *Biomedical engineering letters 1*, 1 (2011), 11–20. Springer.

[51] Maeda, Y., Sekine, M., and Tamura, T. **The advantages of wearable green reflected photoplethysmography**. *Journal of medical systems 35*, 5 (2011), 829–834. Springer.

[52] Martinez, L. F. C., Paez, G., and Strojnik, M. **Optimal wavelength selection for noncontact reflection photoplethysmography**. In *22nd Congress of the International Commission for Optics : Light for the Development of the World* (2011), vol. 8011, p. 801191. International Society for Optics and Photonics.

[53] Poh, M.-Z., McDuff, D. J., and Picard, R. W. **Advancements in noncontact, multiparameter physiological measurements using a webcam**. *IEEE transactions on biomedical engineering 58*, 1 (2011), 7–11. IEEE.

[54] Sepponen, R. **Development of a fractional multi-wavelength pulse oximetry algorithm**, 2011. Aalto University.

[55] Sun, Y., Hu, S., Azorin-Peris, V., Zheng, J., Greenwald, S., Chambers, J., and Zhu, Y. **Detection of physiological changes after exercise via a remote optophysiological imaging system**. In *Design and Quality for Biomedical Technologies IV* (2011), vol. 7891, p. 78910E. International Society for Optics and Photonics.

[56] Sun, Y., Papin, C., Azorin-Peris, V., Kalawsky, R., Greenwald, S., and Hu, S. **Comparison of scientific cmos camera and webcam for monitoring cardiac pulse after exercise**. In *Applications of Digital Image Processing XXXIV* (2011), vol. 8135, p. 813506. International Society for Optics and Photonics.

[57] Yu, S., Hu, S., Azorin-Peris, V., Chambers, J. A., Zhu, Y., and Greenwald, S. E. **Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise**. *Journal of biomedical optics 16*, 7 (2011), 077010. International Society for Optics and Photonics.

[58] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., and others. **Slic superpixels compared to state-of-the-art superpixel methods**. *IEEE transactions on pattern analysis and machine intelligence 34*, 11 (2012), 2274–2282. Institute of Electrical and Electronics Engineers, Inc., 345 E. 47 th St. NY NY 10017-2394 United States.

[59] Kwon, S., Kim, H., and Park, K. S. **Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone**. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (2012), pp. 2174–2177. IEEE.

[60] Lenskiy, A. A., and Lee, J.-S. **Driver's eye blinking detection using novel color and texture segmentation algorithms**. *International journal of control, automation and systems 10*, 2 (2012), 317–327. Springer.

[61] Li, Z., Wu, X.-M., and Chang, S.-F. **Segmentation using superpixels : A bipartite graph partitioning approach**. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012), pp. 789–796. IEEE.

[62] Neubert, P., and Protzel, P. **Superpixel benchmark and comparison**. In *Proc. Forum Bildverarbeitung* (2012), vol. 6.

[63] Sun, Y., Azorin-Peris, V., Kalawsky, R., Hu, S., Papin, C., and Greenwald, S. E. **Use of ambient light in remote photoplethysmographic systems : comparison between a high-performance camera and a low-cost webcam**. *Journal of biomedical optics 17*, 3 (2012). International Society for Optics and Photonics.

[64] Tsouri, G. R., Kyal, S., Dianat, S. A., and Mestha, L. K. **Constrained independent component analysis approach to nonobtrusive pulse rate measurements**. *Journal of biomedical optics 17*, 7 (2012), 77. International Society for Optics and Photonics.

[65] Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., and Van Gool, L. **Seeds : Superpixels extracted via energy-driven sampling**. In *European conference on computer vision* (2012), pp. 13–26. Springer.

[66] Zhang, Y., d'Aspremont, A., and El Ghaoui, L. **Sparse pca : Convex relaxations, algorithms and applications**. 915–940. Springer.

[67] Aarts, L. A., Jeanne, V., Cleary, J. P., Lieber, C., Nelson, J. S., Oetomo, S. B., and Verkruysse, W. **Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study**. *Early human development 89*, 12 (2013), 943–948. Elsevier.

[68] Bousefsaf, F., Maaoui, C., and Pruski, A. **Remote assessment of the heart rate variability to detect mental stress**. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on* (2013), pp. 348–351. IEEE.

[69] Chan, E. D., Chan, M. M., and Chan, M. M. **Pulse oximetry : understanding its basic principles facilitates appreciation of its limitations**. *Respiratory medicine 107*, 6 (2013), 789–799. Elsevier.

[70] Chang, J., Wei, D., and Fisher, J. W. **A video representation using temporal superpixels**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), 2051–2058.

[71] De Haan, G., and Jeanne, V. **Robust pulse rate from chrominance-based rppg**. *IEEE Transactions on Biomedical Engineering 60*, 10 (2013), 2878–2886. IEEE.

[72] Elhamifar, E., and Vidal, R. **Sparse subspace clustering : Algorithm, theory, and applications**. *IEEE transactions on pattern analysis and machine intelligence 35*, 11 (2013), 2765–2781. IEEE.

[73] Gibert, G., D'Alessandro, D., and Lance, F. **Face detection method based on photoplethysmography**. In *Advanced Video and Signal Based Surveillance (AVSS), 10th IEEE International Conference* (2013), p. 449–453. IEEE.

[74] Ibraheem, N. A., Khan, R. Z., and Hasan, M. M. **Comparative study of skin color based segmentation techniques**. *International Journal of Applied Information Systems (IJAIS) 5*, 10 (2013).

[75] Park, S., Ko, D., Whang, M., and Lee, E. C. **Vision based body dither measurement for estimating human emotion parameters**. In *International Conference on Human-Computer Interaction* (2013), pp. 346–352. Springer.

[76] Reso, M., Jachalsky, J., Rosenhahn, B., and Ostermann, J. **Temporally consistent superpixels**. In *Proceedings of the IEEE International Conference on Computer Vision* (2013), pp. 385–392. IEEE.

[77] Allen, J., and Howell, K. **Microvascular imaging : techniques and opportunities for clinical physiological measurements**. *Physiological measurement 35*, 7 (2014), R91. IOP Publishing.

[78] Apvrille, L., Tanzi, T., and Dugelay, J.-L. **Autonomous drones for assisting rescue services within the context of natural disasters**. In *General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI* (2014), pp. 1–4. IEEE.

[79] de Freitas Pereira, T., Komulainen, J., Anjos, A., De Martino, J. M., Hadid, A., Pietikäinen, M., and Marcel, S. **Face liveness detection using dynamic texture**. *EURASIP Journal on Image and Video Processing 2014*, 1 (2014), 2. Springer.

[80] De Haan, G., and Van Leest, A. **Improved motion robustness of remote-ppg by using the blood volume pulse signature**. *Physiological measurement 35*, 9 (2014), 1913. IOP Publishing.

[81] Hadid, A. **Face biometrics under spoofing attacks : Vulnerabilities, countermeasures, open issues, and research directions**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014), pp. 113–118.

[82] Liu, Z., Zhang, X., Luo, S., and Le Meur, O. **Superpixel-based spatiotemporal saliency detection**. *IEEE transactions on circuits and systems for video technology 24*, 9 (2014), 1522–1540. IEEE.

[83] McDuff, D., Gontarek, S., and Picard, R. W. **Improvements in remote cardiopulmonary measurement using a five band digital camera**. *IEEE Transactions on Biomedical Engineering 61*, 10 (2014), 2593–2601. IEEE.

[84] Nitzan, M., Romem, A., and Koppel, R. **Pulse oximetry : fundamentals and technology update**. *Medical Devices (Auckland, NZ) 7* (2014), 231. Dove Press.

[85] Papageorgiou, A., and de Haan, G. **Adaptive gain tuning for robust remote pulse rate monitoring under changing light conditions**. Technische Universiteit Eindhoven.

[86] Tamura, T., Maeda, Y., Sekine, M., and Yoshida, M. **Wearable photoplethysmographic sensors—past and present**. *Electronics 3*, 2 (2014), 282–302. Multidisciplinary Digital Publishing Institute.

[87] Tarassenko, L., Villarroel, M., Guazzi, A., Jorge, J., Clifton, D., and Pugh, C. **Non-contact video-based vital sign monitoring using ambient light and auto-regressive models**. *Physiological measurement 35*, 5 (2014), 807. IOP Publishing.

[88] Tasli, H. E., Gudi, A., and Den Uyl, M. **Integrating remote ppg in facial expression analysis framework**. In *Proceedings of the 16th International Conference on Multimodal Interaction* (2014), pp. 74–75. ACM.

[89] Villarroel, M., Guazzi, A., Jorge, J., Davis, S., Watkinson, P., Green, G., Shenvi, A., McCormick, K., and Tarassenko, L. **Continuous non-contact vital sign monitoring in neonatal intensive care unit**. *Healthcare technology letters 1*, 3 (2014), 87–91. IET.

[90] Zhang, Z. **Heart rate monitoring from wrist-type photoplethysmographic (ppg) signals during intensive physical exercise**. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on* (2014), pp. 698–702. IEEE.

[91] Blackford, E. B., and Estepp, J. R. **Effects of frame rate and image resolution on pulse rate measured using multiple camera imaging photoplethysmography**. In *Medical Imaging 2015 : Biomedical Applications in Molecular, Structural, and Functional Imaging* (2015), vol. 9417, p. 94172D. International Society for Optics and Photonics.

[92] Choe, J., Chung, D., Schwichtenberg, A., and Delp, E. J. **Improving video-based resting heart rate estimation : A comparison of two methods**. In *Circuits and Systems (MWSCAS), 2015 IEEE 58th International Midwest Symposium on* (2015), pp. 1–4. IEEE.

[93] Chouchene, M., Sayadi, F. E., Bahri, H., Dubois, J., Miteran, J., and Atri, M. **Optimized parallel implementation of face detection based on gpu component**. *Microprocessors and Microsystems 39*, 6 (2015), 393–404. Elsevier.

[94] Cui, Y., Fu, C.-H., Hong, H., Zhang, Y., and Shu, F. **Non-contact time varying heart rate monitoring in exercise by video camera**. In *Wireless Communications & Signal Processing (WCSP), 2015 International Conference on* (2015), pp. 1–5. IEEE.

[95] Fan, X., and Wang, J. **Bayesheart : A probabilistic approach for robust, low-latency heart rate monitoring on camera phones**. In *20th Int. Conf. Intell. User Interfaces* (2015), pp. 405–416.

[96] Feng, L., Po, L.-M., Xu, X., Li, Y., and Ma, R. **Motion-resistant remote imaging photoplethysmography based on the optical properties of skin**. *IEEE Transactions on Circuits and Systems for Video Technology 25*, 5 (2015), 879–891. IEEE.

[97] Fernando, S., Wang, W., Kirenko, I., de Haan, G., Bambang Oetomo, S., Corporaal, H., and van Dalfsen, J. **Feasibility of contactless pulse rate monitoring of neonates using google glass**. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare* (2015), pp. 198–201. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[98] Guazzi, A. R., Villarroel, M., Jorge, J., Daly, J., Frise, M. C., Robbins, P. A., and Tarassenko, L. **Non-contact measurement of oxygen saturation with an rgb camera**. *Biomedical optics express 6*, 9 (2015), 3320–3338. Optical Society of America.

[99] Kamshilin, A. A., Nippolainen, E., Sidorov, I. S., Vasilev, P. V., Erofeev, N. P., Podolian, N. P., and Romashko, R. V. **A new look at the essence of the imaging photoplethysmography**. *Scientific reports 5* (2015), 10494. Nature Publishing Group.

[100] Kumar, M., Veeraraghavan, A., and Sabharwal, A. **Distanceppg : Robust non-contact vital signs monitoring using a camera**. *Biomedical optics express 6*, 5 (2015), 1565–1588. Optical Society of America.

[101] Kuo, J., Koppel, S., Charlton, J. L., and Rudin-Brown, C. M. **Evaluation of a video-based measure of driver heart rate**. *Journal of safety research 54* (2015), 55–e29. Elsevier.

[102] Kwon, S., Kim, J., Lee, D., and Park, K. **Roi analysis for remote photoplethysmography on facial video**. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (2015), pp. 4938–4941. IEEE.

[103] Ren, C. Y., Prisacariu, V. A., and Reid, I. D. **gSLICr : SLIC superpixels at over 250Hz**. *ArXiv e-prints* (2015).

[104] Sugita, N., Obara, K., Yoshizawa, M., Abe, M., Tanaka, A., and Homma, N. **Techniques for estimating blood pressure variation using video images**. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (2015), pp. 4218–4221. IEEE.

[105] Vicente, F., Huang, Z., Xiong, X., De la Torre, F., Zhang, W., and Levi, D. **Driver gaze tracking and eyes off the road detection system**. *IEEE Transactions on Intelligent Transportation Systems 16*, 4 (2015), 2014–2027. IEEE.

[106] Wang, W., Stuijk, S., and De Haan, G. **Unsupervised subject detection via remote ppg**. *IEEE Transactions on Biomedical Engineering 62*, 11 (2015), 2629–2637. IEEE.

[107] Wedekind, D., Malberg, H., Zaunseder, S., Gaetjen, F., Matschke, K., and Rasche, S. **Automated identification of cardiac signals after blind source separation for camera-based photoplethysmography**. In *Electronics and Nanotechnology (ELNANO), 2015 IEEE 35th International Conference on* (2015), pp. 422–427. IEEE.

[108] Blackford, E. B., Estepp, J. R., Piasecki, A. M., Bowers, M. A., and Klosterman, S. L. **Long-range non-contact imaging photoplethysmography : cardiac pulse wave sensing at a distance**. In *Optical Diagnostics and Sensing XVI : Toward Point-of-Care Diagnostics* (2016), vol. 9715, p. 971512. International Society for Optics and Photonics.

[109] Bobbia, S., Benezeth, Y., and Dubois, J. **Remote photoplethysmography based on implicit living skin tissue segmentation**. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (2016), pp. 361–365. IEEE.

[110] Boulkenafet, Z., Komulainen, J., and Hadid, A. **Face spoofing detection using colour texture analysis**. *IEEE Transactions on Information Forensics and Security 11*, 8 (2016), 1818–1830. IEEE.

[111] Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S. R., Kowdle, A., Escolano, S. O., Rhemann, C., Kim, D., Taylor, J., and others. **Fusion4d : Real-time performance capture of challenging scenes**. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 114. ACM.

[112] Hanfland, S., and Paul, M. **Video format dependency of ppgi signals**. In *Proceedings of the International Conference on Electrical Engineering* (2016).

[113] Huang, C.-R., Wang, W.-A., Lin, S.-Y., and Lin, Y.-Y. **Useq : Ultra-fast superpixel extraction via quantization**. In *Pattern Recognition (ICPR), 2016 23rd International Conference on* (2016), pp. 1965–1970. IEEE.

[114] Iozzia, L., Cerina, L., and Mainardi, L. **Relationships between heart-rate variability and pulse-rate variability obtained from video-ppg signal using zca**. *Physiological measurement 37*, 11 (2016), 1934. IOP Publishing.

[115] Jingjing, L., Ying, C., Cheng, Z., Hua, Y., and Li, Z. **Tracking using superpixel features**. In *Measuring Technology and Mechatronics Automation (ICMTMA), 2016 Eighth International Conference on* (2016), pp. 878–881. IEEE.

[116] Kumar, M., Suliburk, J., Veeraraghavan, A., and Sabharwal, A. **Pulsecam : High-resolution blood perfusion imaging using a camera and a pulse oximeter**. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (2016), pp. 3904–3909. IEEE.

[117] Liu, S., Yang, B., Yuen, P. C., and Zhao, G. **A 3d mask face anti-spoofing database with real world variations**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2016), pp. 100–106.

[118] Liu, S., Yuen, P. C., Zhang, S., and Zhao, G. **3d mask face anti-spoofing with remote photoplethysmography**. In *European Conference on Computer Vision* (2016), pp. 85–100. Springer.

[119] Rahman, H., Ahmed, M. U., Begum, S., and Funk, P. **Real time heart rate monitoring from facial rgb color video using webcam**. In *The 29th Annual Workshop of the Swedish Artificial Intelligence Society (SAIS), 2–3 June 2016, Malmö, Sweden* (2016), no. 129. Linköping University Electronic Press.

[120] Rasche, S., Trumpp, A., Waldow, T., Gaetjen, F., Plötze, K., Wedekind, D., Schmidt, M., Malberg, H., Matschke, K., and Zaunseder, S. **Camera-based photoplethysmography in critical care patients**. *Clinical hemorheology and microcirculation 64*, 1 (2016), 77–90. IOS press.

[121] Rumiński, J. **Reliability of pulse measurements in videoplethysmography**. *Metrology and Measurement Systems 23*, 3 (2016), 359–371. De Gruyter Open.

[122] Senouci, B., Charfi, I., Heyrman, B., Dubois, J., and Miteran, J. **Fast prototyping of a soc-based smart-camera : a real-time fall detection case study**. *Journal of Real-Time Image Processing 12*, 4 (2016), 649–662. Springer.

[123] Sun, Y., and Thakor, N. **Photoplethysmography revisited : from contact to noncontact, from point to imaging**. *IEEE Transactions on Biomedical Engineering 63*, 3 (2016), 463–477. IEEE.

[124] Tayibnapis, I. R., Koo, D.-Y., Choi, M.-K., and Kwon, S. **A novel driver fatigue monitoring using optical imaging of face on safe driving system**. In *Control, Electronics, Renewable Energy and Communications (ICCEREC), 2016 International Conference on* (2016), pp. 115–120. IEEE.

[125] Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J. F., and Sebe, N. **Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2396–2404.

[126] Wang, W., Stuijk, S., and De Haan, G. **A novel algorithm for remote photoplethysmography : Spatial subspace rotation.** *IEEE Trans. Biomed. Engineering 63*, 9 (2016), 1974–1984.

[127] Wu, B.-F., Chu, Y.-W., Huang, P.-W., Chung, M.-L., and Lin, T.-M. **A motion robust remote-ppg approach to driver's health state monitoring**. In *Asian Conference on Computer Vision* (2016), pp. 463–476. Springer.

[128] Yang, Y., Liu, C., Yu, H., Shao, D., Tsow, F., and Tao, N. **Motion robust remote photoplethysmography in cielab color space**. *Journal of biomedical optics 21*, 11 (2016), 117001. International Society for Optics and Photonics.

[129] Blackford, E. B., and Estepp, J. R. **A multispectral testbed for cardiovascular sensing using imaging photoplethysmography**. In *Optical Diagnostics and Sensing XVII : Toward Point-of-Care Diagnostics* (2017), vol. 10072, p. 100720R. International Society for Optics and Photonics.

[130] Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., and Dubois, J. **Unsupervised skin tissue segmentation for remote photoplethysmography**. *Pattern Recognition Letters* (2017). Elsevier.

[131] Csillik, O. **Fast segmentation and classification of very high resolution remote sensing data using slic superpixels**. *Remote Sensing 9*, 3 (2017), 243. Multidisciplinary Digital Publishing Institute.

[132] et al., S. C. **Isprs annals of the photogrammetry, remote sensing and spatial information sciences**. *Gottingen IV-2*, W3 (2017), 9–16.

[133] Kamshilin, A. A., and Margaryants, N. B. **Origin of photoplethysmographic waveform at green light**. *Physics Procedia 86* (2017), 72–80. Elsevier.

[134] Kevat, A. C., Bullen, D. V., Davis, P. G., and Kamlin, C. O. F. **A systematic review of novel technology for monitoring infant and newborn heart rate**. *Acta Paediatrica 106*, 5 (2017), 710–720. Wiley Online Library.

[135] Temko, A. **Accurate heart rate monitoring during physical exercises using ppg**. *IEEE Transactions on Biomedical Engineering 64*, 9 (2017), 2016–2024. IEEE.

[136] Wang, W., den Brinker, A. C., Stuijk, S., and de Haan, G. **Algorithmic principles of remote ppg**. *IEEE Transactions on Biomedical Engineering 64*, 7 (2017), 1479–1491. IEEE.

[137] Zhang, Q., Wu, Q., Zhou, Y., Wu, X., Ou, Y., and Zhou, H. **Webcam-based, non-contact, real-time measurement for the physiological parameters of drivers**. *Measurement 100* (2017), 311–321. Elsevier.

[138] Benezeth, Y., Li, P., Macwan, R., Nakamura, K., Gomez, R., and Yang, F. **Remote heart rate variability for emotional state monitoring**. In *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (2018).

[139] Bobbia, S., Luguern, D., Benezeth, Y., Nakamura, K., Gomez, R., and Dubois, J. **Real-time temporal superpixels for unsupervised remote photoplethysmography**. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2018), pp. 1341–1348.

[140] Macwan, R., Benezeth, Y., and Mansouri, A. **Remote photoplethysmography with constrained ica using periodicity and chrominance constraints**. *Biomedical engineering online 17*, 1 (2018), 22. BioMed Central.

[141] Marroquin, R., Dubois, J., and Nicolle, C. **Ontology for a panoptes building : exploiting contextual information and a smart camera network**. *Semantic Web Journal, Special Issue on Semantic Technologies and Interoperability in the Built Environment* (2018). IOS Press.

[142] Richard, M., Serge, B., Yannick, B., Julien, D., and Alamin, M. **Periodic variance maximization using generalized eigenvalue decomposition applied to remote photoplethysmography estimation**. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2018).

Table des figures

1.1 Exemple de système pour la mesure cardiaque. Dans un cadre médical (gauche) : oxymètre de pouls, et dans un cadre récréatif (droite) : montre connectée. 2

1.2 Exemple d'installation d'un dispositif ECG mobile *HOLTER*. 5

2.1 Représentation de l'interaction lumière - peau. 12

2.2 Forme caractéristique du signal PPG avec les phénomènes de diastole et de systole [84] pour une mesure PPG en transmitance. 13

2.3 Courbes d'absorption de l'oxyhémoglobine (HbO_2), de la déoxyhémoglobine (Hb), de la carboxyhémoglobine ($COHb$) et de la méthémoglobine ($MetHb$) en fonction de la longueur d'onde [69]. 14

2.4 Exemple d'amplitude AC, interprétable comme une mesure de qualité du signal, pour 4 sujets en fonction de la force au contact. Les cercles indiquent une amplitude maximale [28]. 16

2.5 Installation pour la mesure rPPG telle que présenté par Poh et al. [47]. 17

2.6 Moyenne des coefficients d'inter-corrélations de la pression artérielle systolique et des variations de phases PD_i [104]. 20

2.7 Dispositif expérimental utilisé pour déterminer la qualité de la réponse du signal rPPG pour l'estimation du rythme cardiaque et du rythme réspiratoire en fonction de la longueur d'onde [52]. 21

2.8 Énergie du signal cardiaque (gauche) et du signal respiratoire (droite) en comparaison à l'énergie du bruit mesurée en fonction de la longueur d'onde [52]. 21

2.9 Exemple de l'estimation de la qualité de la réponse rPPG pour les 3 différents canaux R, G et B d'une caméra polychromatique UI-2220SE-C, vis à vis de la réponse en réflectance de la peau, de la courbe de réponse établie Figure 2.8 et avec une source de lumière halogène [80]. La convolution des différentes courbes (à gauche) permet d'établir l'amplitude normalisée du signal rPPG en fonction des canaux de la caméra utilisé pour la mesure (à droite). 22

3.1 Le flux vidéo d'entrée est segmenté afin d'isoler le visage du sujet, les pixels sont moyennés dans les trois canaux afin de construire les signaux $x_i(t)$ (*e.g.* les observations). La méthode ICA permet d'extraire autant de signaux de sortie qu'il y a d'observations, ici trois. Le signal rPPG est alors contenu dans l'un des trois signaux [47]. 28

3.2 Schéma fonctionnel pour la mesure rPPG tel que réalisée dans la littérature. En encadré, les fonctions de la chaîne du processus de mesure rPPG, et en flottant les informations issues des différentes fonctions. 34

3.3 Segmentation de la zone d'intérêt pour la mesure rPPG. Détection et segmentation du visage puis segmentation de la peau. 35

4.1 Qualité du signal rPPG pour chaque superpixel temporel avec de gauche à droite : image d'entrée, segmentation en superpixels temporels et qualité du signal rPPG (la qualité de la mesure est meilleur dans les tons jaunes/oranges, elle est exprimée en dB). 43

4.2 Les vecteurs propres similaires entre eux en énergie et pour les différents niveaux de segmentation sont fusionnés afin de déterminer la zone d'intérêt finale [106]. . 45

4.3 Schéma fonctionnel de la méthode proposée. (1) les images de la vidéo en entrée sont segmentées en superpixels temporels. (2) les signaux rPPG *potentiels* $p^i(t)$ sont extraits depuis les multiples signaux $C^i(t)$ fournis par les superpixels. (3) Une mesure de la qualité du signal est effectuée afin de déterminer les superpixels qui contribuent le mieux à l'information rPPG, les poids w_i sont définis à partir de la mesure de *SNR*. (4) Une moyenne pondérée des signaux est effectuée afin d'obtenir un signal final. 47

4.4 Exemple de décomposition fréquentielle pour des signaux rPPG *potentiels* respectivement issus d'un TSP de peau et d'un TSP de l'arrière-plan. Avec en rouge l'information utile du signal et en bleu le bruit. 49

4.5 Exemples de valeur de SNR en dB et leurs superpixels temporels correspondants. 50

4.6 Forme caractéristique du signal rPPG [95]. 52

4.7 Etats du modèle de Markov pour une modélisation en 2 ou 4 états [95]. Chaque état correspond à une étape du cycle cardiaque. 53

4.8 Observations possibles sur le signal mesuré. 53

4.9 Etats et transitions possibles pour les deux modèles de Markov caché utilisés pour la mesure rPPG [95]. 54

4.10 Segmentation pour deux résolutions de superpixels temporels différentes $K = 500$ et $K = 1000$ avec correspondance entre la valeur de *SNR* et *BIC*. Plus la teinte tend vers le jaune plus la qualité du signal est élevée. 56

5.1 Exemple d'images issues des vidéos constituant l'ensemble de données *dataset simple* de la base UBFC-rPPG. 58

5.2 Exemple d'images issues des vidéos constituant l'ensemble de données *dataset réaliste*. 59

5.3 Exemple de zones d'intérêt déterminées par les trois méthodes de segmentation *face* en ligne 1, *crop* en ligne 2 et *skin* en ligne 3. 60

5.4 Exemple de segmentation du visage et de la peau pour la mesure rPPG depuis le *dataset simple*. 62

5.5 Précision de la mesure rPPG avec une erreur maximale de 2.5 bpm et 5 bpm pour une résolution en superpixels temporels variant de 50 à 1000 TSP. 63

5.6 Corrélation entre les données issues de la mesure rPPG et de la vérité terrain. Avec à gauche les résultats de notre méthode et à droite les résultats de la méthode de référence. L'abscisse représente les données de la mesure rPPG et en ordonnée la vérité terrain. 63

5.7 Représentation de Bland-Altman pour toutes les estimations issues de toutes les vidéos. Avec à gauche les résultats de notre méthode et à droite les résultats de la méthode de référence. L'abscisse représente la vérité terrain et en ordonnée la différence entre le rythme cardiaque HR depuis le capteur au doigt (HR_{PPG}) et le rythme cardiaque HR estimé par la mesure rPPG (HR_{rPPG}). 64

5.8 Valeur de *SNR* calculé pour chaque TSP, pour des valeurs de *SNR* comprises dans la bande [0; 8] dB. La première ligne montre l'image d'entrée et sa décomposition en superpixels temporels. La seconde ligne montre les valeurs de *SNR* pour les méthodes de rPPG (de gauche à droite) : $Green$, $Green - Red$, PCA, $CHROM$ et POS. 66

5.9 Résultats des métriques : corrélation et précisions pour une erreur maximale de 2.5 bpm et 5 bpm pour un nombre de TSP variant de 100 à 600. 67

5.10 Résultats des métriques : erreur quadratique (RMSE) et SNR moyen pour un nombre de TSP variant de 100 à 600. 68

5.11 Métrique *SNR* appliquée à la segmentation en TSP. La première ligne montre la segmentation en TSP tandis que la ligne 2 montre la correspondance avec la valeur de *SNR* et la ligne 3 montre les valeurs de *SNR* en fonction des différents TSP. Pour toutes les lignes, de gauche à droite, les résolutions en superpixels varient avec 100, 150, 200, 400 et 600 TSP. 68

5.12 Exemples d'erreurs dans le suivi temporel des superpixels de la méthode TSP. La première colonne montre les images d'entrée, la seconde la segmentation en TSP et la troisième les superpixels qui ont été remis à zéro (en jaune) dans la vidéo. . 69

5.13 Pour chacune des vidéos de l'ensemble de données *dataset simple* : variations des valeurs du seuil x en fonction de la résolution en superpixels temporels. 71

5.14 Segmentation en TSP pour la métrique de qualité *SNR* à gauche et la métrique de qualité BIC à droite. 73

6.1 Exemple de segmentation par superpixels avec la méthode IBIS avec à gauche la segmentation initiale pour initialiser la méthode et à droite la segmentation finale. 79

6.2 Exemple de segmentation par superpixels avec la méthode IBIS avec à gauche la segmentation finale superposée à l'image d'entrée et à droite l'identification en jaune des pixels de l'image utilisés pour calculer la segmentation. 80

6.3 Exemple d'itérations dans le déroulement de la méthode IBIS avec un bloc unique. Les zones cohérentes sont éliminées du processus afin de réduire la quantité de pixels utiles à l'identification des frontières entre les superpixels. 82

6.4 Schéma fonctionnel de la méthode IBIS. (1) Les *seeds* sont initialisées avec une décomposition en grille fixe. (2) La méthode IBIS est appliquée et les itérations permettent d'affiner l'identification des frontières et (3) le post-traitement permet de s'assurer de la contiguïté et de la taille minimale des superpixels. 82

6.5 Les blocs de plus haut niveau de taille τ_0 ne se chevauchent pas, ainsi ils ne partagent pas de pixels en commun, cela permet une meilleure optimisation de la répartition des opérations en prévision d'une implémentation sur une plateforme avec plusieurs cœurs de calculs. 83

6.6 Évolution de la valeur de la différence absolue entre les valeurs des *seeds* aux différentes itérations et la valeur finale en fonction du nombre de superpixels défini par l'utilisateur pour les méthodes *SLIC* à gauche et *IBIS* à droite. 84

6.7 Exemple dans l'application des blocs de plus haut niveau avec $\alpha = 2$, les blocs se chevauchent et les pixels calculés par chacun se limite à la bordure droite inférieure. Ainsi le premier bloc a son origine (en haut à gauche) en coordonnées $(-1; -1)$ dans l'image. 85

6.8 Pourcentage des pixels qui sont nécessaires à la segmentation en superpixels avec la méthode *IBIS* en fonction de la résolution en superpixels pour l'implémentation initiale (en bleu), l'implémentation optimisée (en rouge) et la différence des deux (jaune). 86

6.9 Evaluation du nombre de pixels attribués en fonction de la taille du bloc définit par l'itération notée α. À gauche les courbes, avec une échelle logarithmique pour mieux visualiser la progression, et à droite le gain en pourcentage défini par le nombre de pixels économisés dans l'attribution d'un bloc de taille maximale τ_0 en fonction de α. 87

6.10 Exemple d'images de l'ensemble de données *BSD500*. 89

6.11 Variation du pourcentage de pixels à allouer individuellement en fonction du nombre de superpixels dans l'image sur l'ensemble de données *BSD500* pour une résolution en entrée de 240x320 pixels. 90

6.12 Variations du pourcentage de pixels à allouer individuellement en fonction du nombre de superpixels dans l'image et du nombre d'itérations autorisées. 91

6.13 Évolution du temps de calcul de la méthode IBIS en fonction de la résolution en pixels de l'image. En rouge, la progression linéaire de la quantité de pixels dans l'image et en jaune la progression du temps de calcul pour la méthode *IBIS* avec notre implémentation optimisée. 92

6.14 Évolution du coefficient de proportionnalité (*i.e.* rapport des deux courbes de la figure 6.13) de la méthode *IBIS* comparé à une évolution linéaire du nombre de pixels. 92

6.15 Schéma logique de la propagation temporelle pour la segmentation en superpixels temporels *IBIS* telle qu'elle fut développée pour la mesure rPPG dans [139]. . . . 94

6.16 Hiérarchie des blocs utilisés dans l'implémentation initiale de la méthode *IBIS* sur CPU. La parallélisation de l'algorithme s'effectue au niveau des blocs de plus haut niveau. 95

6.17 Algorithme propre à tous les blocs de niveau *i* pour une exécution à l'itération *j*. . 95

6.18 Répartition des temps de calcul en fonction de la résolution en superpixels dans l'image. A gauche, la répartition globale moyennée sur toutes les résolutions en superpixels et à droite l'évolution de la répartition des temps de calcul en fonction de la quantité de superpixels pour 50, 100, 250, 500, 1000 et 2500 superpixels. . 96

6.19 Comparaison de *IBIS* et *IBIS* optimisée. À gauche le *Boundary Recall* et à droite la métrique *Undersegmentation Error*. Les métriques sont calculées pour les résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et pour une compacité de 20. 97

6.20 Résultats pour la qualité de la segmentation par superpixels pour les méthodes *SLIC*, *SEEDS*, *IBIS* et *USEQ*. Les métriques sont calculées pour les résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et pour une compacité de 20 pour les méthodes qui utilisent ce paramètre. 98

6.21 Temps de calcul obtenus pour la segmentation en superpixels des méthodes *SLIC*, *SEEDS*, *IBIS* et *USEQ*. L'exécution est réalisée sur 1 cœur de calcul à gauche et à droite sur les 4 disponibles sur le processeur i7-4790 utilisé pour l'expérience. . . 98

6.22 Temps de calcul relatifs à la méthode *IBIS* comme référence. 99

6.23 Comparaison qualitative pour les méthodes de segmentation par superpixels *IBIS*, *SLIC*, *USEQ* et *SEEDS*. Chaque image est séparée en deux régions qui sont segmentées avec à gauche 250 superpixels et à droite 500 superpixels. 100

7.1 Représentation de la répartition des *threads* au sein des *blocs* eux mêmes répartis dans la *grille* [2]. 105

7.2 Les différents types de mémoire accessibles suivant le niveau d'abstraction dans la géométrie des *threads* CUDA [2]. 106

7.3 Répartition du temps de calcul nécessaire à la segmentation en superpixels avec la méthode *IBIS* en fonction du type d'opérations effectuées. 107

7.4 Évolution des temps de calcul des méthodes *IBIS* et *gSLICr* en fonction de la résolution en pixels de l'image. En rouge, la progression de la quantité de pixels dans l'image (qui représente la complexité linéaire), en bleu la progression du temps de calcul pour la méthode *gSLICr* et en jaune la progression du temps de calcul pour la méthode *IBIS* avec notre implémentation optimisée. 109

7.5 Évolution des temps de calcul des méthodes *IBIS* et *gSLICr*, sur la plateforme GPU *Jetson TX1*, en fonction de la résolution en pixels de l'image et pour les résolutions en superpixels 50, 100, 250, 500, 1000 et 2500. A gauche la méthode *IBIScuda* et à droite la méthode *gSLICr*. 110

7.6 Évolution des temps de calcul des méthodes *IBIScuda* et *gSLICr*, sur la plateforme GPU *GTX 1080 ti*, en fonction de la résolution en pixels de l'image et pour les résolutions en superpixels 50, 100, 250, 500, 1000 et 2500. A gauche la méthode *IBIScuda* et à droite la méthode *gSLICr*. 110

7.7 Évolution des temps de calcul des méthodes *IBIScuda* et *gSLICr* en fonction de la résolution en pixels de l'image. A gauche sur la plateforme GPU *Jetson TX1* et à droite sur la plateforme GPU *GTX 1080 ti*. 111

7.8 Évolution de la qualité de la segmentation pour les métriques *Boundary recall* et *Undersegmentation error* sur l'ensemble de données *BSD500* pour des résolutions en superpixels de 50, 100, 250, 500, 1000 et 2500 et avec un nombre d'itérations limité à 1, 2, 3 et non limité avec $\alpha = 3$ (soit 4 itérations). 113

7.9 Évolution des temps de calcul des méthodes *gSLICr*, *IBIScuda* et *IBIScuda* avec α fixé à 0, en fonction de la résolution en pixels moyennée sur les résolutions en superpixels 50, 100, 250, 500, 1000, 2500. A gauche sur la plateforme GPU *Jetson TX1* et à droite sur la plateforme GPU *GTX 1080 ti*. 114

7.10 Rapport des temps de calcul pour la méthode *IBIScuda* et la méthode *IBIScuda* avec α fixé à 0. En rouge sur la plateforme GPU *Jetson TX1* et en bleu sur la plateforme *GTX 1080 ti*. 114

7.11 Répartition du temps de calcul pour la méthode *IBIScuda* sur la plateforme GPU *Jetson TX1* en fonction du nombre d'itérations et moyenné pour les résolutions en superpixels de 50 à 2500 et pour des résolutions d'images de 250x250 pixels à 4000x4000. A gauche le nombre d'itérations non limité et à droite α est fixé à 0. . 115

7.12 Répartition du temps de calcul pour la méthode *IBIScuda* sur la plateforme GPU *GTX 1080 ti* en fonction du nombre d'itérations et moyenné pour les résolutions en superpixels de 50 à 2500 et pour des résolutions d'images de 250x250 pixels à 4000x4000. A gauche le nombre d'itérations non limité et à droite α est fixé à 0. . 116

7.13 Fréquence image pour la segmentation en superpixels pour les méthodes *IBIS*, *gSLICr* et *IBIS* avec α fixé à 0 en fonction du nombre de pixels dans l'image. A gauche pour la plateforme GPU *Jetson TX1* et à droite pour la plateforme GPU *GTX 1080 ti*. 117

8.1 Résumé des travaux de thèse en fonction des étapes de la mesure rPPG. De gauche à droite : segmentation en superpixels temporels en temps réel, métriques de qualité pour les signaux issus des méthodes de rPPG et fusion pondérée des signaux rPPG afin de maximiser la qualité de l'estimation du rythme cardiaque. 121

Liste des tables

5.1 Résumé de l'ensemble des résultats obtenus dans la comparaison de la méthode proposée avec deux résolutions de superpixels temporels N=200 et N=400 et de la méthode de référence notée *skin*. 64

5.2 Comparaison des méthodes de segmentation $FACE$, $CROP$ et $SKIN$ décrites en section 5.1.2 appliquées aux 5 méthodes de rPPG : $Green$, $Green - Red$, PCA, $CHROM$ et POS décrites en section 3.2. Les métriques (présentées en section 5.1.3) sont calculées pour une résolution en superpixels de 150 TSP. 65

5.3 Comparaison des temps de calcul des étapes de prétraitement et du calcul des méthodes de rPPG sur un signal de 20 secondes. 66

5.4 Résultat pour l'estimation du rythme cardiaque avec la métrique de qualité BIC pour la vraisemblance avec les modèles de cachés de Markov en comparaison avec la métrique de qualité SNR. 72

6.1 Résultats pour l'estimation du rythme cardiaque avec la métrique de qualité SNR pour les méthodes *IBIS* et *IBIS adjacent* en comparaison avec la méthode de référence évaluée au chapitre 5. 101

6.2 Résultats pour l'estimation du rythme cardiaque avec la métrique de qualité BIC pour les méthodes *IBIS* et *IBIS adjacent* en comparaison avec la méthode de référence évaluée au chapitre 5. 101

7.1 Spécifications pour les cartes de développement *Jetson TX1* et *Jetson TX2*. L'intégration sur GPU a été réalisée sur la plateforme *Jetson TX1*. 104

Liste des définitions

rPPG : est l'acronyme de *Remote Photoplethysmography* qui est l'appellation anglaise pour la mesure de photopléthysmographie sans contacts.

SpO_2 : défini la concentration en oxygène dans le sang.

superpixels : défini un groupe de pixels homogènes, dont les caractéristique spatiales et chromatiques sont similaires.

voxels : défini un groupe de pixels homogènes dans le temps, dont les caractéristique spatiales et chromatiques sont similaires dans le temps.

SNR : est l'acronyme de *Signal on Noise Ratio* qui est l'appellation anglaise du rapport signal sur bruit tel qu'il est défini dans la science du traitement du signal et permet la mesure de la qualité d'un signal.

BIC : est l'acronyme de *Bayesian Information Criterion* qui est l'appellation anglaise du critère d'informations bayésien ou critère de vraisemblance et permet une estimation de la vraisemblance entre un jeu d'observations et un modèle donné.

HMM : est l'acronyme de *Hidden Markov Model* qui est l'appellation anglaise des modèles de Markov caché.

GPU : est l'acronyme de *Graphical Processor Unit* qui est l'appellation anglaise des processeurs graphiques qui sont des unités matérielle dédié au traitement d'image et plus généralement aux calculs sur les matrices.

Document réalisé avec LaTeX et :
le style LaTeX pour Thèse de Doctorat créé par S. Galland — http://www.multiagent.fr/ThesisStyle
la collection de paquets `tex-upmethodology` — http://www.arakhne.org/tex-upmethodology/