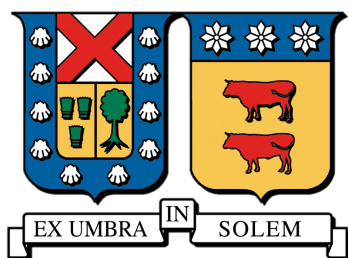


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



**COMPENSACIÓN DE CAMBIOS DE
ILUMINACIÓN PARA ESTIMACIÓN DE
SIGNOS VITALES MEDIANTE
FOTOPLETISMOGRAFÍA**

FELIPE IGNACIO SÁNCHEZ PIZARRO

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
ELECTRÓNICO**

PROFESOR GUÍA: MARCOS ZÚÑIGA BARRAZA.
PROFESOR CORREFERENTE: NICOLÁS JARA.

ABRIL - 2024

AGRADECIMIENTOS

En primer lugar, quiero expresar mi sincero agradecimiento a mis padres por su apoyo inquebrantable a lo largo de estos años. Han estado presentes en cada etapa de mi vida, ofreciéndome su amor y aliento en los momentos difíciles y compartiendo mi alegría en los momentos de felicidad. También agradezco a mis padrinos por ser una presencia constante en mi vida, por su apoyo continuo y su fe en mí.

Quiero expresar mi sincero agradecimiento a mis amigos y amigas. A pesar de la distancia, siempre estuvieron presentes para apoyarme y brindarme su amistad. Les agradezco a todos por su apoyo. Ustedes saben quiénes son.

Quiero expresar mi gratitud a mi profesor guía por brindarme la oportunidad de explorar un tema tan fascinante. Su constante apoyo, orientación y motivación fueron fundamentales para lograr este trabajo.

Resumen

La Fotopletismografía remota (rPPG) es una técnica utilizada para identificar cambios de intensidad de color en secuencias de imágenes a lo largo del tiempo, con el fin de crear señales de información condensada del rostro de una persona para su posterior análisis y obtención de signos vitales. En el contexto de la telemedicina, donde controlar los eventos y las perturbaciones lumínicas es un desafío, se requiere una metodología robusta para obtener estimaciones precisas de los signos vitales mediante la rPPG.

Para alcanzar este objetivo, se desarrolla una metodología de compensación de perturbaciones, junto con un método robusto para estimar el Heart Rate (HR). Estas metodologías se combinan para llevar a cabo un análisis espacio-temporal de la escena, segmentando la Región de Interés (ROI) del rostro en varias sub-ROIs, así como áreas específicas del fondo. Esto facilita el análisis del comportamiento de la iluminación y las perturbaciones en el rostro, además de compensar las perturbaciones temporales que afectan en común al rostro y al fondo.

La metodología desarrollada se comparó con trabajos previos, desde el estado del arte, que abordan la compensación de perturbaciones temporales y con métodos tradicionales de estimación de HR que se basan en la recuperación de la señal de Blood Volume Pulse (BVP) a partir de la señal rPPG. En la evaluación con tres Dataset diferentes, se destacaron dos metodologías: la propuesta en este trabajo y CHROM adaptado con una metodología específica para la estimación de HR.

Se comparan y analizan los dos mejores métodos mencionados, evaluando su combinación. Se revelan ventajas y limitaciones. CHROM resalta componentes no deseados al no considerar la dinámica en la escena. La metodología desarrollada se adapta a cada escena pero presenta debilidades ante perturbaciones de iluminación en el rostro debido al movimiento del sujeto. Además, el desempeño general de los dos métodos es similar, con la metodología desarrollada mostrando mejores resultados en los rangos bajo y normal de HR, mientras que CHROM destaca en el rango alto de HR, aunque con una diferencia mínima en comparación con la metodología desarrollada. Por lo tanto, no es concluyente elegir sólo uno de ellos.

Abstract

Remote photoplethysmography (rPPG) is a technique used to identify color intensity changes in image sequences over time, in order to create condensed information signals from a person's face for further analysis and vital sign acquisition. In the context of telemedicine, where monitoring events and light disturbances is a challenge, a robust methodology is required to obtain accurate estimates of vital signs using rPPG.

To achieve this goal, a perturbation compensation methodology is developed, along with a robust method to estimate the Heart Rate (HR). These methodologies are combined to perform a spatio-temporal analysis of the scene, segmenting the Region of Interest (ROI) of the face into several sub-ROIs, as well as specific areas of the background. This facilitates the analysis of illumination behavior and disturbances on the face, as well as compensating for temporal disturbances that affect both the face and the background.

The methodology developed was compared with previous works, from the state of the art, that address the compensation of temporal disturbances and with traditional methods of HR estimation based on the recovery of the Blood Volume Pulse (BVP) signal from the rPPG signal. In the evaluation with three different Datasets, two methodologies stood out: the one proposed in this work and CHROM adapted with a specific methodology for HR estimation.

The two best methods mentioned above are compared and analyzed, evaluating their combination. Advantages and limitations are revealed. CHROM highlights unwanted components by not considering the dynamics in the scene. The developed methodology adapts to each scene but presents weaknesses in the face illumination perturbations due to subject motion. Furthermore, the overall performance of the two methods is similar, with the developed methodology showing better results in the low and normal HR ranges, while CHROM excels in the high HR range, albeit with a minimal difference compared to the developed methodology. Therefore, it is not conclusive to choose only one of them.

Índice general

1. Introducción	1
1.1. Contexto	1
1.2. Planteamiento del problema	2
1.3. Hipótesis	2
1.4. Acercamiento a la solución	2
1.5. Objetivos	3
1.6. Estructura del Documento	3
2. Estado del arte	5
2.1. Marco Teórico	5
2.2. Alternativas de solución	13
2.3. Discusión	26
3. Diseño de la solución	28
3.1. Desarrollo de la primera metodología exploratoria	28
3.2. Desarrollo de la segunda metodología exploratoria	30
3.3. Desarrollo de la tercera metodología exploratoria	32
3.4. Desarrollo de la cuarta metodología exploratoria	45
4. Validación	55
4.1. Análisis cualitativo	60
4.2. Análisis cuantitativo	83
4.3. Evaluación	91

Capítulo 1

Introducción

1.1. Contexto

La creciente incorporación de la tecnología en todos los ámbitos de la vida ha influido significativamente en la forma en que las personas llevan a cabo sus actividades diarias. Esta influencia se refleja en un desarrollo integral que abarca una gran variedad de campos, que van desde la creación de arte digital y asistentes virtuales personalizados, hasta aspectos de desarrollo complejos como máquinas autónomas que intentan replicar tareas humanas.

En este contexto, las nuevas tecnologías aplicadas en el área de la salud han experimentado un crecimiento exponencial en los últimos años. Esto ha dado lugar a diversas aplicaciones que van desde la creación de robots con la capacidad de operar en pacientes de manera autónoma, hasta el desarrollo de diversas aplicaciones en el campo de la prevención y detección temprana de enfermedades.

Bajo esta línea de implementación tecnológica robusta en el área de la salud, en los últimos años se han realizado esfuerzos para diversificar y poner al alcance de todo público herramientas de monitoreo y análisis remoto sin necesidad de requerir implementación técnica especializada. Uno de estos avances se refiere a la medición de signos vitales de forma remota utilizando un dispositivo que prácticamente todas las personas en esta era poseen: un smartphone. Este proceso se lleva a cabo mediante la cámara que tienen integrada estos dispositivos, la cual graba un segmento de video del rostro de aproximadamente 20 a 30 segundos y posteriormente calcula la señal de Fotopletismografía generada, en su mayoría, a partir de las mejillas.

El problema principal que se debe abordar para que la medición de signos vitales en contextos no estructurados como la telemedicina, es la iluminación. Los cambios repentinos, distintas fuentes de luz y la interferencia de armónicos que ocurren en la captura de los videos son ejemplos de problemas inherentes en los distintos entornos donde se desarrollará la sesión de telemedicina, los cuales no se pueden controlar, pero se estima

que pueden compensarse.

1.2. Planteamiento del problema

La Fotopletismografía remota (rPPG) es una técnica utilizada para identificar cambios de intensidad de color en secuencias de imágenes a lo largo del tiempo, con el fin de crear señales de información condensada del rostro de una persona para su posterior análisis y obtención de signos vitales. En el contexto de las sesiones de telemedicina, resulta desafiante controlar por completo los eventos que suceden en la escena. Por consiguiente, se hace imperativo establecer protocolos que regulen el comportamiento durante estas sesiones, al mismo tiempo que se abordan las perturbaciones de iluminación que puedan impactar tanto en el rostro como en el fondo. Esta gestión es fundamental para obtener una estimación precisa de los signos vitales utilizando rPPG en el rostro. Por ende, se plantea la necesidad de desarrollar una metodología robusta para la compensación de estas perturbaciones de iluminación en la escena, con el fin de lograr una estimación precisa de los signos vitales requeridos.

1.3. Hipótesis

Se plantea la siguiente Hipótesis: ***"Es posible compensar los efectos de la iluminación global en la escena para obtener de forma confiable una señal de Fotopletismografía remota en el contexto de sesiones de telemedicina."***

Preguntas de trabajo:

- ¿Este proceso de compensación se puede hacer de forma rápida?
- ¿Se puede obtener una alta precisión en la inferencia de signos vitales con esta nueva metodología?
- ¿Cuáles son los escenarios en donde se maximiza la precisión?

1.4. Acercamiento a la solución

El enfoque propuesto en esta memoria presenta un acercamiento hacia la solución del problema de compensación de perturbaciones de iluminación de manera espacio-temporal. Este enfoque se implementa como un bloque de pre-procesamiento con el objetivo de restaurar las señales libres de perturbaciones, lo que facilita la identificación clara de la información relevante sobre los signos vitales por parte de los algoritmos. Para alcanzar este propósito, se emplea una estrategia que implica el uso de múltiples

sub-ROIs, las cuales son segmentaciones más pequeñas de la ROI general, que se enfoca en el rostro. Estas sub-ROIs se utilizan para caracterizar de forma espacio-temporal la información del rostro que contiene información sobre los signos vitales y perturbaciones tanto locales como globales de la escena. Además, se extrae información de manera similar desde el fondo de la escena con el fin de establecer comparaciones respecto a la caracterización espacio-temporal del rostro. Estas comparaciones son para establecer si hay algún grado de similitud que deba ser compensado y de esta manera es posible recuperar la señal de rPPG en un entorno con presencia de cambios de iluminación para distintos tonos de piel.

1.5. Objetivos

Objetivo General: Analizar y compensar los cambios de iluminación, con el fin de aplicar técnicas de Fotoplethysmografía en ambientes de iluminación no controlada.

Objetivos específicos:

- Proponer una metodología de compensación de iluminación, identificando el rango de funcionamiento de la metodología propuesta junto con la PPG.
- Minimizar el error de cálculo de Heart Rate (HR), Heart Rate Variability (HRV) y Oxigenación en sangre (SpO₂).
- Protocolizar el uso y toma de datos de video en una sesión de telemedicina para la estimación óptima de los signos vitales mediante PPG.
- Cumplir con la restricción de tiempo real para la estimación de signos vitales, con la incorporación de la metodología de compensación de iluminación propuesta.

1.6. Estructura del Documento

- Capítulo 2, Estado del arte: Se realiza un estudio de las técnicas y metodologías actuales que más se acerquen a la resolución del problema de compensación de perturbaciones de iluminación. Para acotar el problema, se adoptan dos enfoques: compensación temporal y compensación espacial.
- Capítulo 3, Desarrollo de la solución: Con ideas de los artículos analizados en el estado del arte, se desarrollan cuatro metodologías exploratorias para abordar la problemática principal de compensación de perturbaciones de iluminación con un enfoque espacio-temporal. La primera y segunda metodología exploratoria, corresponden a los métodos adaptados desde los artículos correspondientes,

enfoque espacial y temporal, mientras que la tercera y cuarta metodología, corresponden a los algoritmos propuestos de compensación de perturbaciones con enfoque espacio-temporal y a la detección robusta de HR.

- Capítulo 4, Validación: Se realiza la validación de las metodologías exploratorias del Capítulo 3. Para caracterizar completamente las metodologías propuestas, se realiza un análisis cualitativo y cuantitativo, por Dataset y en conjunto, para obtener una caracterización general del funcionamiento.
- Conclusiones: se expone un resumen generalizado donde se menciona lo desarrollado en este trabajo y la metodología propuesta juntos con sus limitaciones y un análisis general respecto a los mejores métodos seleccionados.

Capítulo 2

Estado del arte

2.1. Marco Teórico

2.1.1. Fotopletismografía (PPG)

La PPG es una técnica óptica no invasiva que se aplica sobre la piel para detectar los cambios volumétricos de la sangre en las venas. Se realiza utilizando luces infrarrojas o LEDs junto con sensores fotodetectores, que registran las fluctuaciones de luz absorbida por la hemoglobina a lo largo del tiempo.

2.1.2. Fotopletismografía remota (rPPG)

El objetivo de la rPPG es el mismo que el de la PPG: detectar los cambios volumétricos de la sangre a lo largo del tiempo. Sin embargo, la forma en que se capturan estos cambios difiere. En el caso de la rPPG, se utilizan imágenes secuenciales, típicamente un video en el espacio de color RGB, para identificar las Regiones de Interés (ROI), las cuales suelen ubicarse en las mejillas y la frente. Posteriormente, se realiza un promedio de las intensidades de color en cada fotograma para construir una señal en el tiempo que refleje las fluctuaciones volumétricas de la sangre en el rostro.

Esta técnica se basa en el principio de que los cambios en el flujo sanguíneo en la superficie de la piel afectan la absorción y reflexión de la luz visible, lo que se traduce en variaciones en la intensidad del color capturado por la cámara. Estas variaciones son tan pequeñas que cualquier perturbación en la iluminación, gestos faciales o movimientos involuntarios pueden provocar un gran ruido en la señal de rPPG, lo que perjudica en gran medida la obtención de signos vitales.

Es importante destacar que la rPPG es una técnica emergente que busca detectar los cambios en el flujo sanguíneo utilizando cámaras convencionales en lugar de sensores

físicos colocados sobre la piel. Aunque esta técnica ofrece la ventaja de ser no invasiva y permitir mediciones a distancia, su precisión puede verse comprometida por factores ambientales, movimientos del sujeto o condiciones de iluminación inconsistentes. Por lo tanto, el desarrollo de algoritmos de procesamiento de señales robustos y la optimización de las condiciones experimentales son cruciales para mejorar la calidad de la señal de rPPG y maximizar su utilidad en aplicaciones médicas y de monitoreo fisiológico.

2.1.3. Signos vitales

- *Heart Rate* (HR): Es el número de veces que late el corazón en un minuto y se mide en Beats por Minuto (BPM). La obtención de este signo vital se puede hacer mediante dispositivos especializados de contacto o mediante rPPG a través de un video, que es la manera que se implementa en este documento.
- *Heart Rate Variability* (HRV): Es la variación temporal entre latidos cardiacos consecutivos. Una alta variabilidad entre las mediciones, indica un funcionamiento irregular del corazón, lo que es altamente correlacionado con problemas de arritmia y en caso extremo un paro cardiaco.
- Oxigenación en Sangre (SpO₂): Es un indicador de la cantidad de oxígeno transportado por la hemoglobina en la sangre, expresado como un porcentaje. Es una medida importante de la función pulmonar y la oxigenación del cuerpo.

2.1.4. Región de Interés (ROI)

En lo que respecta a la ROI donde se desenvuelve todo el análisis de la rPPG, se utilizan dos herramientas para segmentar las zonas objetivo, rostro y fondo.

- FaceMesh [14]: Es una herramienta desarrollada por Google que se utiliza para realizar un seguimiento y reconocimiento del rostro según puntos faciales. Estos puntos faciales pueden ser ojos, nariz, cejas y otros puntos que indiquen otras zonas dentro del rostro. Este seguimiento y reconocimiento se pueden realizar en tiempo real dependiendo del dispositivo y las características de captura del video. Específicamente en este trabajo, se emplea en conjuntos de datos de videos offline, con la intención de una implementación local para una respuesta en tiempo real. Facemesh se utiliza para identificar las zonas objetivo para la obtención de múltiples señales de rPPG. Esta herramienta es la base para la generación de señales de color a través del tiempo.
- Super Pixels - SEEDS [9]: Este método de segmentación se basa en la idea de muestrear píxeles iniciales o "semillas" en la imagen y luego expandir estas semillas para formar regiones homogéneas basadas en criterios de similitud de color y textura.

La aplicación específica en este trabajo es identificar las zonas del fondo similares y, a partir de áreas alrededor del rostro, construir una caracterización espacio-temporal de la escena mediante los cambios de color a lo largo del tiempo. De esta manera, es posible identificar, en caso de ser viable, las perturbaciones temporales de iluminación y compensarlas en las señales del rostro obtenidas a través de las segmentaciones logradas por Facemesh.

2.1.5. Herramientas de análisis de señales

En este trabajo, se hace énfasis en análisis de señales en el tiempo y en frecuencia. Dado estos dos enfoques, a continuación se presentan las herramientas más usadas:

2.1.5.1. Análisis espectral

- *Discrete Fourier Transform* (DFT) : Es una herramienta matemática para analizar señales en el dominio de la frecuencia. Es la versión discreta de la Transformada de Fourier (FT).

La DFT se utiliza para análisis de señales digitales, donde las señales están discretizadas a intervalos constantes de tiempo. A continuación, se presenta el par de transformaciones para trabajar en frecuencia (Ecuación 2.1) y para volver del espectro de frecuencia al tiempo (Ecuación 2.2).

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn}, \quad k = 0, \dots, N-1 \quad (2.1)$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{-\frac{2\pi i}{N} kn} \quad (2.2)$$

En este trabajo se utilizará como herramienta base de análisis espectral, de manera que sobre esta herramienta se construirán una serie de algoritmos de análisis tanto para la compensación espectral como para el cálculo del Heart Rate (HR).

- Método de Welch: el método de Welch se utiliza para estimar la densidad espectral de potencia (PSD) de una señal, lo que implica encontrar en qué frecuencias se concentra la energía. La densidad espectral de potencia proporciona información sobre cómo se distribuye la energía de la señal en diferentes frecuencias. Por lo tanto, al aplicar el método de Welch, se puede identificar en qué frecuencias específicas se encuentra la mayor energía en la señal.

2.1.5.2. Análisis temporal

Para análisis temporal, se emplean dos herramientas de descomposición de señales con enfoques distintos:

- Ensemble Empirical Mode Decomposition (EEMD) [33]: Sea $x(t)$ la observación o la señal medida; $s(t)$ y $n(t)$, la señal verdadera y ruido, respectivamente. Se tiene:

$$x(t) = s(t) + n(t) \quad (2.3)$$

Como se indica en la Ecuación 2.3, todos los datos son la fusión de señal y ruido. Para mejorar la precisión de las mediciones, la media conjunta (ensemble mean) es un enfoque, donde los datos se recopilan mediante observaciones separadas, cada una de las cuales contiene un ruido diferente. Para generalizar esta idea de conjunto, se introduce ruido en un único conjunto de datos, $x(t)$, como si en realidad se estuviera realizando observaciones separadas como analogía a un experimento físico que podría repetirse muchas veces. El ruido blanco agregado se trata como el posible ruido aleatorio que se encontraría en el proceso de medición. En tales condiciones, la i -ésima observación "artificial" sería:

$$x_i(t) = x(t) + w_i(t) \quad (2.4)$$

En el caso de una sola observación, cada conjunto de observaciones múltiples se imita agregando realizaciones no arbitrarias sino diferentes de ruido blanco, $w_i(t)$, a esa única observación como se indica en 2.4. Aunque agregar ruido puede dar como resultado una SNR mas pequeña, el ruido blanco proporcionará una distribución de escala de referencia relativamente uniforme para facilitar la EMD. Por lo tanto, la baja SNR no afecta al método de descomposición pero en realidad lo mejora para evitar el modo de mezcla (mixing mode), que se refiere a la situación en la que una IMF resultante de la descomposición EMD tiene componentes de diferentes frecuencias. Con base en este argumento, se da un paso adicional al argumentar que agregar ruido blanco puede ayudar a extraer las señales verdaderas en los datos, un método denominado EEMD.

El algoritmo EEMD es el siguiente:

1. Añadir una serie de ruido blanco a los datos de destino, señales de rPPG en este caso.
2. Descomponer las señales con ruido blanco añadido en IMFs.
3. Repita los pasos 1 y 2, una y otra vez pero con diferentes series de ruido blanco cada vez.
4. Obtener las medias (conjunto) de las IMFs correspondientes de las descomposiciones como resultado final.

El efectos de la descomposición utilizando EEMD son que las series de ruido blanco agregadas se cancelan entre sí en la media final de las IMFs correspondientes. Esto ayuda a que las descomposiciones sean más estables, apuntando a la obtención de una señal de rPPG limpia de perturbaciones.

- Singular Spectrum Analysis (SSA) [10]: El objetivo de SSA es el de descomponer una señal de interés en varios componentes aditivos que típicamente pueden interpretarse como "tendencia" (es decir, partes de la serie suaves y que varían lentamente), varios componentes "oscilatorios"(quizás con amplitudes variables) y componentes de "ruido".

Para conseguir la descomposición, se requieren cuatro pasos:

1. Incrustación: La serie unidimensional se representa como una serie multi-dimensional cuya dimensión se llama longitud de ventana. La serie de tiempo multi-dimensional(que es una secuencia de vectores) forma la matriz de trayectoria. El único (y muy importante) parámetro de este paso es la longitud de ventana, que afecta directamente a la resolución de la descomposición. A mayor ventana, mayor descomposición y viceversa.
2. Singular Value Decomposition(SVD): se lleva a cabo la descomposición en valores singulares de la matriz de trayectoria en una suma de matrices bi-ortogonales de rango uno.
3. Agrupación: corresponde a dividir las matrices, calculadas en el paso SVD, en varios grupos y sumar las matrices dentro de cada grupo. El resultado del paso es una representación de la matriz de trayectoria como una suma de varias matrices resultantes.
4. Este último paso transfiere cada matriz resultante a una serie de tiempo, que es un componente aditivo de la serie inicial. La operación correspondiente se llama promediado diagonal. Es una operación lineal y asigna la matriz de trayectoria de la serie inicial.

De esta forma se consigue una descomposición de la serie inicial en varios componentes aditivos.

2.1.6. Blind Source Separation (BSS)

Blind Source Separation (BSS) [25] es el término general utilizado para referirse a las técnicas que permiten la estimación o recuperación considerable de las señales fuente responsables de las observaciones en algún sistema sólo con la información que se encuentra dentro de estos sets de señales. En el contexto específico del desarrollo de esta memoria, las observaciones son señales de color en el tiempo y en el espacio de color RGB, que se obtienen a partir de la técnica rPPG aplicada en el rostro de una persona. Esta técnica permite obtener señales del rostro, que es nuestra Región de Interés (ROI),

para la identificación de signos vitales. Los métodos de BSS permiten, dentro de lo posible, separar las fuentes y encontrar, dentro de alguna de ellas, la información pulsátil que se requiere.

A continuación, se presentan las técnicas utilizadas dentro de este trabajo de título:

- **Principal Component Analysis (PCA):** En primer lugar, para poder identificar las fuentes, se llevan los datos o señales a un espacio estadístico. Este espacio estadístico brinda nuevas herramientas de análisis de los datos que nos proporcionan información importante. Bajo este concepto, PCA utiliza estadísticas de segundo orden, como el promedio y la varianza, para poder identificar y estimar las fuentes del conjunto de datos. Esto sucede porque se maximiza la varianza entre los datos, de manera que sirva como referencia para la separación; a mayor varianza, menor similitud de los datos, lo que genera ejes o vectores de datos ortogonales (producto punto = 0).
- **Independent Component Analysis (ICA):** Para identificar los vectores fuente de un set de datos, ICA utiliza estadísticas de orden superior: Asimetría (Skewness) y Curtosis, tercer y cuarto momento, respectivamente. Estos momentos estadísticos se utilizan debido a que ICA se basa en que las señales fuente cumplen con la no-gaussianidad de los datos. Debido a esto, se utiliza la Curtosis para identificar la forma de distribución de los datos en el histograma construido a partir de los datos y la Asimetría.

2.1.7. Joint Blind Source Separation (JBSS)

Joint Blind Source Separation (JBSS) [2] se presenta como la extensión de BSS a múltiples conjuntos de datos (dos o más). El objetivo de JBSS es agrupar los componentes fuente de cada conjunto de señales en vectores, organizando aquellos que presenten alguna relación dentro del mismo vector. En consecuencia, si se cuenta con múltiples vectores, estos se consideran independientes entre sí, dado que buscan maximizar la varianza vectorial. En el contexto específico del desarrollo de este trabajo, se emplea el enfoque de JBSS debido a la necesidad de manejar múltiples segmentaciones de la ROI del rostro. Cada sub-ROI segmentada se considera como un conjunto de datos independiente. Al realizar un análisis de todas las sub-ROIs, surge un problema de múltiples conjuntos de datos que pueden ser abordados mediante JBSS para identificar los vectores de características fuente. Estos vectores pueden representar perturbaciones de movimiento, de iluminación y la señal objetivo: la señal pulsátil, necesaria para la obtención de HR y otros signos vitales.

- **Independent Vector Analysis (IVA)[16]:**
El problema que resuelve IVA es el siguiente:

Dadas las observaciones:

$$x_i = \sum_{j=0}^L a_{ij} \circ s_j \quad (2.5)$$

Encontrar los vectores fuentes (Source Components)

$$s_i \approx \hat{s}_i = \sum_{j=0}^M w_{ij} x_j \quad (2.6)$$

Donde "o" denota producto por elementos, y L y M es el numero de fuentes y observaciones correspondientes.

Suposiciones:

- Los elementos de un vector fuente son mutuamente independientes de los elementos de otros vectores fuente.
- Dentro de un vector fuente, los elementos dependen en gran medida de los demás.
- El número de fuentes es menor o igual al numero de observaciones.

Fácilmente, se puede tratar este problema como varios problemas de ICA, por que (2.5) se puede reescribir como:

$$X^1 = A^1 S^1; X^2 = A^2 S^2; \dots X^K = A^K S^K$$

Sin embargo, una vez que el algoritmo ICA se aplica por separado a cada elemento de un vector, los elementos de los vectores fuente recuperados se ordenan aleatoriamente. En este caso, después se debe decidir qué componente pertenece a cada vector fuente.

Así que, en lugar de aplicar ICA por separado, abordamos el problema definiendo la dependencia entre componentes multivariados y derivando un algoritmo para el problema de IVA directamente.

Aterrizando el uso de la metodología a la problemática de este trabajo, se definen las estructuras de procesamiento de K Datasets.

$$X^{[k]} = A^{[k]} S^{[k]}, k = 1, 2, \dots, K \quad (2.7)$$

donde $X^{[k]} \in R^{N \times T}$ son los datos observados en el k-ésimo dataset, $S^{[k]} \in R^{N \times T}$ es la matriz fuente que está compuesta por N componentes fuente (source components) con T muestras, y $A^{[k]} \in R^{N \times N}$ es la matriz de mezcla. El objetivo principal de IVA es estimar las matrices fuente $U^{[k]} = W^{[k]} X^{[k]} \in R^{N \times T}$ donde $W^{[k]} \in R^{N \times N}$ es la matriz de desmezcla (demixing) para el k-ésimo dataset.

Se define el i-ésimo vector de componentes fuente (SCV) S_n , concatenando el n-ésimo SCV de cada $S^{[k]}$ como:

$$S_i = \left[\left(s_n^{[1]} \right)^T \left(s_n^{[2]} \right)^T \dots \left(s_n^{[K]} \right)^T \right]^T \in R^{K \times T}$$

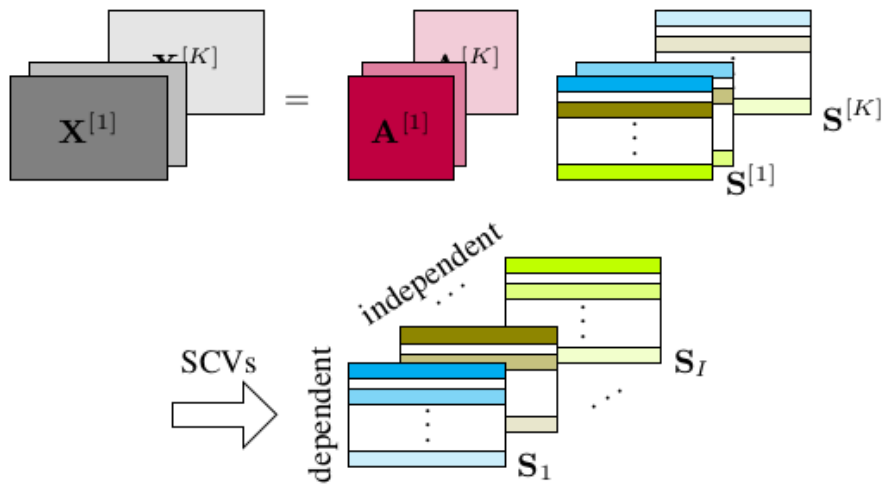


Figura 2.1: Ilustración de funcionamiento de IVA creada por los autores en la cita [19].

2.2. Alternativas de solución

Para la problemática en cuestión, es posible encontrar soluciones en el Estado del Arte que van de la mano directamente con el cálculo del *Heart Rate* (HR).

En trabajos anteriores, la obtención de la señal de PPG para el cálculo de signos vitales, se desarrolla directamente en el rostro del paciente, lo cual corresponde a la ROI desde la cual se pueden apreciar los cambios más significativos de *Blood Volume Pulse* (BVP), los que se ven evidenciados como pequeñas variaciones quasi-periódicas en la señal de PPG construída a partir del video. Sin embargo, estos estudios se desarrollaron sobre datasets que presentan condiciones de laboratorio, o sea, completamente estables en cuanto a iluminación y comportamiento del sujeto de análisis sin perturbaciones, lo que dista completamente de la realidad de la aplicación.

Para abordar de forma realista el problema, diversos autores identificaron las problemáticas principales de perturbaciones a abordar [35][17], que en esencia son tres:

- Cambios de iluminación: Perturbaciones que van desde encender una luz hasta cambios de iluminación provocados por los videos capturados desde una pantalla de teléfono o pc.
- Movimientos Rígidos: Contemplan los movimientos de la cabeza en cualquier dirección, tanto voluntarios como involuntarios.
- Movimiento No-Rígidos: Se consideran las perturbaciones que se generan en la cara al hablar, reírse, pestañear o cualquier movimiento en general que no involucre movimientos rígidos pero que perturben la superficie de la cara al realizar una acción.

Los cambios de iluminación son la problemática más compleja a abordar, de las problemáticas previamente mencionadas, y es el eje principal de este trabajo. Con esto, algunos autores proponen tomar el fondo como referencia [7][34] asumiendo que este se comporta de manera similar a la cara en cuanto a reflectancia, por lo que se podría aislar el componente común en el *BackGround* y restarlo de las señales de PPG obtenidas en el ROI de la cara para poder contrarrestar el efecto de la iluminación. Bajo este concepto, de considerar el *BackGround* como referencia, diversos autores construyeron sobre esta base distintas maneras de abordar el problema de la iluminación pero usando distintas técnicas de procesamiento de las señales en *BackGround* y rostro [7][34][37], así como trabajar con distintos espacios de color [28] para evaluar en cual es más robusta la señal de PPG a las perturbaciones de iluminación.

A continuación, se presentan los trabajos más relevantes en cuanto a técnicas:

En la cita [35], los autores proponen una metodología para abordar la medición de HR bajo condiciones reales tanto de movimiento como perturbaciones de iluminación. El método considera cuatro ejes específicos:

- ROI tracker con Landmarks, para disminuir el efecto de los movimientos rígidos.
- Reducir la interferencia causada por la iluminación, segmentando el Background con Distance Regularized Level Set Evolution (DRLSE) y aplicando PPG en la zona para trabajar con el canal G del modelo RGB, que es el que contiene mayor información lumínica.
- Reducir las perturbaciones producidas por los movimientos no-rígidos de la cara, descartando los segmentos donde esto ocurre.
- Filtros temporales para dejar sólo el rango objetivo 0.7 - 4 HZ, luego aplicando Welch para estimar HR.

El objetivo de reducir la interferencia causada por la iluminación es uno de los temas más complejos de abordar, debido a la naturaleza de la iluminación en la escena. Gracias a esta problemática, surgen distintas metodologías para abordar las perturbaciones de iluminación para obtener HR de manera robusta.

Por otro lado, otros autores [17], proponen un enfoque distinto para calcular HR, desarrollando un sistema de *random patches* dentro de una ROI que está compuesta por la zona de las mejillas. Dentro de esta zona, se seleccionan dos *patches* de manera aleatoria, para aplicar PPG en cada uno, obteniendo las señales respectivas para luego aplicar BSS, obtener la señal estadísticamente independiente y aplicar el método de Welch para tener una caracterización del PSD. Para identificar si la selección de patches satisface los requerimientos para una inferencia adecuada de HR, se seleccionan los dos mayores valores de potencia y se saca una relación entre ellos. Si la relación es mayor un *threshold* pre-definido, se calcula el HR multiplicando por 60 la frecuencia correspondiente al mayor peak de potencia y se almacena en un histograma de valores de HR, que los autores llaman *majority vote*. El efecto de este histograma, provoca que se obtenga gran cantidad de estimaciones de valores de HR y que se elija la adecuado dependiendo de los "votos" que tenga.

La calidad de este método depende de la cantidad de iteraciones del método y de las condiciones específicas que deben cumplirse en la escena para que los *patches*, de donde se extrae la información, se consideren válidos. Por ejemplo, es importante que los *patches* estén iluminados por la misma fuente de luz, lo cual no suele cumplirse en situaciones reales.

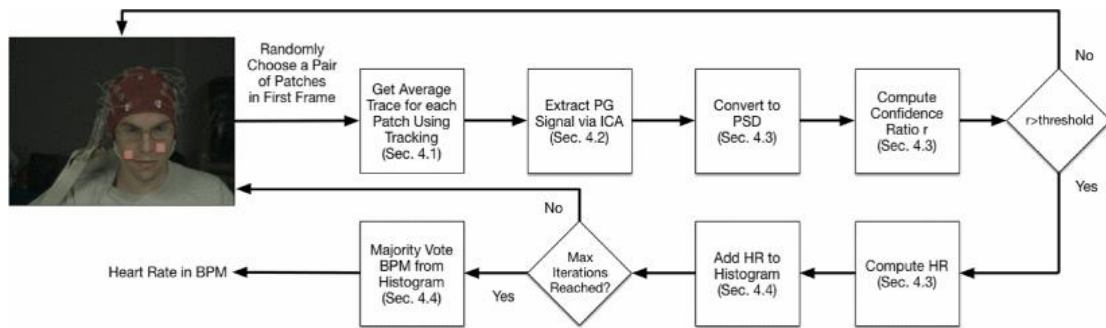


Figura 2.2: Metodología propuesta por los autores en la cita [17].

Sin embargo, esta metodología no es capaz de seguir los cambios de iluminación del entorno, debido a que sólo se centra en análisis dentro del rostro, sin preocuparse de la interacción de la iluminación con la escena.

La falta de análisis de la interacción de la iluminación con el entorno es abordada en la cita [7], donde proponen un enfoque robusto para abordar los desafíos de la iluminación utilizando Join Blind Source Separation (JBSS) y Ensemble Empirical Mode Decomposition (EEMD)[33]. Esta estrategia se fundamenta en la observación de que tanto el rostro como el fondo presentan variaciones de iluminación similares. El fondo se considera una referencia de ruido que se debe eliminar de la señal facial. En este punto, JBSS entra en juego para identificar la componente común entre las señales del rostro y el fondo. Una vez identificada esta componente común entre ambos conjuntos de datos, se anula para eliminar los efectos de la iluminación y, posteriormente, se reconstruyen las señales de PPG del rostro.

Luego, se realiza un análisis del canal G en el espacio de color RGB reconstruido y se aplica EEMD. Esto tiene como objetivo identificar los modos intrínsecos que componen la señal y determinar en qué modo se encuentran los valores más altos de la frecuencia cardíaca (HR).

Por otro lado, crearon un Dataset con el propósito de estudiar los efectos de las perturbaciones de iluminación. Este Dataset incluye personas iluminadas con una luz LED conectada a un generador de señales que produce señales sinusoidales, cuadradas, de rampa y pulsos a frecuencias de 0.1, 0.5, 1 y 3 Hz. Estas frecuencias se eligieron dentro del rango de operación de la frecuencia cardíaca (0.7 a 3 Hz).

Al aplicar la técnica de Independent Vector Analysis (IVA) [16] en JBSS, es posible recuperar señales de perturbación que son comunes a todas las fuentes y presentan una alta correlación. Sin embargo, cuando existe ruido similar a la frecuencia cardíaca, o cuando se encuentra dentro de un rango específico, este ruido contribuye a aumentar la energía en el espectro de frecuencias, por lo que se acoplan los métodos que consideran la mezcla lineal y no pueden extraer esta perturbación.

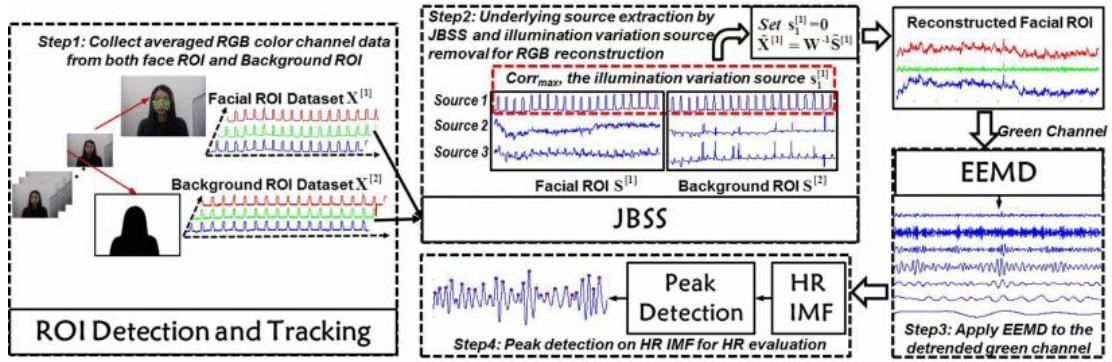


Figura 2.3: Metodología propuesta por los autores en la cita [7].

En la cita [34], los autores proponen nuevas herramientas para eliminar el efecto de las perturbaciones de iluminación, siguiendo la misma línea que se presenta en el trabajo previamente descrito[7]. En este caso, para identificar las fuentes de perturbación de iluminación, utilizaron Partial Least Squares (PLS). Este método se utiliza para obtener los componentes principales de un conjunto de datos.

Posteriormente, se identificaron las Intrinsic Mode Functions (IMFs) de la señal utilizando la técnica de Multivariate Empirical Mode Decomposition (MEMD). Es importante destacar que MEMD puede analizar más de un canal a la vez, en contraste con EEMD, que analiza un sólo canal. Según los autores, en comparación con el enfoque que utiliza JBSS + EEMD, esta nueva metodología muestra un mejor rendimiento, mejorando aproximadamente un 5 % según los resultados reportados por los autores.

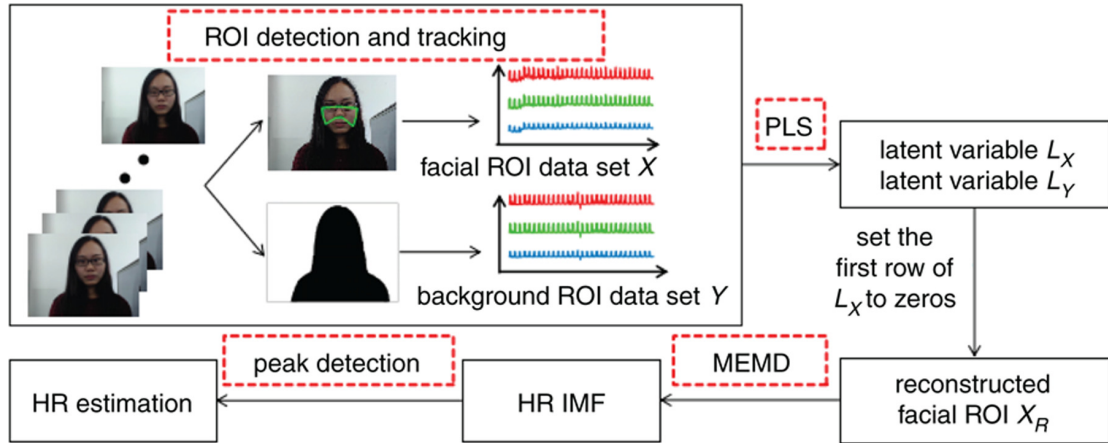


Figura 2.4: Metodología propuesta por los autores en la cita [34].

En la cita [24], los autores proponen un modelo para la adquisición simultánea de datos RGB y *Near-InfraRed* (NIR). Este enfoque se centra en la medición de la frecuencia cardíaca (HR) en situaciones que involucran cambios drásticos de iluminación y movimientos abruptos o prolongados. En primer lugar, se utiliza un *tracker* de rostro seguido

de Facemesh [14], que permite identificar 48 regiones de interés, excluyendo los labios y los ojos. A partir de estas regiones, se generan 48 señales de PPG.

Posteriormente, se trabaja con matrices de observación tanto para NIR como para RGB. Estas últimas pueden estar afectadas por cambios en la iluminación y variaciones en la señal de PPG. Se realiza análisis mediante *Robust Principal Component Analysis* (RPCA) [6] para identificar los componentes periódicos de baja frecuencia en la señal que se separan en una matriz de observación de rango bajo, mientras que los elementos restantes se consideran ruido. Para determinar la calidad de las señales PPG en estas regiones, se calcula la *Signal Noise Ratio* (SNR) en el espectro de frecuencia. Si no supera un umbral predefinido, se descartan estas señales.

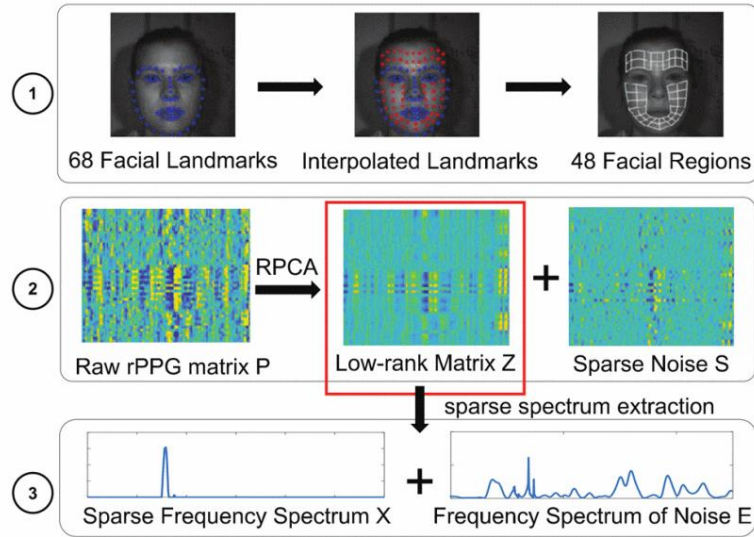


Figura 2.5: Metodología propuesta en por los autores en la cita [24].

En la cita [20], los autores proponen un algoritmo denominado Adaptive SpatioTemporal Homomorphic Filtering (ASTHF), centrado en la separación de las señales de frecuencia cardíaca (HR) y el ruido inducido por las fuentes de iluminación. En primer lugar, emplean un filtro homomórfico para hacer que el ruido de iluminación sea linealmente separable. Esto se fundamenta en la hipótesis de que una imagen es una combinación de componentes de iluminación y textura, de manera que al aplicar el logaritmo a esta relación, se obtiene una forma lineal y, al aplicar la FFT para realizar la compensación, se conservan estas cualidades separables. Luego, se aplica un filtro espacio-temporal para eliminar el ruido espacial y temporal en el video.

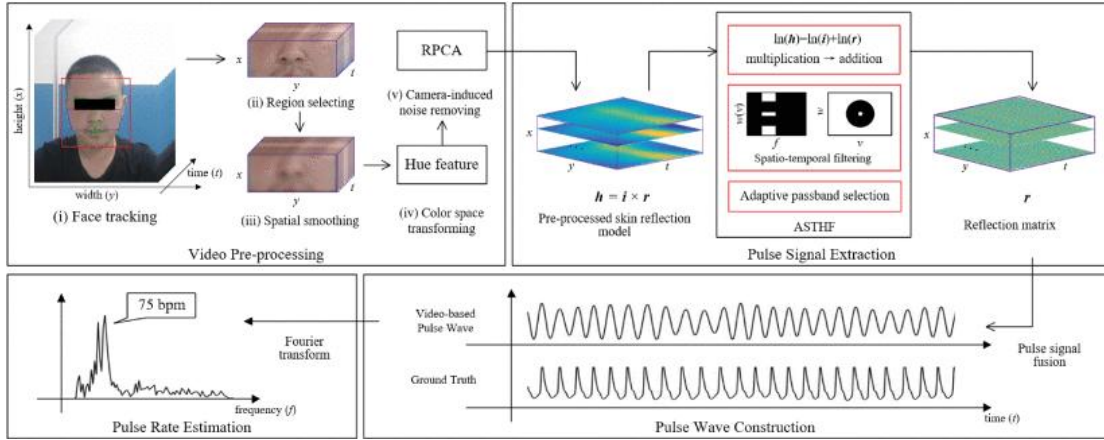


Figura 2.6: Metodología propuesta por los autores en la cita [20].

En la cita [36], los autores proponen un método robusto para condiciones desiguales de iluminación en la escena. En primer lugar, se identifica la ROI en el rostro de la persona, se recorta y se utilizan las señales de color RGB para establecer un modelo tridimensional de rPPG. Luego, aplican sobre la imagen el logaritmo para transformar la relación entre la iluminación y la reflectancia, convirtiendo una multiplicación en una suma, con el fin de hacer que los efectos sean linealmente separables. Una vez separado el componente de reflectancia, aplican EEMD para dividir este componente en múltiples IMFs y luego identifican aquellas que contienen una mayor potencia mediante el método de Welch para obtener la señal que contenga el HR.

Además, en este trabajo mencionan que los problemas de iluminación pueden dividirse en problemas espaciales y temporales. El hecho de abordar estos dos tipos de problemas hace que el modelo sea robusto, lo cual, se aplica al separar los componentes de iluminación y reflexión en la imagen.

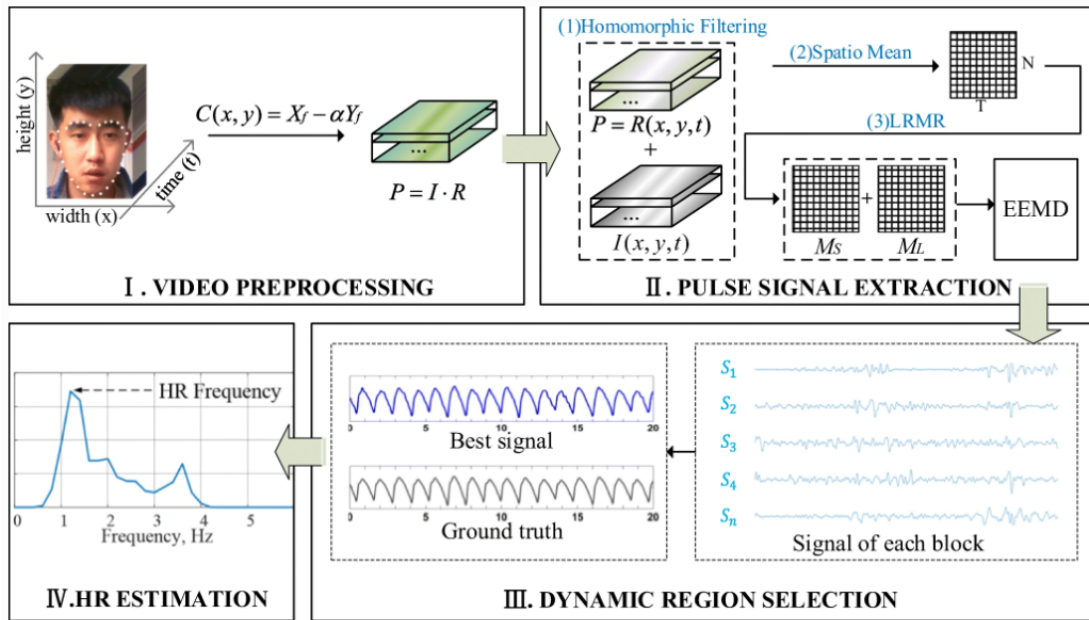


Figura 2.7: Metodología propuesta por los autores en la cita [36].

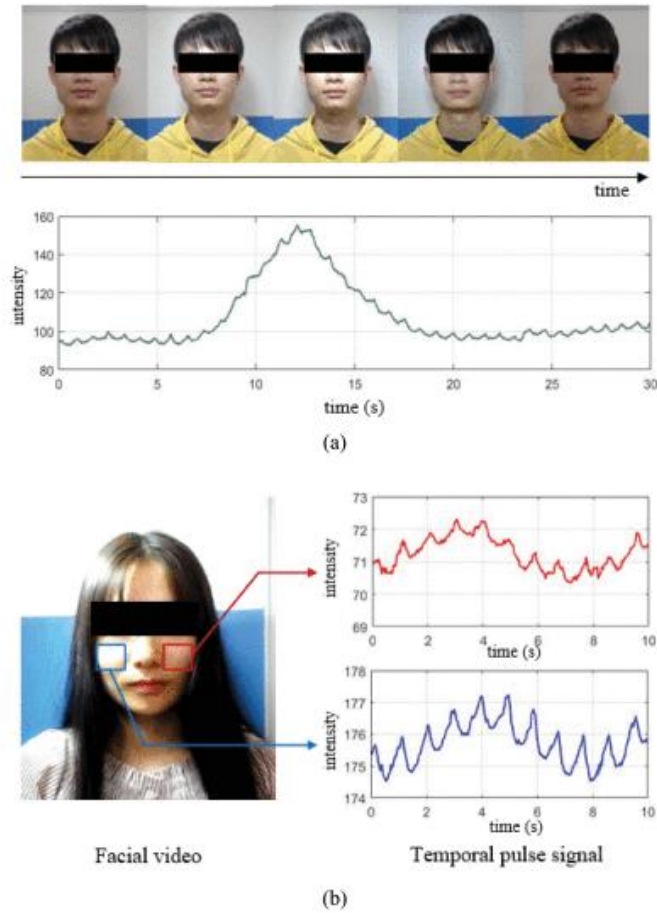


Figura 2.8: Perturbación temporal y perturbación espacial presentada por los autores en la cita [36].

En la cita [28], los autores proponen un sistema robusto para abarcar la problemática de la iluminación espacialmente desigual utilizando Ensemble Empirical Mode Decomposition (EEMD) + Multiset Canonical Correlation Analysis (MCCA). Para lograr esto, dentro de la ROI objetivo que es el rostro, generan una grilla de $N \times N$ sub ROIs, que en este caso es de 4×4 , para obtener una caracterización espacio-temporal local de la iluminación. De cada sub ROI se obtienen señales de rPPG que son sometidas a un proceso de selección que consta de:

- Pasar las señales del espacio de color RGB al espacio LAB, seleccionando el canal L para calcular el promedio y su desviación estándar. Estos valores sirven para la selección de las mejores sub ROIs, donde el promedio corresponde a la intensidad de iluminación y la desviación estándar a la variación de iluminación. De esta manera, se ordenan de manera descendente los valores de intensidad de iluminación y de manera ascendente las variaciones de iluminación para luego

interseccionar estos valores y seleccionar las regiones que contengan las condiciones descritas.

- Con las zonas pre-seleccionadas, se procede a calcular el SNR del canal G del espacio RGB de cada sub ROI y se seleccionan los con mayor valor, en este caso cuatro zonas.

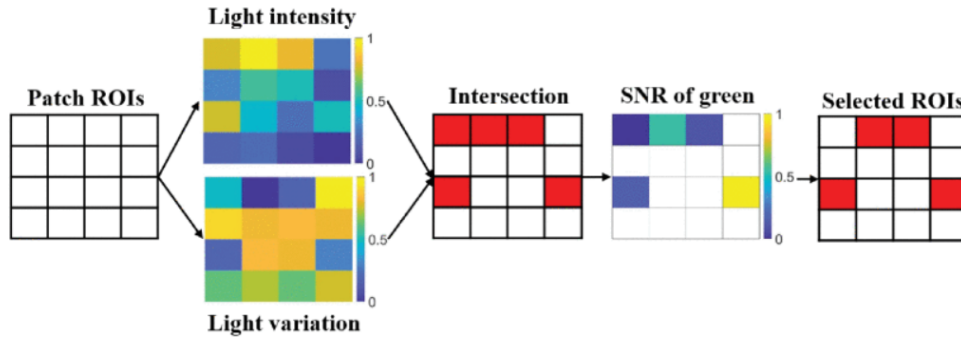


Figura 2.9: Proceso de selección de ROIs de metodología propuesta en la cita [28].

Con este proceso de pre-selección, se logra mitigar el efecto de la iluminación espacialmente desigual al igual que una parte de los efectos temporales de las perturbaciones de iluminación.

Para poder obtener una señal de rPPG más limpia, se procede a aplicar EEMD a cada región seleccionada para obtener las Intrinsic Mode Functions (IMFs) y seleccionar sólo las que se encuentran en el rango de $[0.7, 2.5]$ Hz, obteniendo así cuatro conjuntos de IMFs. Luego estos conjuntos de señales se ingresan a MCCA para obtener las señales correlacionadas entre los sets de datos, que son las aproximaciones obtenidas de las variaciones del volumen de sangre a lo largo de las mediciones. De este conjunto de datos se selecciona la primera señal, que es la con mayor correlación cruzada, y luego se aplica FFT para calcular HR.

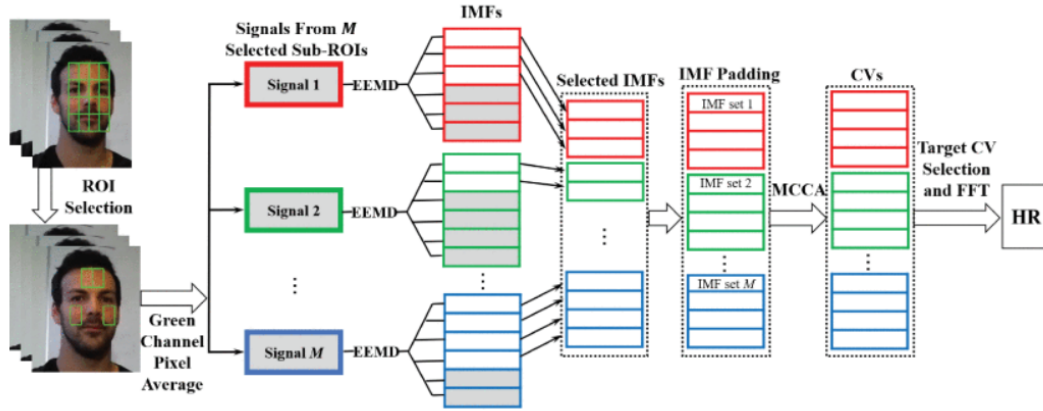


Figura 2.10: Metodología propuesta por los autores en la cita [28].

En la cita [37], los autores proponen una metodología que utiliza JBSS y un proceso de proyección basado en el modelo de reflexión de iluminación en la piel. Para lograr obtener una señal de rPPG lo menos contaminada posible, se sigue el mismo proceso descrito en otras metodologías: aplicar JBSS a los datos obtenidos de las señales de Background y de la ROI en el rostro para luego extraer la información común, que en este caso son las perturbaciones de iluminación a lo largo del tiempo. Luego, se suprime este componente de las señales obtenidas del rostro y se vuelven a reconstruir las señales de la ROI, obteniendo señales temporalmente limpias de perturbaciones de iluminación.

Con las señales de rPPG reconstruidas, se procede a aplicar el modelo de reflexión de iluminación en la piel. Al usar este modelo, se trabaja con la suposición de que en la señal de rPPG hay dos componentes principales: La reflexión difusa y la reflexión especular. La reflexión difusa es la señal que contiene la información de cómo varía en el tiempo el volumen sanguíneo que a su vez nos entrega la información pulsátil de frecuencia cardíaca (HR). Por otro lado, la reflexión especular corresponde a la señal que es reflejada por la piel directamente, que depende exclusivamente de cuánta melatonina tenga la persona. Para poder separar estas señales se aplica ICA y luego, para identificar cuál es el componente pulsátil, se hace una correlación con la señal del canal G del espacio de color RGB, que es la señal que tiene mayor información pulsátil, para ver cual es la que tiene mayor similitud y se selecciona como candidato para el cálculo de HR. Con la señal seleccionada, se aplica un filtrado entre 0.7 y 4 Hz para luego aplicar FFT y seleccionar el componente que tenga mayor potencia.

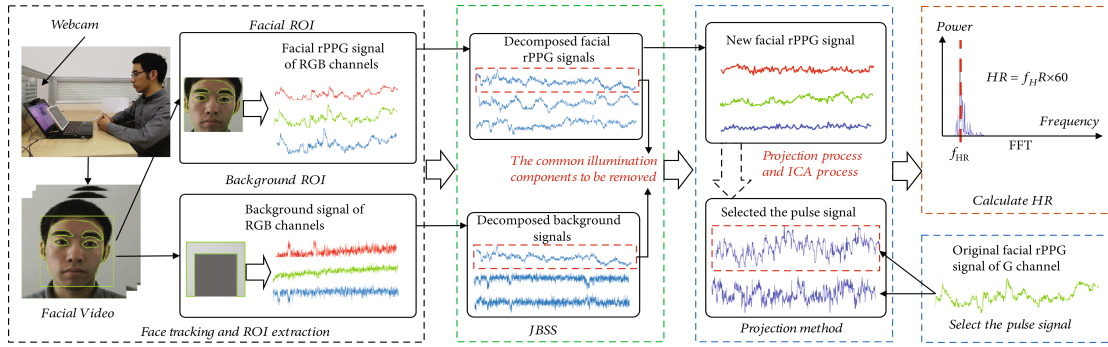


Figura 2.11: Metodología propuesta por los autores en la cita [37].

Como se propone en [36], diferentes de perturbaciones afectan la iluminación tanto de manera temporal como espacial en un secuencia video. Los trabajos presentados por los autores en las citas [7][34][28][37], sólo abordan la problemática desde la perspectiva temporal. Al abordarla de esta manera, la señal de referencia obtenida sólo proporciona información temporal sobre los cambios de iluminación en la escena, sin tener en cuenta la direccionalidad de la iluminación ni su distribución en la escena, lo que limita la caracterización espacio-temporal completa de las perturbaciones causadas por la iluminación a lo largo del video.

Por lo tanto, considerando que la solución más cercana en el Estado del Arte no ha sido exhaustiva en cuanto a la compensación espacial de la escena, la rama más cercana que se acerca a este aspecto es la compensación de la iluminación en imágenes. Al incorporar esta rama, se abren nuevas posibilidades para combinar técnicas y desarrollar una metodología que permita la compensación espacio-temporal de la iluminación y, como resultado, obtener una señal de PPG coherente.

Dentro del Estado del Arte de la compensación de iluminación en imágenes, se encuentran los algoritmos de Ecualización de Histograma (HE). Las técnicas más comunes en este caso son:

- **Global HE (GHE)[11][31]:** Utiliza el histograma de la imagen para transformar todos los píxeles, distribuyendo los niveles de iluminación en toda la imagen.
- **Adaptive HE (AHE)[11][31]:** Mejora la ecualización global al centrarse en análisis de vecindad. En vez de usar sólo un histograma para toda la imagen, ocupa múltiples histogramas que dependen del sector donde se analiza. Esto genera una adaptación localizada. Sin embargo, debido a esta característica, puede sobre-

plificar el ruido en regiones homogéneas, lo que puede resultar en una apariencia granulada o ruidosa.

- **Contrast Limited AHE (CLAHE)[11]:** Es una mejora de AHE diseñada específicamente para abordar el problema de la amplificación del ruido en imágenes. Este método se basa en la adaptación local del histograma y utiliza un parámetro llamado "clipLimit" para controlar la amplificación excesiva de los valores en el histograma. Cuando un valor en el histograma supera el límite establecido por "clipLimit", los píxeles excedentes se redistribuyen uniformemente entre todos los valores de gris. Dado que CLAHE trabaja con un sólo canal a la vez, esta redistribución ayuda a mejorar el contraste y resaltar los detalles en la imagen. La capacidad de limitar la amplificación del contraste es especialmente útil para evitar la exageración de ruido en áreas de baja variabilidad de intensidad.

Se utilizará esta última técnica para evaluar su efectividad en la compensación de la iluminación debido a su implementación sencilla y su enfoque en analizar la imagen en bloques pequeños y aplicar la ecualización de histograma en cada bloque de forma independiente. Además, CLAHE tiene la ventaja de tener un límite que evita la saturación del contraste y al mismo tiempo limita el ruido. Por lo tanto, aplicado a la problemática específica, resulta útil ya que equilibra el contraste local y el control del ruido para mejorar cualitativamente la imagen de manera global.

Por otro lado, existen algoritmos basados en la teoría Retinex [18] que proponen una descomposición de la imagen en dos componentes distintos: la reflectancia y la luminancia. Esta teoría tiene como objetivo lograr una percepción constante del color, independientemente de las variaciones en la iluminación. Aplicando esta teoría a la problemática en cuestión, al separar el componente de luma de la imagen, es posible mejorar la iluminación a nivel local sin comprometer la coherencia espacial en relación a la parte global, lo que resulta en un método robusto.

La teoría Retinex ha dado lugar a las siguientes bases de implementación:

- **Single-Scale Retinex (SSR)[11]:** Aplica la teoría Retinex en una única escala espacial. Se basa en la idea de que la percepción del color y la luminosidad se pueden mejorar al realizar una compensación local de la iluminación. Para lograrlo, el algoritmo utiliza técnicas de manipulación de histogramas y filtrado espacial. SSR es adecuado para mejorar la calidad de la imagen en términos de contraste y detalles locales.
- **Multi-Scale Retinex (MSR)[11]:** A diferencia de SSR, el algoritmo MSR opera en múltiples escalas espaciales. Reconoce que las variaciones de iluminación pueden ocurrir a diferentes escalas y utiliza una combinación ponderada de las respuestas Retinex a diferentes escalas para lograr una mejora de la imagen más

robusta y global. MSR se destaca por su capacidad para conservar detalles finos y realzar la estructura de la imagen.

- **Multi-Scale Retinex algorithm with Color Restoration (MSRCR)[11]:** Este algoritmo es una extensión del MSR que también considera la restauración del color. Reconoce que la variación de iluminación puede afectar la percepción del color y busca restaurar la apariencia cromática original de la imagen. MSRCR utiliza técnicas de corrección de ganancia de color y compensación de balance de blancos para lograr una mejora completa de la imagen en términos de iluminación y color.

Estas implementaciones forman la base teórica sobre la cual se pueden desarrollar algoritmos específicos para abordar problemas particulares, manteniendo el enfoque principal de la teoría. Gracias a esta flexibilidad, es posible seleccionar los algoritmos que mejor se adapten a la problemática en cuestión y presentarlos como métodos viables para el análisis de la Hipótesis planteada.

Por otra parte, están los enfoques de mejora de imagen con baja iluminación o sobre exposición, que utilizan métodos distintos a los mencionados anteriormente. En este caso, los enfoques son los siguientes:

- **Low-Light Image Enhancement via Illumination Map Estimation (LIME)[12][1]:** El enfoque principal de LIME es modificar y resaltar características de una imagen con baja iluminación, pero evitando la saturación. Opera de manera individual cada píxel de cada canal, en una imagen con formato RGB. Este método interpreta que la mayor fuente viene dada por el valor máximo de alguno de los tres canales y la fuente de iluminación más baja esta dada por el mínimo de uno de los tres canales. Luego de un proceso de normalización, dividiendo cada valor píxel por el valor del máximo calculado anteriormente y se refina el mapa de iluminación inicial con la aplicación del método LIME.
- **Dual Illumination Estimation for Robust Exposure Correction (DUAL)[12][1][38]:** Este método utiliza un enfoque de proyección separada y corrección de sobre exposición como estimación de iluminación trivial de la imagen de entrada y la imagen de entrada invertida. Al usar la estimación de iluminación dual, se obtienen dos resultados de corrección de exposición intermedios para la imagen de entrada, una corrección sobre las regiones sub-expuestas y una restauración de las regiones sobre-expuestas y utilizando una técnica de fusión de imágenes de exposición múltiple, en la cual se combinan de forma adaptativa las partes visualmente mejor expuestas en las dos imágenes de corrección de exposición intermedia junto con la imagen original, dando como resultado, una imagen globalmente bien expuesta.

Con los algoritmos descritos anteriormente, tanto para el cálculo de HR robusto como para la compensación de iluminación en imágenes, se espera poder desarrollar una metodología viable para verificar si la hipótesis es correcta o no.

2.3. Discusión

Analizados los trabajos del estado del arte, se pueden identificar los puntos de interés para abordar la problemática principal en el desarrollo de la sección siguiente.

En solo tres trabajos ([7], [34], [37]), específicamente en la parte de estimación de HR robusto, se plantea la idea de comparar la información del fondo con la del rostro para compensar los componentes comunes y obtener señales de rPPG libres de perturbaciones. Sin embargo, estos trabajos no consideran la caracterización espacial de la iluminación tanto en el rostro como en el fondo, lo que limita su capacidad para compensar perturbaciones complejas.

Por otro lado, hay trabajos ([20], [28]) que se centran únicamente en la información del rostro y proponen metodologías robustas para la estimación de HR. Sin embargo, al no recopilar información del entorno del sujeto, se pierde información valiosa en términos de distribución espacial y temporal de la iluminación.

Además, existen trabajos ([11], [38], [12]) que abordan la compensación de la escena de manera espacial, frame a frame, con el objetivo de normalizar la iluminación. Esta aproximación logra compensar la iluminación de manera espacial, pero puede perder la caracterización temporal debido a la individualidad de la compensación en cada frame.

Considerando los problemas identificados al analizar el estado del arte, se desarrollarán cuatro metodologías exploratorias en este estudio para abordar las perturbaciones de iluminación en ambientes no controlados. Las dos primeras se basan en ideas y metodologías propuestas en el estado del arte, pero se adaptan con el objetivo de resolver la compensación de iluminación. Las otras dos corresponden a metodologías desarrolladas específicamente para abordar la problemática de las perturbaciones espacio-temporales.

En el siguiente capítulo, la Metodología Exploratoria uno y dos abordarán la compensación espacial y temporal, respectivamente, desde los trabajos analizados en el estado del arte. Sin embargo, estas implementaciones no lograron compensar las perturbaciones ni calcular la frecuencia cardíaca (HR) con una precisión aceptable. Más bien, se observó una tendencia a cometer errores significativos en la inferencia de HR.

Es por esto que implementaron las Metodología Exploratoria tres y cuatro, donde se adoptó un enfoque de caracterización espacio-temporal. Se generaron señales distribuidas en el fondo y en el rostro para identificar las direcciones de la iluminación de manera espacial. Además, se aplicó la correlación con el fondo de manera conjunta para compensar las perturbaciones de iluminación temporales.

Este enfoque de caracterización espacio-temporal demostró ser más prometedor para abordar las perturbaciones de iluminación en ambientes no controlados en comparación con las metodologías anteriores. Sin embargo, aún se requiere un trabajo adicional para mejorar la precisión y la robustez de estas metodologías en entornos del mundo real.

Capítulo 3

Diseño de la solución

Para diseñar una solución que aborde tanto los problemas espaciales como temporales asociados con la perturbación de iluminación en la señal de rPPG, se proponen tres metodologías exploratorias. Estas metodologías tienen como objetivo analizar las propuestas presentadas en el Estado del Arte y, al mismo tiempo, servir como componentes fundamentales de una metodología más completa.

Para realizar un análisis previo al capítulo de Validación (Capítulo 4), la primera metodología exploratoria será analizada con en una porción del Dataset NAM-EDGE, mientras que las demás serán evaluadas en casos específicos de interés.

3.1. Desarrollo de la primera metodología exploratoria

Esta línea de desarrollo, se centra en la compensación de iluminación de manera espacial, de forma independiente en cada frame para ver el efecto que produce sobre la señal de rPPG. Para identificar los efectos y caracterizarlos, hay que centrarse en los métodos que mantengan la estructura del color y que, en lo posible, afecten sólo a la iluminación de la escena para que sea coherente respecto al video original.

Así que, en esta parte, se seleccionarán los mejores algoritmos del estado del arte, que se analizaron en la sección previa (Sección 2.2), para poder ver si compensando una iluminación no homogénea en la zona, background y rostro, se logra aumentar la precisión de la medición de signos vitales mediante rPPG.

Para ver si es factible una implementación, en primera instancia, se deben cumplir dos restricciones:

1. Tiempo de compensación de cada *frame*: La compensación de un segmento de video de 20 [s] a 30 FPS no debe ser mayor a 30 [s]. Llevándolo a compensación de un sólo *frame*, el tiempo no debe ser mayor a 50 [ms] en promedio.

2. Restricciones de Hardware: Debido a que se trabajará con algoritmos, estos tienden a tener grandes iteraciones que consumen gran cantidad de memoria RAM (en algunos casos) para mantener cálculos de los distintos niveles. Es por esto que se analizarán las restricciones de Hardware asociadas a los algoritmos seleccionados al usar distintos tamaños de frame: High Definition (HD) 1280 x 720 píxeles , Full HD (FHD) 1920 x 1080 píxeles y Ultra HD (UHD) 3840 x 2160 píxeles.

A continuación, se muestra un cuadro comparativo donde se aplicaron las dos restricciones. Se fijó un tamaño de fotograma estándar FHD para una comparación uniforme.

Método	Resolución de <i>frame</i>	Tiempo [s]
CLAHE	FHD	0.018322
LIME	FHD	67.521936
DUAL	FHD	129.016972
CLAHE-DUAL	FHD	0.062321
MSRCR	FHD	23.722719

Cuadro 3.1: Análisis Primera y Segunda Restricción.

Respecto a la restricción de tiempo, lo únicos métodos que lograron cumplir con la restricción son CLAHE y CLAHE-DUAL. Los demás métodos, si bien muestran un incremento considerable en la calidad de la imagen a nivel cualitativo, no logran cumplir con la restricción de tiempo asociada al procesamiento esperado para proporcionar una respuesta en un tiempo coherente.

Se muestra a continuación, una evaluación sobre un subgrupo de videos del Dataset NAM-EDGE (Descrito en Capitulo 4). Se destaca que se utiliza un subgrupo para determinar si es viable continuar con el análisis en la sección de Evaluación.

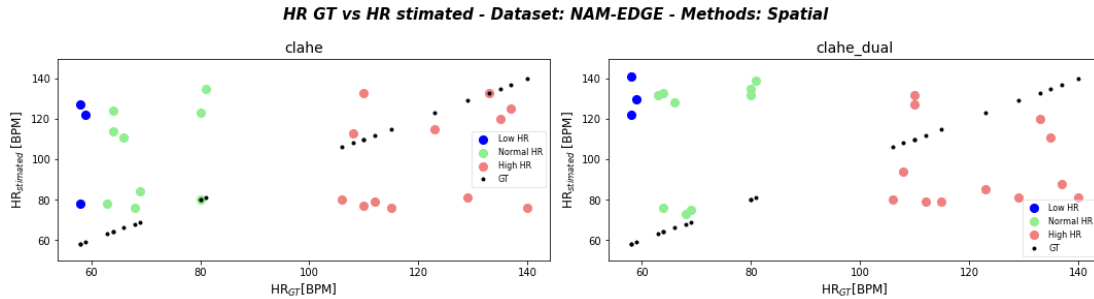


Figura 3.1: Gráfico HR_{GT} vs $HR_{Estimated}$ de estimaciones sobre un subgrupo de videos de Dataset NAM-EDGE.

Respecto a la evaluación del subgrupo de videos, en la Figura 3.1 se presentan los resultados. La línea de color negro representa la línea de Ground Truth (GT). Cuanto

más se acerquen los valores estimados a la línea de GT, mayor será la precisión. Es evidente que al utilizar estos métodos, que son mejores en términos de tiempo pero no tanto en calidad de imagen cualitativa, no es posible estimar adecuadamente un Heart Rate (HR) con buena precisión. En su lugar, se observa un error generalizado que se traduce en sobreestimaciones y subestimaciones, para todos los rangos de análisis: *Low HR*, *Normal HR* y *High HR*.

3.2. Desarrollo de la segunda metodología exploratoria

Este desarrollo se centra en la compensación de iluminación de manera temporal, utilizando los métodos y herramientas propuestas en el Estado del Arte. En esta línea, se ocuparán herramientas Avanzadas de Procesamiento de Señales, así como, métodos estadísticos para poder identificar y caracterizar completamente el comportamiento de las señales de rPPG bajo perturbaciones temporales de iluminación.

Para llevar a cabo este análisis, es necesario primero identificar cuáles son los métodos y combinaciones propuestas más relevantes del Estado del Arte (Sección 2.2):

- **Ensemble Empirical Mode Decomposition (EEMD)[33] + Independent Vector Analysis (IVA)[16]**
- **IVA + Skin Reflection Model**
- **Partial Least Square Correlation (PLSC) + Multi-set Empirical Mode Decomposition(MEMD)**
- **EEMD + Multi-set Canonical Correlation Analysis(MCCA)**
- **Robust Principal Component Analysis (RPCA) + EEMD**

Para poder evaluar el desempeño de los métodos propuestos en el estado del arte, es necesario identificar las partes fundamentales de cada uno y agrupar las similares para poder tener una base sólida para la construcción en código de estos métodos, ya que ningún paper dejó disponible de manera pública el código utilizado para el desarrollo de sus metodologías.

Con lo anterior en consideración, se desarrollaron las metodologías utilizando bibliotecas de Python específicas para métodos particulares, como IVA[16], EEMD[33], MCCA[15][29], que no necesariamente son las mismas utilizadas en los papers, ya que cada método presenta variaciones en su composición. Por otro lado, para implementar los métodos no encontrados inicialmente en bibliotecas, se recurrió al pseudocódigo descrito en los papers, así como a los diagramas generales que explican el método, con el fin de desarrollar las herramientas faltantes. Esto podría introducir un error inicial

entre los modelos adaptados en esta sección y los modelos originales de los papers [17][37][28][34].

Para identificar y analizar el desempeño de las metodologías desarrolladas según las aproximaciones mencionadas anteriormente, de manera preliminar, es necesario verificar si se cumple la robustez ante perturbaciones de iluminación, temporal y/o espacial. Para ello, se analizarán dos casos particulares: cambios bruscos de iluminación y *Heart Rate* (HR) mayor a 120 BPM.

3.2.1. Caso 1: Perturbaciones de iluminación

En este caso, se produce una perturbación de iluminación del tipo escalón. Alrededor del segundo 7.5, se apaga la luz de los tubos fluorescentes en la escena, dejando sólo luz natural. Esto afecta directamente a las señales del fondo y del rostro, como se puede observar.

Bajo esta particularidad, se evalúan sobre este caso todos los métodos mencionados anteriormente para obtener una idea respecto al comportamiento de las metodologías frente a perturbaciones agresivas de iluminación.

En la Tabla 3.2, se observa que el método IVA+*Skin reflection* exhibe el menor error, mostrando una subestimación del valor real de la GT, pero manteniéndose cercano con un 12 % de error. Por otro lado, las demás metodologías presentan una sobrestimación del valor de HR, lo cual es perjudicial para fines prácticos. En el peor de los casos, el método IVA+EEMD muestra un error del 184 % en comparación con la GT.

	GT	IVA+EEMD	IVA+skin reflection	PLS+MEMD	EEMD+MCCA	RPCA+EEMD
HR [BPM]	58	165	51	84	95,14	123
Error [%]	0	184,48	-12,07	44,83	64,03	112,07
Error [BPM]	0	107	-7	26	37,14	65

Cuadro 3.2: Tabla comparativa de errores para caso 1.

3.2.2. Caso 2: *Heart Rate* alto mayor a 120 BPM

Para verificar la capacidad de las metodologías para calcular HR superiores a 120 BPM, se realiza un análisis utilizando el video de un sujeto que presenta el HR más alto entre todos los Datasets disponibles para este estudio [3][13][39]. Este sujeto registró un HR de 170 BPM después de realizar ejercicio de manera intensa.

Respecto a las condiciones ambientales y las del sujeto, en el video se observa una iluminación LED estable, un fondo invariable y el sujeto mantiene la cabeza inmóvil con un dispositivo especial. Esto facilita la obtención de una señal limpia de rPPG sin la interferencia de movimientos rígidos de la cabeza.

En la Tabla 3.3, se observa que la metodología EEMD+MCCA exhibe el menor error, con un valor de 1.72 %, que prácticamente es el valor esperado. Por otro lado, se encuentran los demás métodos, que en general, lograron buenas estimaciones en comparación con el caso anterior. Esto posiblemente se deba a las condiciones de entorno y toma de datos diferentes a las del video del caso 1, lo que induce a menores errores en el cálculo de HR.

	GT	IVA+EEMD	IVA+skin reflection	PLS+MEMD	EEMD+MCCA	RPCA+EEMD
HR [BPM]	170,04	162	156	159	172,29	183
Error [%]	0	-4,73	-8,26	-6,49	1,32	7,62
Error [BPM]	0	-8,04	-14,04	-11,04	2,25	12,96

Cuadro 3.3: Tabla comparativa de errores para caso 2.

En general, este breve análisis de las metodologías desarrolladas a partir de los papers originales sirve para evaluar su efectividad frente a casos particulares. El análisis completo de los datasets disponibles se encuentra en la Sección 4, Evaluación.

3.3. Desarrollo de la tercera metodología exploratoria

Esta metodología de compensación temporal, toma las herramientas más relevantes de la ME2 y las organiza de tal manera que se saque el máximo provecho a estas.

Para abordar la problemática de la compensación y posterior cálculo de los signos vitales, esta metodología se separa en dos partes: compensación de la iluminación y calculo de signos vitales.

Compensación de perturbaciones temporales de iluminación

Para abordar la problemática de las perturbaciones de iluminación en el entorno que afectan al rostro, se propone como base de comparación el algoritmo IVA. Este algoritmo permite procesar K sub-ROIs con N observaciones, que en este caso son los canales del espacio de color RGB obtenidos mediante rPPG, y donde cada observación contiene T muestras. En este caso en particular, se tienen K sub-ROIs del rostro y $K1$ sub-ROIs del fondo, con $K \neq K1$ y manteniendo la misma cantidad de muestras T e igual numero de observaciones N , lo que da origen a una matriz de observaciones del rostro con una forma de (N, T, K) y del fondo $(N, T, K1)$. Para más detalles de la forma de los datos de entrada y salida de IVA, revisar Sección 2.1: Marco Teórico .

Para poder identificar cuál es la relación que mantienen las observaciones del rostro en conjunto con las observaciones del fondo, se aplica IVA a cada matriz de datos, fondo y rostro, de manera independiente para obtener una matriz S , para cada uno, que

contiene las señales fuente ordenadas y que mantienen la misma forma que los datasets de entrada, (N, T, K) para el rostro y $(N, T, K1)$ para el fondo.

Estas matrices S contienen los Vectores de Componentes Fuente (SCV). Estos vectores agrupan, dentro de un mismo vector, señales fuente que tienen dependencia entre sí, como algún grado de correlación, y que entre vectores maximizan la varianza multi-dimensional para asegurar independencia vectorial. Se destaca que, las señales fuente dentro de cada SCV, son las que generan las observaciones a través de una matriz de mezcla, similar a lo que sucede con ICA o PCA.

Con los SCVs estimados del rostro y el fondo, se procede a realizar un análisis de correlación de cada SCV del rostro con cada SCV del fondo. Con esto es posible identificar si existe algún grado de correlación entre las perturbaciones de iluminación del entorno que afecten directamente a las señales de rPPG obtenidas y así poder compensarlas.

El procedimiento específico de análisis se detalla a continuación:

1. Obtener los datos de las ROIs, tanto del fondo como del rostro, para obtener una caracterización general de la escena en la que se lleva a cabo la captura de la señal de rPPG. Con el fin de lograr una buena caracterización, la ROI del rostro se segmenta mediante Landmarks, obtenidos de la biblioteca Mediapipe [21] en Python [30], en donde se seleccionaron 31 sub-ROIs para la obtención de las señales de rPPG. Por otro lado, la ROI del fondo se segmenta mediante Super Pixels, SEEDS [24] específicamente, y se seleccionan 6 sub-ROIs alrededor del rostro, sin incluirlo, para obtener una caracterización espacio-temporal que tome en cuenta puntos importantes de la escena. Con las señales obtenidas de cada sub-ROI, se construyen las matrices de observaciones correspondientes a las ROIs en análisis, fondo y rostro, que tienen la forma (N, T, K) y $(N, T, K1)$, respectivamente.
2. Seleccionar las mejores sub-ROIs del rostro para la obtención de señales de color del espacio de color RGB de buena calidad. El criterio de selección se basa en la caracterización de la calidad del canal G, que es el canal que contiene la mayor información sobre la iluminación en la escena y la información pulsátil de la señal de rPPG objetivo que se proponen en los trabajos del estado del arte. La caracterización se logra utilizando el promedio de la señal temporal de cada sub-ROI junto con su valor mínimo y máximo para identificar cuál sub-ROI sirve o no. Esto se debe a que no se deben seleccionar sub-ROIs con diferencias muy marcadas entre los valores mínimos y máximos, ni que saturen la señal G. Seleccionadas las mejores sub-ROIs que cumplan con estos criterios, se forma una matriz de observaciones X de forma (N, T, M) con $M \in]0, K]$
3. Aplicar IVA al conjunto de observaciones del rostro X y del fondo, que fueron construidas en el paso anterior, de manera independiente para recuperar S , matriz de señales fuente que mantiene la misma forma de las observaciones.

4. Aplicar descomposición mediante Singular Spectrum Analysis (SSA)[10] a cada señal fuente $n \in N$ de cada sub-ROI contenida en la matriz S , tanto del fondo como el rostro. Esto origina dos matrices de la forma (N, n_decomp_face, T, K) y $(N, n_decomp_bg, T, K1)$, respectivamente, donde n_decomp_face y n_decomp_bg son el número de descomposiciones de cada señal fuente n de cada sub-ROI k . El número de descomposiciones es crítico, ya que cuantas más se apliquen, más estables serán las señales obtenidas.
5. Aplicar correlación completa entre todos los componentes de las matrices fuente descompuestas del rostro y del fondo, n_decomp_face y n_decomp_bg , respectivamente, de cada fuente n y de todas las sub-ROIs k y $k1$, del rostro y fondo respectivamente. Si la correlación entre estas señales superan el threshold predefinido, se pasa a una segunda etapa de descomposición y compensación. Esta etapa contempla una descomposición mediante Ensemble Empirical Mode Decomposition (EEMD)[33] para la señal del rostro debido a que es posible que la señal de rPPG objetivo siga enmascarada en la señal descompuesta en el paso anterior por SSA. A las descomposiciones por EEMD se le aplica análisis mediante correlación de la señal del fondo, que no fue descompuesta una segunda vez, y si supera el threshold, se elimina. Cuando se terminan de analizar y compensar estas descomposiciones, se vuelve a reconstruir de manera ascendente, reconstruir las descomposiciones de EEMD y luego las de SSA, hasta reconstruir la matriz de se señales fuente compensada S' de la forma (N, T, K) .
6. Reconstruir las observaciones compensadas (X') a través de la matriz de señales fuentes compensadas S' usando la matriz de mezcla A en el punto tres.
7. Analizar el canal G de cada sub-ROI de las observaciones reconstruidas X' .

Para evidenciar los rangos de compensación del algoritmo, se seleccionaron casos particulares de todos los conjuntos de datos en estudio. El objetivo es determinar si es factible compensar en los casos extremos de interés. Para ello, se llevaron a cabo pruebas individuales en sujetos de interés, describiendo el entorno y el proceso de captura de datos durante el transcurso del video.

3.3.1. Caso 1: Perturbaciones de iluminación

Para verificar si el algoritmo de compensación propuesto previamente logra contrarrestar las perturbaciones de iluminación, se presenta un caso de estudio en el que se produce una perturbación de iluminación tipo escalón. En este caso, alrededor del segundo 7.5, se apaga la luz de los tubos fluorescentes en la escena, dejando solo luz natural. Esto afecta directamente a las señales del fondo y del rostro, como se puede observar en la Figura 3.5.

Para caracterizar la escena, se presenta en la Figura 3.2 el primer frame del vídeo que

se encuentra segmentado por las sub-ROIs del fondo (borde rojo) y sub-ROIs del rostro (borde verde), y en la Figura 3.3 el template construido a partir de las señales del fondo y rostro obtenidas mediante rPPG.



Figura 3.2: Primer frame segmentado por sub-ROIs del fondo y del rostro.

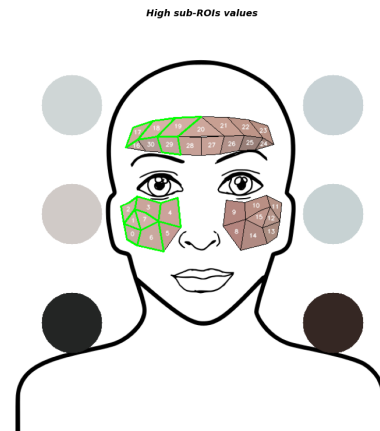


Figura 3.3: Template construido con la información recopilada mediante rPPG

En la Figura 3.4 se muestra el criterio de selección de las mejores sub-ROIs utilizando el promedio de la señal temporal junto con el análisis de borde superior e inferior de la intensidad del canal G en el tiempo. Como se mencionó en un principio, para este caso, existe una perturbación del tipo escalón que se puede para todas las sub-ROIs en los valores mínimos y máximos, que tienen una diferencia aproximada de 30 niveles de intensidad del canal G.

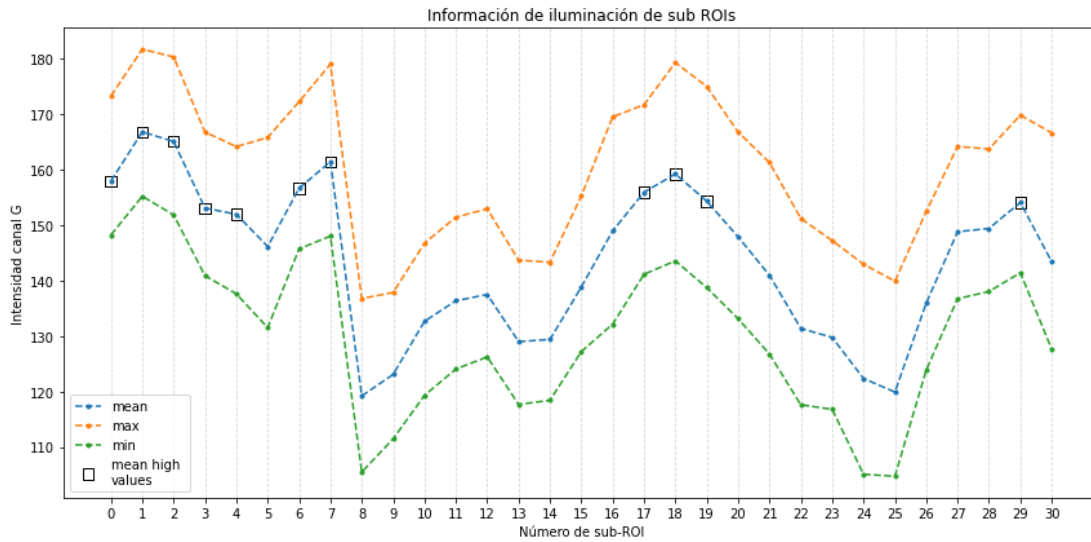


Figura 3.4: Información de iluminación de cada señal de color G de cada sub-ROI del rostro.

Para evidenciar si el algoritmo propuesto es capaz de compensar la perturbación de iluminación, se analizarán dos sub-ROIs, 1 y 2, que cumplen con la condición de tener buena iluminación y que se encuentran en la mejilla derecha(ver Figura 3.3).

Señales rPPG rostro antes de compensar

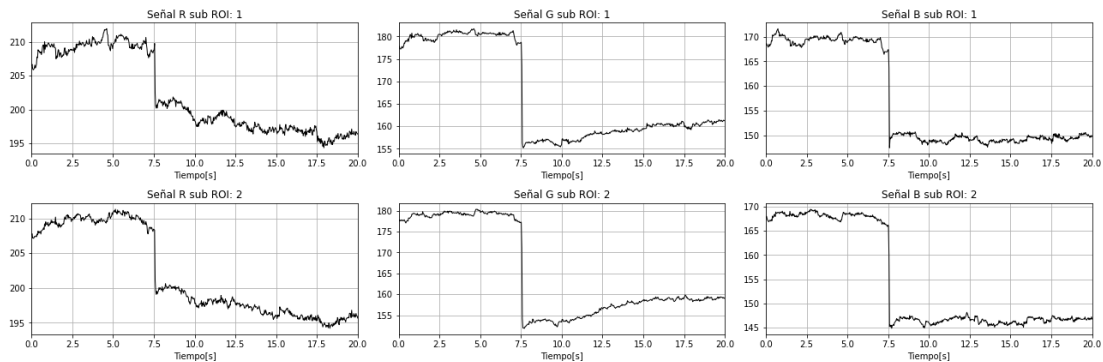


Figura 3.5: Señales RGB construidas mediante rPPG a partir de las sub-ROIs 1 y 2 que corresponden a la mejilla derecha.

En las Figuras 3.6 y 3.7, se puede apreciar que la señal fuente del rostro (gráfico superior izquierdo) presenta una alta correlación con la señal fuente del fondo (primer gráfico fila del medio). La zona de alta correlación corresponde a una zona de la pared (zona superior izquierda en Figuras 3.3 y 3.2) que es bastante homogénea y se puede inferir que la perturbación es de iluminación general y no local. Es por esto que es posible que tenga una alta correlación con la perturbación y se pueda compensar de

buena manera.

Por otro lado, el espectro en frecuencia de la señal fuente compensada resultante (gráfico inferior derecho), presenta una tendencia notoria al peak de GT del HR. Sin embargo, esta es sólo una compensación de una señal fuente, por lo que hay que ver si el efecto de tendencia al peak de GT se sigue manteniendo al reconstruir las observaciones compensadas de cada sub-ROI.

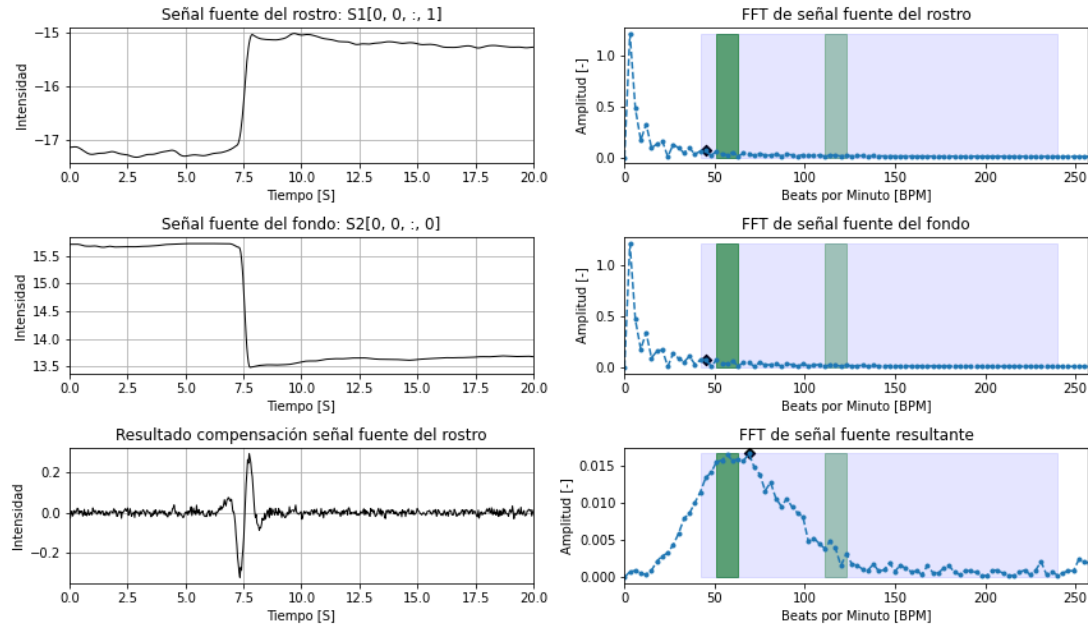


Figura 3.6: Compensación de señal fuente de sub-ROI 1.

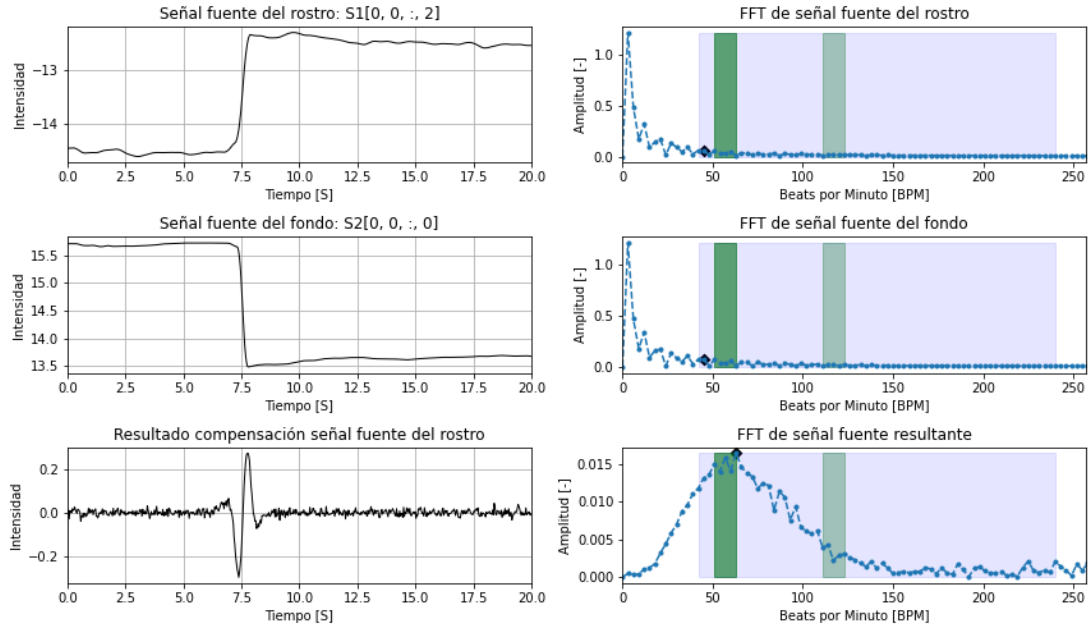


Figura 3.7: Compensación de señal fuente de sub-ROI 2.

Compensadas y reconstruidas las sub-ROIs en estudio, se presentan las señales RGB en la Figura 3.8, donde se puede ver claramente que las señales de color RGB fueron compensadas en comparación a la Figura 3.5.

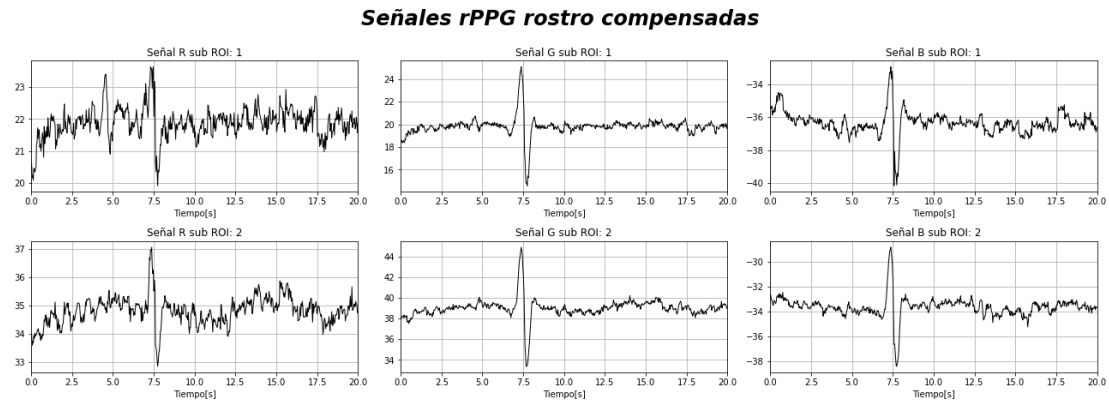


Figura 3.8: Señales RGB compensadas para sub-ROI 1 y 2.

Como sólo nos interesa el canal G para ver la señal de rPPG compensada respecto a la perturbación, se muestra en la Figura 3.9 y 3.10, las comparaciones directas de la señal G en la sub-ROI antes de compensar y compensada, junto con sus respectivos espectros en frecuencia. En estos espectros, se aprecia que la señal fue compensada de los elementos de baja frecuencia asociados a la perturbación de iluminación de tipo escalón, y además se evidencia la tendencia de la señal a la GT de HR en los dos casos.

Se destaca que se realizó un análisis de dos sub-ROIs de las 10 que fueron seleccionadas como objetivo en la etapa de selección según calidad de iluminación, para facilitar la explicación del problema. Las demás sub-ROIs también muestran claros elementos de compensación de iluminación y la tendencia clara a la GT.

Comparación canal G de sub-roi 1

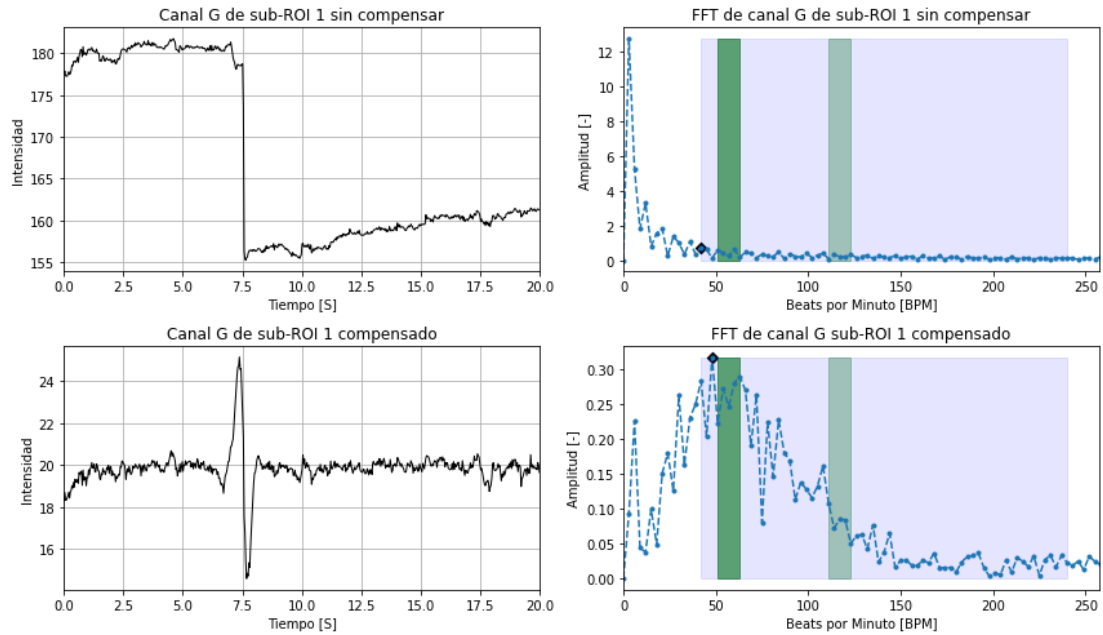


Figura 3.9: Comparación canal G sub-ROI 1

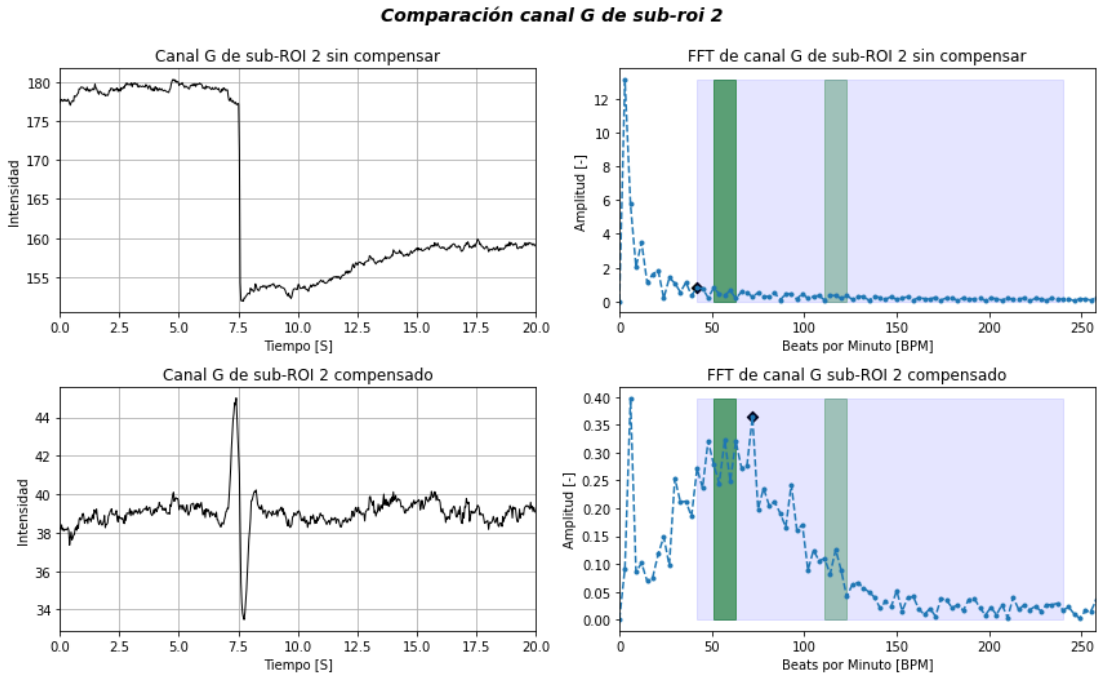


Figura 3.10: Comparación canal G sub-ROI 2

3.3.2. Caso 2: *Heart Rate* alto mayor a 120 BPM

Para verificar la capacidad del algoritmo para calcular HR superiores a 120 BPM, se realiza un análisis utilizando el video de un sujeto que presenta el HR más alto entre todos los datasets disponibles para este estudio [3][13][39]. Este sujeto registró un HR de 170 BPM después de realizar ejercicio de manera intensa.

Respecto a las condiciones ambientales y las del sujeto, en el video se observa una iluminación led estable, un fondo invariable y el sujeto mantiene la cabeza inmóvil con un dispositivo especial. Esto facilita la obtención de una señal limpia de rPPG sin la interferencia de movimientos rígidos de la cabeza.

En la Figura 3.11 se puede apreciar que la distribución de color en el template es bastante similar en el rostro, aún así, sigue marcando como buenas zonas las que están en la frente. Estas zonas, presentan los mejores valores según el criterio de selección, en la Figura 3.12, por lo que se espera una sub-ROI para ser compensada.

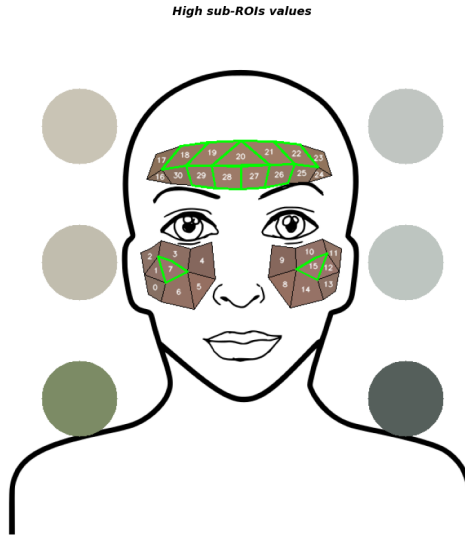


Figura 3.11: Template construido con la información recopilada mediante rPPG.

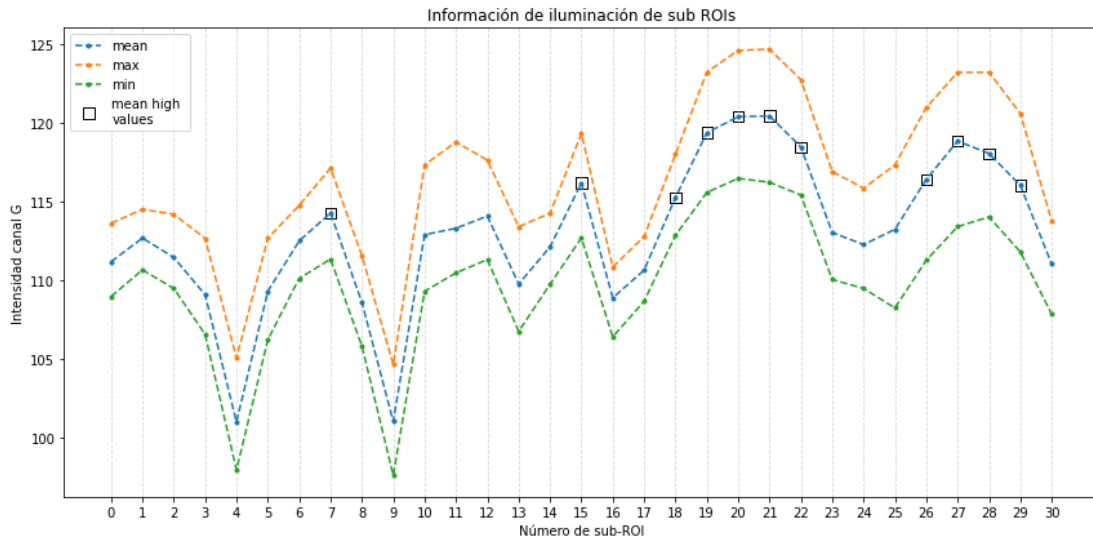


Figura 3.12: Información de iluminación de cada señal de color G de cada sub-ROI del rostro.

Respecto a la compensación, a continuación se muestra un análisis para dos sub-ROIs seleccionadas como las mejores, sub-ROI 20 y 21 que se encuentran en la frente. En las Figuras 3.13 y 3.14, se puede apreciar claramente una compensación en los componentes de baja frecuencia del rostro que son comunes con las señales del fondo. Gracias a esto, después de la compensación, se puede ver claramente la tendencia de la señal entre los 150 y 180 BPM, mostrando peaks de valores posibles para HR. Es importante notar que es posible la presencia de picos similares en un área acotada debido a que

el espectro en frecuencia pierde la característica temporal. Por lo tanto, es posible que haya pequeñas variaciones temporales que se agrupan en un solo espectro.

Comparación canal G de sub-roi 20

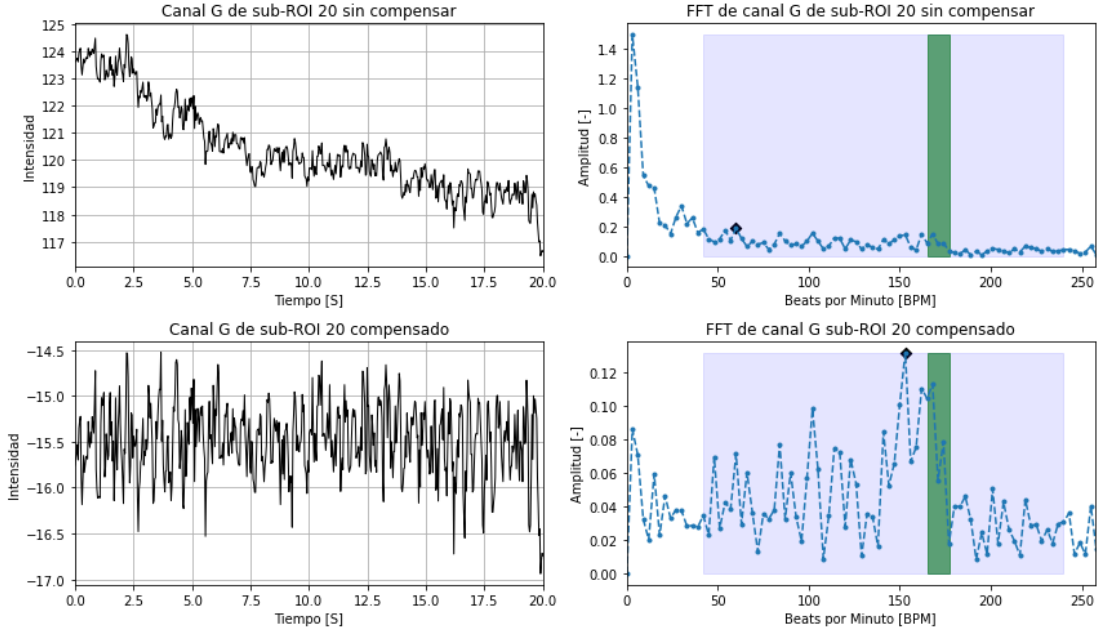


Figura 3.13: Comparación canal G sub-ROI 20.

Comparación canal G de sub-roi 21

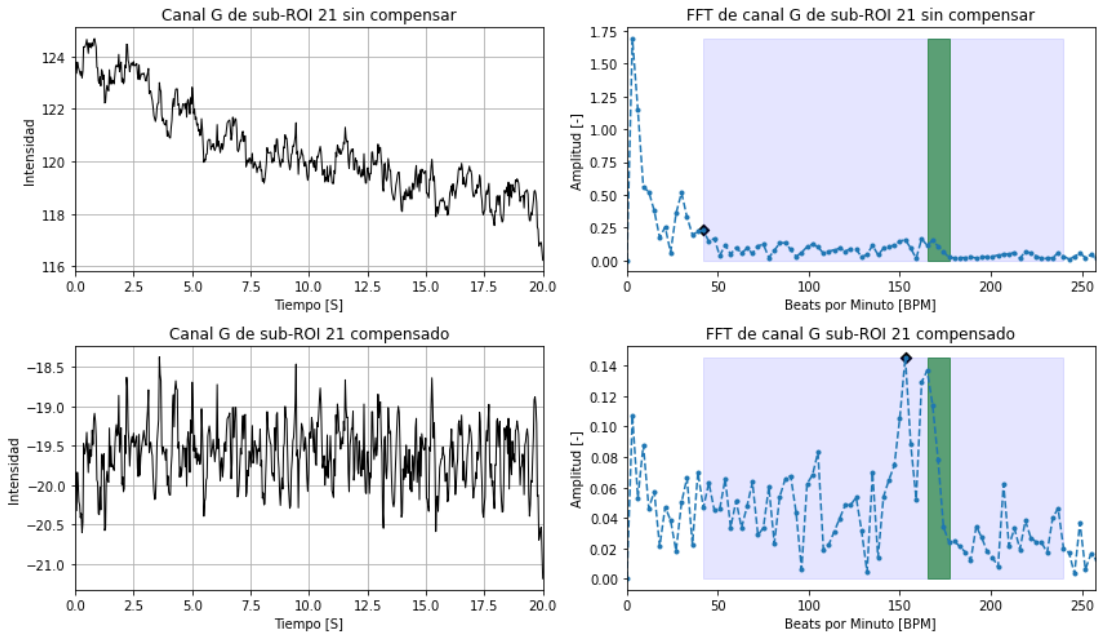


Figura 3.14: Comparación canal G sub-ROI 21.

Cálculo de HR

Esta sección se centra en la estimación de los signos vitales, en específico *Heart Rate*(HR). En primer lugar, para obtener las señales limpias de rPPG, se deben identificar en ellas cuál es el HR dominante dentro de todas las señales obtenidas a través de las sub-ROIs compensadas en el paso anterior de compensación de iluminación. Con el HR calculado, es posible hacer una retro-proyección para filtrar en el rango donde se mueve el HR y así, obtener una señal limpia de perturbaciones, lo que permite calcular los demás signos vitales: *Heart Rate Variability* (HRV), Saturación de Oxígeno (SpO2) y *Respiration Rate* (RR).

El algoritmo se presenta de manera secuencial en los siguientes pasos :

1. Se seleccionan todas las señales de color G, del espacio de color RGB, de cada sub-ROI compensada en el paso anterior. Se elige el canal G porque es el que contiene la mayor información de iluminación de la escena, así como la información pulsátil que refleja los signos vitales.
2. Aplicar filtro temporal a las señales seleccionadas en el rango objetivo de HR, que va desde los 42 [BPM] a 240[BPM], 0.7 [Hz] y 4[Hz], respectivamente.
3. Aplicar análisis espectral por ventanas a cada señal filtrada en el paso anterior. Estas ventanas tienen un tamaño de 10 segundos y presentan un solapamiento entre ellas del 90 %, lo que genera un cantidad de 11 ventanas con una distancia temporal entre ellas de 1 segundo.

La particularidad del uso de ventanas para este análisis radica en el hecho de caracterizar la señal no sólo de manera espectral, sino también mediante un análisis temporal. Esto se debe a que el HR, como se manifiesta en una señal de baja amplitud, puede ser más visible al aplicar ventanas más pequeñas, lo que revelaría su característica cuasi-periódica y facilitaría su identificación de manera automática por el algoritmo.

Con lo anterior en consideración, se identifican el o los mayores valores de magnitud de FFT de cada segmento, junto con su HR correspondiente, y se agregan a una lista general que almacena todos los posibles valores obtenidos para todas las ventanas de todas las señales.

4. Con todos los valores de HR obtenidos, se procede a realizar el cálculo del posible HR final. Este cálculo se hace mediante un histograma que se construye usando toda la información obtenida en el paso anterior, de manera que se identifican los *peaks* de este histograma y se selecciona el que contiene más repeticiones, el *peak* más alto en este caso. Seleccionado este *peak* máximo, se procede a calcular el HR de manera ponderada con el vecino anterior y posterior, para tener un característica más flexible debido a los errores que se generan al calcular el HR utilizando análisis espectral.

Con el algoritmo de cálculo de HR descrito anteriormente, se procede a mostrar los histogramas generados para los dos casos presentados en la parte de compensación de iluminación.

En la Figura 3.15, para el **Caso 1: Perturbaciones de iluminación**, se puede observar una clara tendencia hacia un valor de HR = 63 [BPM], que se repite aproximadamente 30 veces como el valor máximo en todos los segmentos procesados de todas las señales compensadas de las sub-ROIs seleccionadas. Luego, el valor que se obtiene de la ponderación con sus vecinos inmediatos es de un HR = 60[BPM] y considerando que su GT = 58 [BPM], se logra una precisión de 97 %.

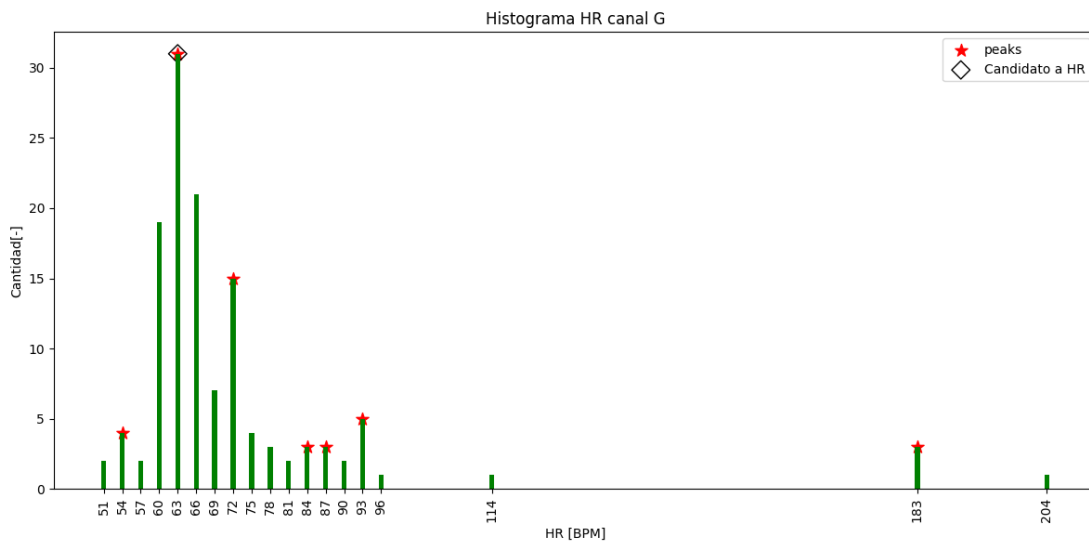


Figura 3.15: Histograma generado para Caso 1: Perturbaciones de iluminación

En la Figura 3.16, para el **Caso 2: HR alto > 120 [BPM]**, se puede observar una clara tendencia hacia un valor de HR = 165 [BPM], que se repite aproximadamente 43 veces como el valor máximo en todos los segmentos procesados de todas las señales compensadas de las sub-ROIs seleccionadas. Luego, el valor que se obtiene de la ponderación con sus vecinos inmediatos es de un HR = 166[BPM] y considerando que su GT = 170 [BPM], se logra una precisión de 98 %.

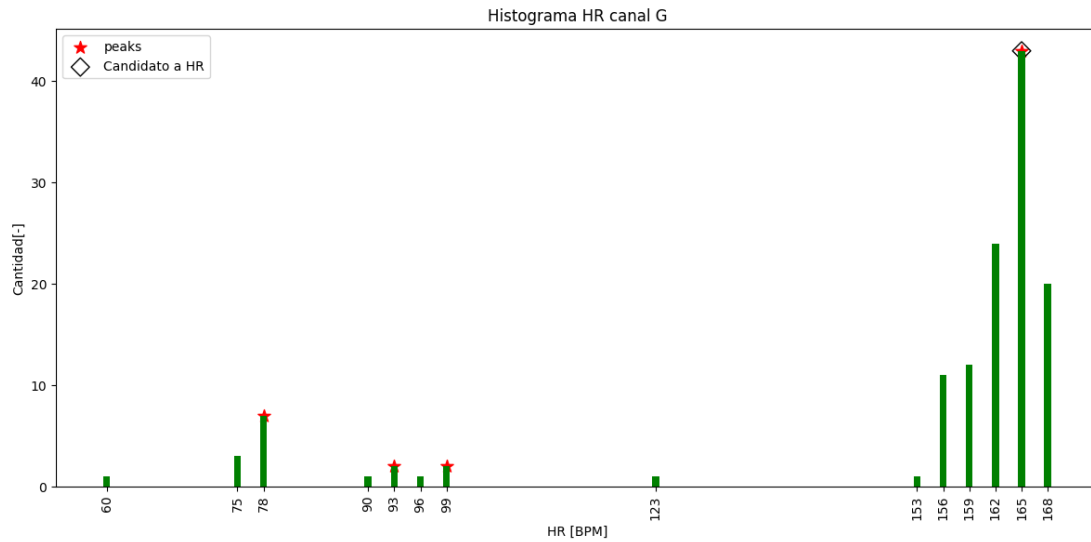


Figura 3.16: Histograma generado para Caso 2: HR alto > 120 [BPM]

3.4. Desarrollo de la cuarta metodología exploratoria

Esta metodología se basa esencialmente en las características de la señal pulsátil de la que se obtienen los signos vitales. Esta señal es continua en el tiempo y muestra los cambios de volumen de sangre en las venas. Por lo tanto, siempre debe estar presente en todo el video a partir del cual se construyen las señales mediante rPPG. Bajo esta premisa, se propone un algoritmo que consta de dos partes: selección de candidatos a HR y estimación de HR final.

Selección de candidatos a *Heart Rate*(HR)

Para abordar la problemática de las perturbaciones de iluminación, se propone un enfoque de análisis de señales de rPPG a distintos niveles de detalle, en relación con el área utilizada en comparación con la ROI total del rostro para generar estas señales. Este enfoque se basa en el hecho de que al promediar una mayor cantidad de píxeles de la ROI, se elimina el ruido espacial y/o temporal que no afecte de manera general a toda la zona del rostro, resaltando así los componentes intrínsecos de la ROI, que son la señal pulsátil de donde se puede calcular el HR, y , en caso de existir, los componentes de ruido que afectan a toda la zona. Por otro lado, se encuentra el análisis de una sub-ROI específica, que es el nivel de detalle más preciso que se utiliza en este trabajo, donde se puede seleccionar una zona que cumpla con características específicas que sean propicias para obtener una señal de rPPG limpia en términos de ruido. Además, para intentar maximizar la posibilidad de encontrar señales de rPPG limpias en los distintos niveles de detalle, se agrega un paso previo de selección el 80 % de las mejores

sub-ROIs según criterio de puntajes que se calculan a partir del promedio, desviación estándar y diferencia entre máximo y mínimo en la señal.

Con las mejores sub-ROIs seleccionadas, se procede crear los distintos niveles de detalle y a compensar la iluminación usando el algoritmo presentado en la Metodología Exploratoria 3 (Sección 3.3), obteniendo así señales compensadas. Luego, a todas estas señales se les aplica enventanado con *stride* de 1[s] y tamaño de ventana de 10[s], y se procede a la identificación de caminos¹ usando criterio de continuidad. Estos caminos se consideran candidatos a HR y se les asigna un puntaje que depende de cuatro factores: promedio, desviación estándar, diferencia entre valor máximo y mínimo, y magnitud relativa del *peak* en frecuencia de la ventana de análisis.

El algoritmo específico de esta sección se describe a continuación:

1. Abrir señales de color de cada sub-ROI del rostro y las señales del fondo.
2. Seleccionar las mejores sub-ROIs según criterio de puntajes utilizando sólo el canal G del espacio de color RGB. Estos puntajes se crean a partir del promedio, desviación estándar y diferencia entre máximo y mínimo global de la señal de color. En particular, este último criterio es para descartar las zonas que tengan cambios muy bruscos de iluminación, que pueden ser originados por movimientos de la cabeza, personas u objetos en movimiento que perturban la iluminación de las sub-ROI en una porción de tiempo. Estos cambios bruscos generan perturbaciones en gran parte del espectro posible de HR, por lo que es necesario eliminarlos.
3. Con las sub-ROIs seleccionadas anteriormente, se proponen tres niveles de análisis para caracterizar qué tan buena es la señal de rPPG en el rostro. Estos niveles permite identificar al detalle la calidad de la señal de rPPG objetivo.
 - *All mean*: Todas las señales de todas las sub-ROIs se promedian, generando sólo una señal por canal de color que tiene la forma general (N, T, K) , donde N es el número de canales de color, RGB en este caso; T es el número total de muestras de las señales y K es el número de sets de señales, quedando la forma característica para este conjunto de señales como $(3, T, 1)$.
 - *Original ROIs*: Se agrupan las sub-ROIs dependiendo de la zona donde pertenecen: mejilla derecha, mejilla izquierda o frente. La forma particular de este conjunto de señales es $(3, T, 3)$.
 - *Optimal ROIs*: Se seleccionan las k mejores señales del conjunto de sub-ROIs seleccionadas. La forma particular de este conjunto de señales es $(3, T, k)$, donde $k \in \{3, 5\}$.

¹El termino de "caminos" se refiere a *peaks* continuos al poner el espectro de las ventanas una al lado de otro.

Se destaca que, a todos los niveles de análisis se le aplica la compensación de iluminación correspondiente, siendo este paso el que ocupa más tiempo de procesamiento.

4. Se aplica enventanado a todas las señales de los tres niveles de análisis, para permitir en los pasos siguientes un análisis espectro-temporal del comportamiento de estas señales.

5. Análisis de señales enventanadas e identificación de candidatos:

En este paso, se realiza un análisis exhaustivo de cada señal utilizando un enfoque espectro-temporal para examinar todas las ventanas de la señal en cuestión. En primer lugar, se aplica el método de Welch para obtener una caracterización espectral de todas las ventanas. Mediante el método de Welch, es posible identificar en qué frecuencias se concentra la energía y gracias a esto es posible identificar todos los peaks de frecuencia en todas las ventanas. Luego, con todos los peaks identificados de todas las ventanas, se procede a buscar, según un criterio de continuidad, cuáles son los peaks válidos que se mantienen a lo largo del tiempo (a través de las ventanas), generando arreglos de candidatos de HR que almacenan la posición de los peaks junto con un puntaje que depende de cada ventana espectral y que tiene que ver con la magnitud relativa de los peaks en esa ventana.

El criterio de continuidad se basa en la característica intrínseca de la señal pulsátil que permite identificar los signos vitales, siempre presente, por lo que el criterio busca la continuidad espectral de esta señal a través de las ventanas desfasadas temporalmente. En la Figura 3.17, se aprecian en color amarillo los caminos identificados por continuidad como posibles candidatos a HR. Sin embargo, se pueden encontrar caminos que no abarcan la totalidad de las ventanas. En este caso, el criterio busca aquellos que tengan una secuencialidad, aunque sea en las ventanas finales. Sin embargo, esta secuencialidad incompleta se penaliza rellenando los valores faltantes con valores nulos, lo que perjudica el puntaje que se calcula en el paso siguiente.

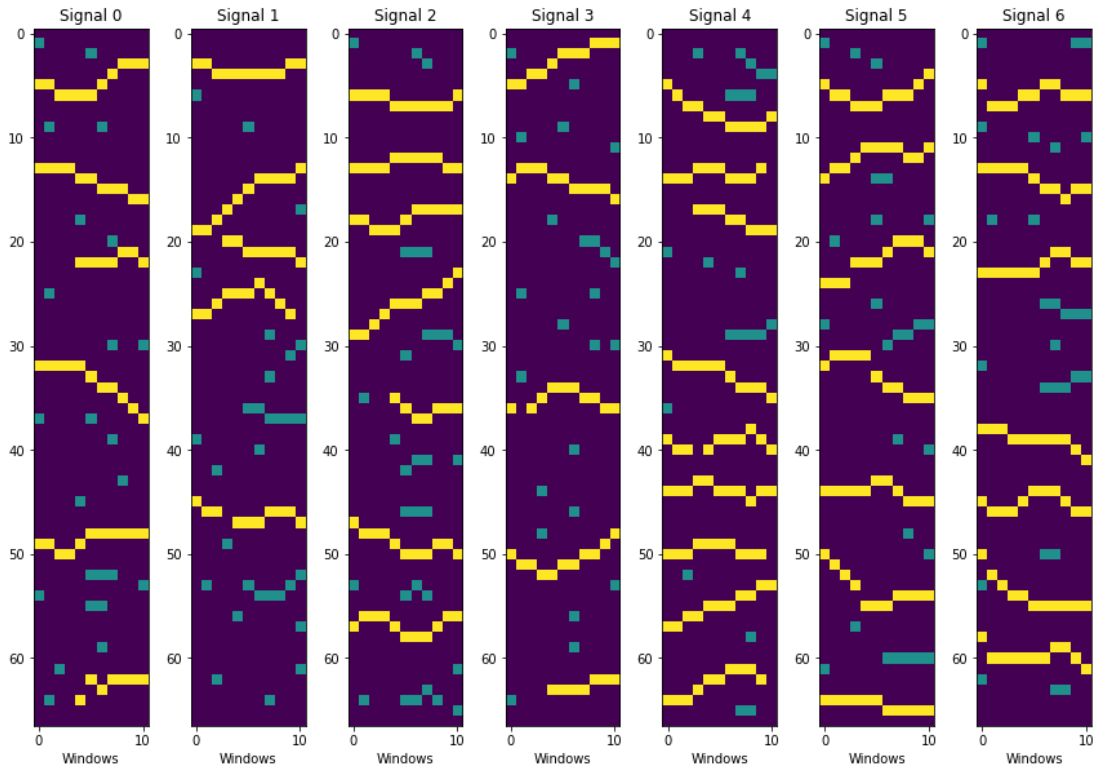


Figura 3.17: Ejemplo del criterio de continuidad, los "caminos" de color amarillo son los identificados por el algoritmo.

Con los candidatos identificados junto a los puntajes correspondientes de cada *peak*, se procede a calcular los puntajes correspondientes de cada candidato utilizando un enfoque espectro-temporal, o sea, calcular un puntaje a través de las ventanas para caracterizar el comportamiento en el tiempo de estos *peaks* que componen un candidato. En particular este puntaje se calcula usando el promedio, la desviación estándar relativa, la diferencia entre máximo y mínimo valor de magnitud relativa junto con la diferencia relativa entre un *peak* y otro en términos de magnitud. Entonces, una vez calculados todos estos valores, se ponderan y se obtiene el puntaje general de cada candidato.

6. Procesar candidatos e identificar *Heart Rate* (HR):

Se seleccionan los mejores puntajes que superan un *threshold* de porcentaje relativo al máximo valor obtenido. Esto se hace con el fin de garantizar que los valores candidatos de HR sean representativos y no produzcan falsos resultados. Para obtener una caracterización general de HR, se ponderan los candidatos según su puntaje específico, expresado como un número entero. Por ejemplo, si el puntaje de un candidato es 0.812, se le asigna un valor de 81 y se agregan 81 veces los valores de ese candidato a la frecuencia cardíaca a un arreglo general que contiene todos los candidatos ponderados. Esto permite construir un histograma

que proporciona información sobre el comportamiento de la frecuencia cardíaca.

Estimación de *Heart Rate*(HR) final

Trabajar con una metodología robusta para el cálculo de HR es fundamental en entornos con muchas perturbaciones. En ambientes no controlados, es difícil obtener una señal fiable para HR. Por lo tanto, se propone lo siguiente:

1. Crear un histograma con el arreglo donde se encuentran todos los candidatos a HR ponderados.
2. Una vez creado el histograma, se procede a identificar todos los *peaks* presentes utilizando el criterio de tres puntos. Este criterio implica el uso de una ventana deslizante que abarca tres posiciones. Se identifica un *peak* cuando se cumplen las siguientes condiciones a la vez: diferencia entre el borde inferior de la ventana y el centro es negativa, y la diferencia entre el borde superior y el centro también es negativa. En ese caso, se considera que hay un *peak* en la posición donde se encuentra el centro de la ventana.
3. Identificados todos los *peaks* del histograma, se procede a seleccionar los mejores en términos de repeticiones respecto a un *threshold* de 85 % relativo al mayor *peak*. Si existe un *peak* único dentro de este margen, se considera como el *peak* final. En caso de que haya dos o más *peaks* dentro del margen, se establece acumulación en una vecindad que depende de una holgura predeterminada, se aplica nuevamente el criterio del umbral y, si aún hay similitudes, se repite el proceso aumentando la holgura y elevando el valor del umbral hasta que sólo quede un candidato que cumpla con la condición.

Ahora que el algoritmo está completamente descrito, se procede a realizar el análisis pertinente respecto a los dos casos de interés.

Para la evaluación de los siguientes casos, se utilizaron nueve señales de análisis², una señal por parte de *all mean*; tres señales por parte de *original ROIs* y cinco señales de *optimal ROIs*, sumando en total nueve señales enumeradas del 0 al 8 en la Figura 3.18, 3.19, 3.21 y 3.22, respectivamente. Además, en estas figuras, se muestra una representación espectro-temporal de cada señal analizada. En el eje horizontal se presentan las ventanas de análisis, mientras que en el eje vertical se muestran las posiciones de los *peaks* en frecuencia de cada ventana. Esta representación permite visualizar de manera clara los patrones generados por el algoritmo de selección de candidatos por continuidad.

²Es importante recordar que se selecciona sólo el canal G de cada nivel de análisis.

3.4.1. Caso 1: Perturbaciones de iluminación

Para este caso, como se abordó en las metodologías exploratorias 2 y 3, se presenta una perturbación de iluminación en el segundo 7.5 aproximadamente. En particular este tipo de perturbación, tipo escalón, generan en frecuencia una serie de armónicos que tienen gran presencia en el rango posible de HR, 0.7 a 4 [Hz]. Con esto en consideración, aunque se promedien todas las sub-ROIs planteadas, como en *all mean*, no es posible suprimir esta perturbación ya que afecta a la escena completa y el promediar las zonas sólo resaltarán los componentes comunes, como la señal pulsátil de signos vitales y la perturbación temporal del ruido.

Por otro lado, siguiendo el análisis del caso, en la Figura 3.18 se muestran todos los candidatos posibles de HR según criterio de continuidad de *peaks*. Estos caminos generados pueden abarcar todas las ventanas o un porcentaje de estas, pero manteniendo la continuidad. Sin embargo, los caminos mas cortos, reciben una penalización al no ser continuos en todas las ventanas a través del tiempo, por lo que, al calcular los puntajes individuales de cada candidato a HR, la mayoría quedará descartado dependiendo del puntaje obtenido.

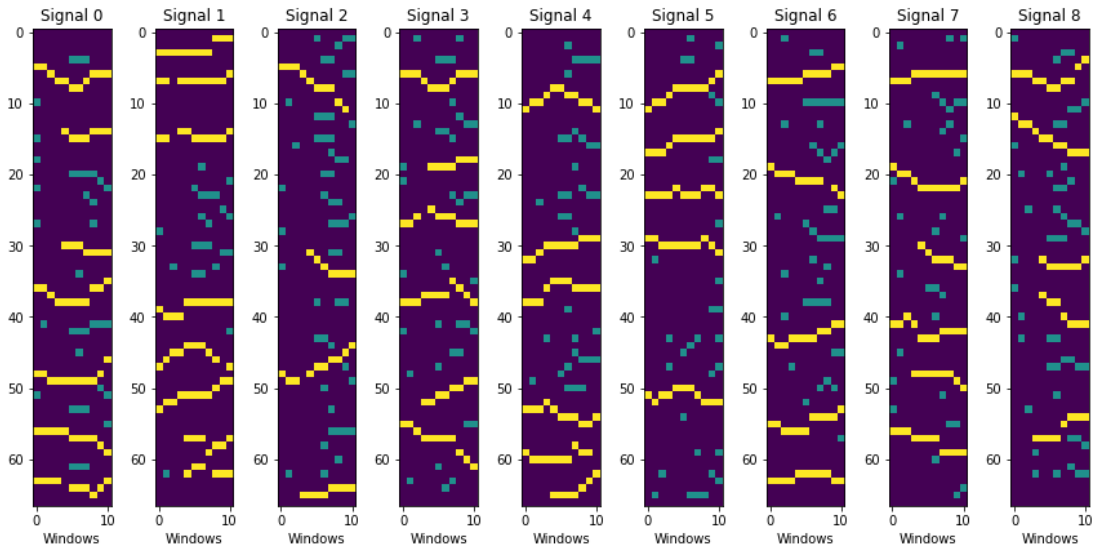


Figura 3.18: Candidatos a HR.

Una vez calculados los puntajes de todos los candidatos a HR, se selecciona el 20 % de los mejores puntajes en relación con el máximo valor entre todos los candidatos. Se espera que estos puntajes seleccionados contengan la información relevante sobre HR, como se muestra en la Figura 3.19, donde se observa que la mayoría de los candidatos seleccionados se encuentran en las posiciones cercanas a 4 y 10, correspondientes a 54 y 72 BPM, respectivamente. Se conoce que la *Ground Truth* (GT) del sujeto en cuestión es de 57 BPM, lo que se aproxima a la posición 5 y que calza con el rango donde se

mueven la mayoría de los candidatos a HR, por lo que la selección de candidatos según puntajes, hasta el momento, ha mostrado que puede identificar los *peaks* de la señal objetivo.

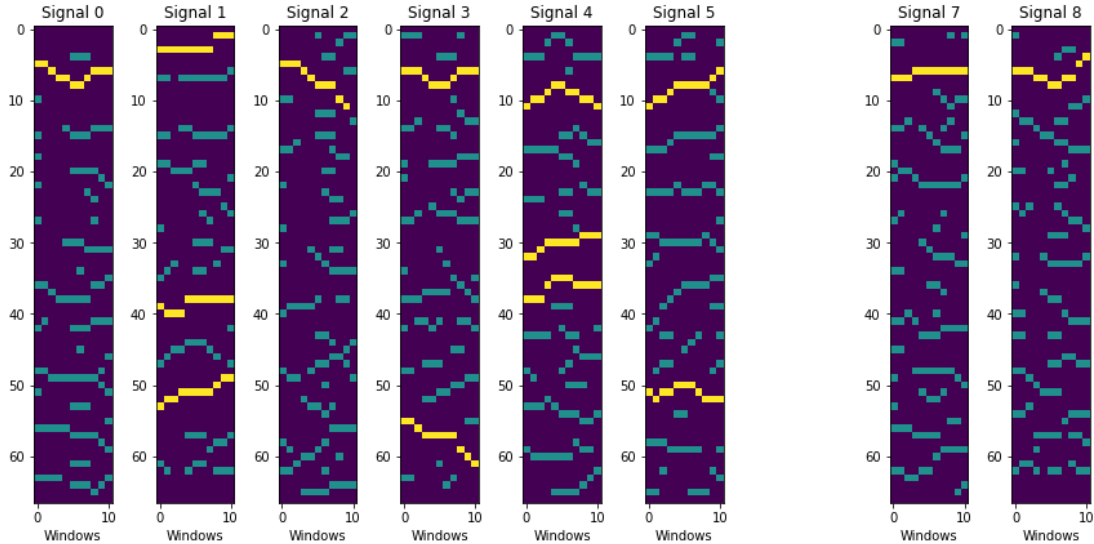


Figura 3.19: Candidatos a HR seleccionados.

Con los mejores candidatos seleccionados, se construye un histograma utilizando los valores ponderados de los candidatos según el puntaje obtenido en las etapas anteriores. En la Figura 3.20 se puede apreciar una clara tendencia de los candidatos a mostrar el HR estimado en 60 BPM, junto con una pequeña vecindad de ± 10 BPM, lo cual es esperable debido a la resolución espectral utilizada. Además, se puede identificar el valor de GT en 57 BPM. Utilizando el criterio de cálculo de HR con los vecinos inmediatos dentro de una holgura establecida de aproximadamente 3 BPM, que corresponde a la diferencia mínima entre dos valores consecutivos en el espectro posible de HR debido al tamaño de ventana, se obtiene un HR estimado de 61 BPM con un error del 4.91 %, valor bastante aceptable debido a la naturaleza de la perturbación de iluminación.

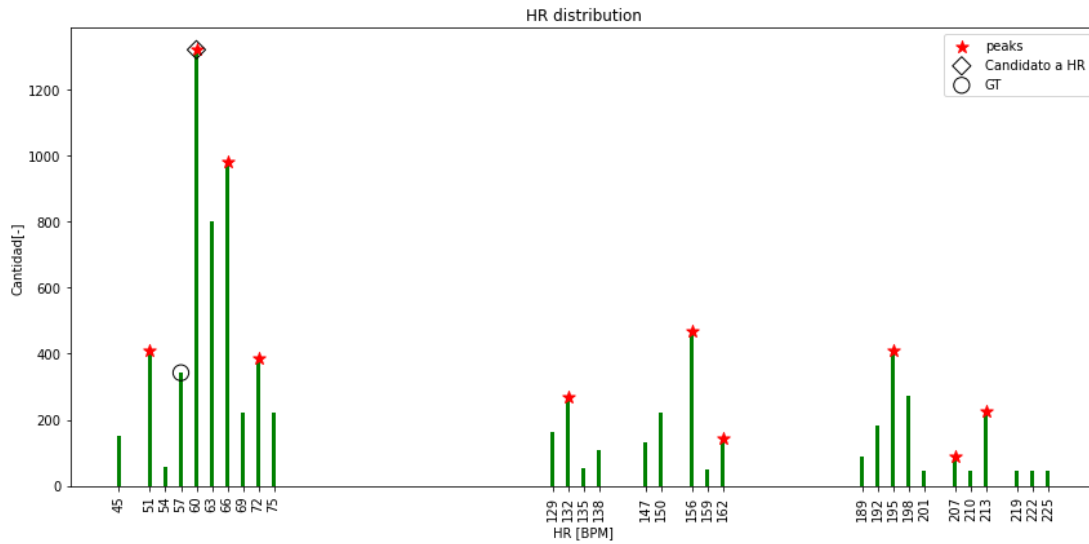


Figura 3.20: Histograma construido a partir de los candidatos a HR seleccionados.

3.4.2. Caso 2: *Heart Rate* alto mayor a 120 BPM

Este caso es de suma importancia debido a que es el mayor valor de HR dentro de los Datasets. La toma del video es en un entorno de laboratorio con condiciones de iluminación estables y movimientos de cabeza mínimos debido a un sistema de sujeción especial. En particular, al realizar el análisis pertinente en la Figura 3.21, se puede apreciar que los candidatos de HR son bastante estables en comparación al caso anterior, Figura 3.18. Esto se debe en particular a las condiciones de la toma de datos.

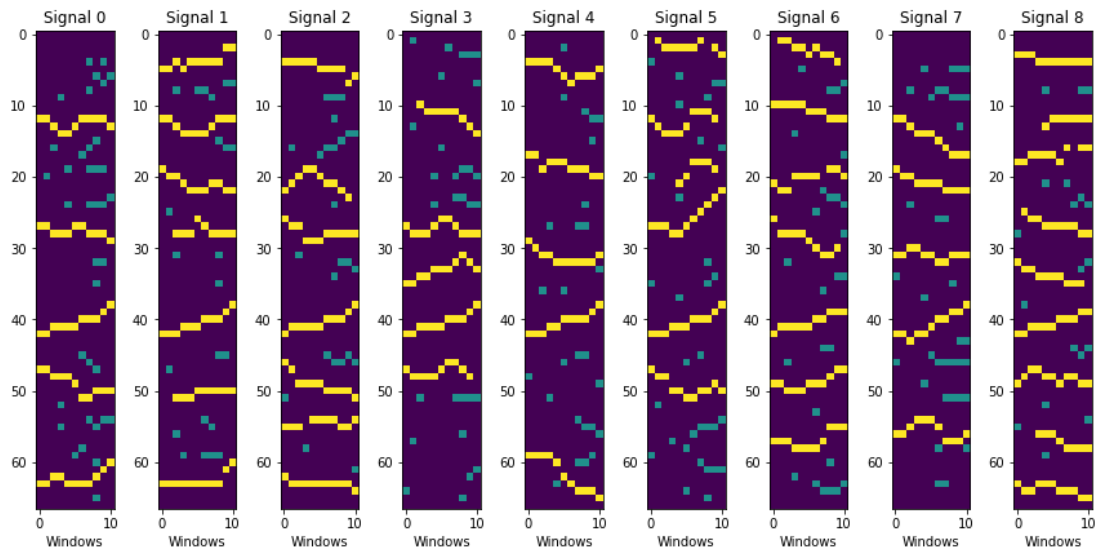


Figura 3.21: Candidatos a HR.

Se calculan los puntajes respectivos de cada candidato de HR de la Figura 3.21 y se selecciona el 20 % de los mejores puntajes en relación con el máximo valor entre todos los candidatos. Los candidatos que cumplen con esta condición se pueden apreciar en la Figura 3.22, donde se ve claramente que todas las señales presentan el mismo candidato en los mismo rangos de HR. Es importante destacar que, dado que el video está grabado después de una sesión intensa de ejercicio, el HR tiende a bajar hasta la estabilidad en reposo. Por esta razón se puede observar un comportamiento decreciente en las posiciones de los peaks a través de las ventanas de tiempo. En específico, durante los 20 [s] el sujeto presenta una diferencia de 5 posiciones espectrales, pasando de 168 a 156 BPM.

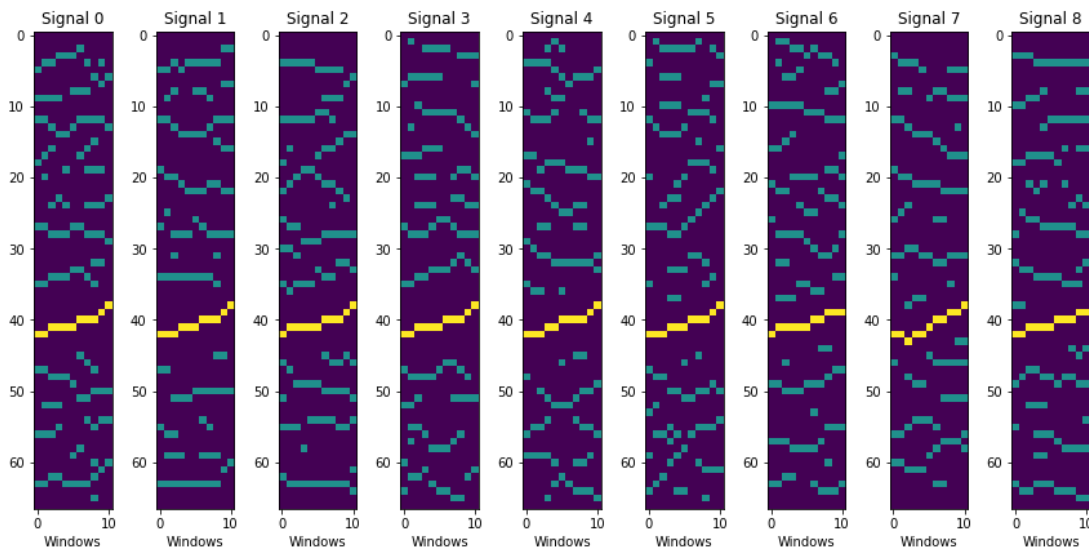


Figura 3.22: Candidatos a HR seleccionados.

Con los mejores candidatos seleccionados, se construye un histograma utilizando valores ponderados de los candidatos según el puntaje obtenido en las etapas anteriores. En la Figura 3.23 se puede apreciar una clara tendencia de los candidatos a mostrar el HR estimado en 165 BPM, junto con una pequeña vecindad, que son los valores que va tomando durante los 20 [s] del video. Además, se puede identificar el valor de HR estimado en los valores del histograma, donde la tendencia es el valor de 165 BPM y GT es 170 BPM, pero como los valores posibles de HR dependen de la resolución espectral, en el gráfico se puede ver GT con un valor de 171 BPM.

Utilizando el criterio de cálculo de HR con los vecinos inmediatos dentro de una holgura establecida de aproximadamente 3 BPM, que corresponde a la diferencia mínima entre dos valores consecutivos en el espectro posible de HR debido al tamaño de ventana, se obtiene un HR estimado de 165 BPM con un error del -3.05 %, valor bastante aceptable debido a la naturaleza del comportamiento de HR después de ejercicio intenso, que tiende al valor de reposo, disminuyendo HR y presión arterial.

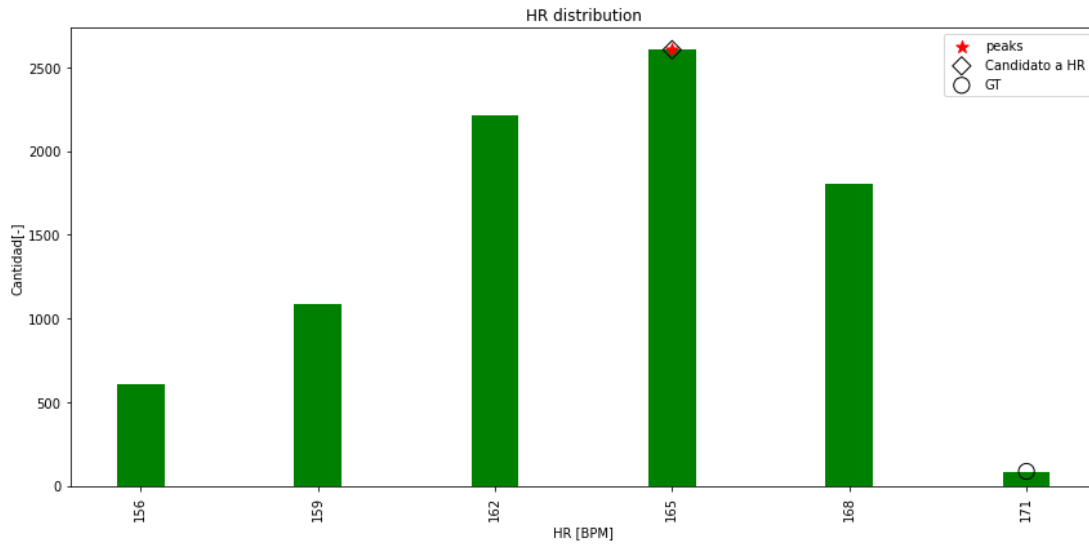


Figura 3.23: Histograma construido a partir de los candidatos a HR seleccionados.

Durante el desarrollo de esta metodología, se exploraron diferentes enfoques para calcular un HR robusto. En un primer momento, se empleó únicamente la metodología de compensación de iluminación para corregir las señales y luego calcular el HR de manera directa basándose en el espectro de las señales seleccionadas, como en la metodología exploratoria 3 (Sección 3.3). Sin embargo, esta metodología por sí sola no logró proporcionar una señal donde se pudieran estimar correctamente los valores de HR. Sin un método robusto de estimación, en entornos no controlados donde existe mucho ruido en las señales debido a movimientos de cabeza, gestos faciales y/o perturbaciones de iluminación, la información de HR se ve afectada de manera muy agresiva en comparación con la información de las perturbaciones, debido a que estas, por lo general, contienen más energía.

Capítulo 4

Validación

En primer lugar, se deben presentar los Datasets que se utilizarán para la validación de los métodos desarrollados anteriormente:

1. **UBFC-rPPG Dataset 2** [3]: Este Dataset está compuesto por 42 videos con una resolución de 640x480 píxeles a 30 FPS, sin codec de compresión, por lo que cada *frame* viene en un tamaño de $640 \times 480 \times 3 = 2.4576$ MB. Cada video está creado a partir de sujetos distintos con un tono de piel claro y cada video tiene una señal de PPG como *Ground Truth* (GT) para la obtención de signos vitales. Por otro lado, la escena donde se desarrolla la toma de los videos es bastante estable en términos de condiciones de iluminación. Los sujetos se encuentran aproximadamente a 1 [m] de distancia de la cámara, mientras que el fondo de la escena, es un pantalla verde opaca. Durante la grabación de los videos, los sujetos están jugando un juego matemático que pretende hacer variar en el tiempo el HR de los participantes, y esta particularidad es la única condición variable en este dataset ,ya que, las condiciones de interés a compensar como perturbaciones de iluminación, no presentan gran desafío. respecto al entorno donde se desarrolla la toma de muestras, debido a la no presencia de perturbaciones de iluminación. Se destaca que no es posible agregar imágenes de la escena debido a limitaciones impuestas por los autores del Dataset.
2. **PBDT-rPPG**[13]: Este Dataset contiene 21 videos con una resolución de 1080x640 píxeles a 30 FPS, usando H.264 como codec de compresión. La totalidad de los videos están generados a partir de tres sujetos, con distintos tonos de piel: claro, intermedio y oscuro, y cada video tiene una señal de Electrocardiograma (ECG) asociada como GT para la obtención de signos vitales. En la parte común de este Dataset, estos sujetos son grabados bajo cinco niveles de iluminación, partiendo del más oscuro (nivel 1) al más claro (nivel 5),

generando así una variedad de videos bajo distintos tipos de iluminación lo que permite, considerando los distintos tonos de piel, abarcar una amplia variedad de escenarios para la evaluación de los algoritmos de rPPG. Por otro lado, en los demás videos, se somete al sujeto 1 (tez clara) a distintas perturbaciones de iluminación, movimientos rígidos de la cabeza y a ejercicio intenso.

Respecto a la escena donde se desarrolla la obtención de los videos, este es un entorno estable, con superficies distintas, no un color homogéneo como en UBFC, sino más bien un entorno con distintos elementos, muebles, relieves varios, etc. Cada sujeto tiene fija la cabeza, en la mayoría de los casos, con la ayuda de un dispositivo de fijación especial y encontrándose a 1[m] aproximadamente de la cámara. Se destaca que no es posible agregar imágenes de la escena debido a limitaciones impuestas por los autores del Dataset.

3. **NAM-EDGE** [39]: Este dataset está compuesto por 69 vídeos con una resolución de 2160x3840 píxeles a 30 FPS, usando H264 como codec de compresión. Gracias a la colaboración de 10 funcionarios del laboratorio WildSense, se recopilaron videos antes y después de actividad física intensa. Estos videos cuentan con una GT generada a partir de dispositivos de auscultación médica, proporcionando valores únicos para el HR, la saturación de oxígeno en la sangre (SpO2) y la presión arterial. A diferencia de los conjuntos de datos anteriores, este conjunto no incluye señales de PPG y ECG como GT.

Respecto a la escena donde se recopilan los datos, se pueden observar en algunos videos perturbaciones de diferentes tipos, destacándose movimientos que afectan la captura adecuada de la señal de rPPG y perturbaciones de iluminación variadas que le otorgan realismo a la escena. Es importante destacar que, algunos funcionarios, ocultan parte de las sub-ROI objetivo, mejillas y frente, debido a pelo que oculta la frente y al uso de lentes.



Figura 4.1: Ejemplo de escena y sujetos que componen a Dataset: NAM-EDGE.

A continuación, se muestra una tabla resumen sobre los Datasets descritos anterior-

mente.

Datasets	Sujetos	Cantidad videos	FPS	Resolución	Codec	Características		
						Cambios iluminación	Movimiento	Tono de piel
UBFC	42	42	30	640x480	-	-	-	-
PBDT-rPPG	3	21	30	1080x640	H.264	X	X	X
NAM-EDGE	10	69	30	2160x3840	H.264	X	X	-

Cuadro 4.1: Tabla comparativa de características entre Datasets

Para lograr obtener una caracterización general, tanto de la escena como el rostro, donde se desarrolla la obtención de los videos que sirven para obtener las señales de rPPG, se presentan las siguientes segmentaciones que se aplican frame a frame a un video de 20 segundos, con un valor de Frames por Segundo (FPS) típico de 30.

- **Original ROIs:** Se encuentran dentro de la ROI del rostro, específicamente se compone de tres sub-ROIs que son: mejilla izquierda, mejilla derecha y frente.
- **All ROIs:** Se encuentran dentro de la ROI del rostro y son segmentaciones más pequeñas de las sub-ROIs presentadas en *Original ROIs*. Cuenta con 8 segmentaciones por mejillas y 15 en la frente, dando un total de 31 sub-ROIs que permiten una caracterización espacio-temporal más precisa de la escena en términos de iluminación y perturbaciones.
- **Background ROIs:** Estas sub-ROIs se encuentran fuera del rostro, incluyendo el fondo y, en algunos casos, prendas de vestir. Se obtienen mediante la segmentación de la escena a través de *Super Pixels: SEEDS* [9], lo que permite agrupar áreas con características similares. Se seleccionan de entre todas las zonas generadas sólo seis, tres a cada lado, para garantizar una caracterización equilibrada de la escena.

Con las segmentaciones anteriores, se procesan los videos de dos formas, utilizando *Original ROIs + Background ROIs* y *All ROIs + Background ROIs*, como se puede apreciar en las Figuras 4.2 y 4.3 respectivamente.

Con los Datasets presentados y las segmentaciones del rostro y la escena caracterizadas, se procede presentar las características relevantes de los métodos desarrollados en la Metodología Exploratoria 2 (ME2) y Metodología Exploratoria 3 (ME3), junto con las características del trabajo previo [39] en la Tabla 4.2. En esta tabla se puede ver el funcionamiento principal de cada método a grandes rasgos y para entenderla mejor, se explican los tópicos principales:

- **Métodos:** Este tópico toma en cuenta los tipos de métodos para la obtención de una señal de rPPG limpia, sin perturbaciones. Bajo este concepto, se agrupan en dos grupos:

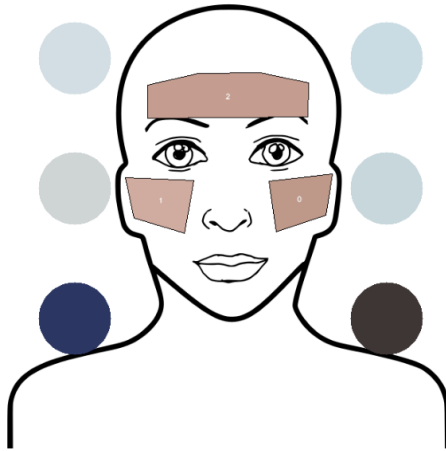


Figura 4.2: *Original ROIs + Background ROIs.*

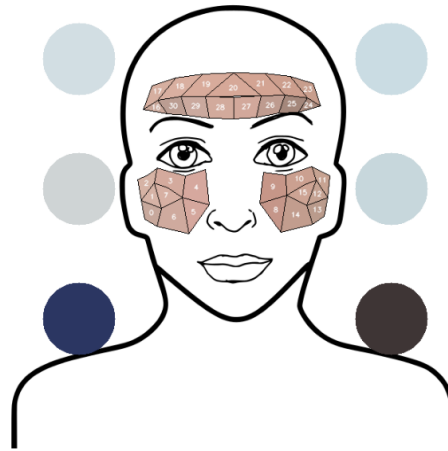


Figura 4.3: *All ROIs + Background ROIs.*

- **Robustos:** Estos métodos utilizan una serie de combinaciones de técnicas algorítmicas para contrarrestar el efecto particular de la iluminación sobre la señal de rPPG, necesaria para el cálculo de los signos vitales bajo condiciones de entorno realistas.
- **Tradicionales:** Estos métodos utilizan algoritmos específicos para obtener una señal de rPPG limpia. Algunos se basan en BSS(PCA e ICA), mientras que otros se fundamentan en las características intrínsecas del comportamiento de la superficie de la piel en relación con su interacción con la iluminación (POS, CHROM).
- **Soporte:** Este tópico toma en cuenta el funcionamiento interno al procesar las señales de color obtenidas de cada sub-ROI del rostro y/o fondo una vez entran al método. Se destacan dos a grandes rasgos:
 - *All mean:* Se refiere a que, una vez que entran las señales, se promedian todos los canales de todas las sub-ROIs, independiente del número. Lo que genera sólo las tres señales RGB como información condensada de todas las sub-ROIs junto a una pérdida de características importante. Se gana por el lado de atenuar los componentes de "ruido", pero se pierde en cuanto a características espaciales y/o temporales.
 - *sub-ROIs:* Se refiere a que procesa de manera independiente cada sub-ROI de entrada, lo que ayuda a obtener información relevante sobre la información espacio-temporal contenida en el rostro, al mismo tiempo que mantiene la información no relevante, que puede considerarse como "ruido".

- **Señales fondo:** Se refiere a si el método toma en cuenta el entorno para la compensación de perturbaciones de movimiento y/o iluminación.
- **Obtención de HR:** Al método principal del cálculo de HR. En especial se pueden identificar tres:
 - FFT :Utiliza como eje principal el calculo de HR identificando el mayor valor en magnitud del espectro en frecuencia al aplicar Fast Fourier Transform(FFT).
 - Peak detection: Utiliza como eje principal el cálculo de HR mediante contar *peaks* de una señal obtenida mediante el procesamiento del método en cuestión.
 - Propio: Utiliza algoritmos propios de cálculo de HR que pueden utilizar varias herramientas de procesamiento de señales (DSP) y/o matemáticas. Se destaca que también usan FFT, en algunos casos, como base de análisis espectral, lo que recae en herramientas de DSP.
- **ROIs a procesar:** Se refiere al tipo de sub-ROIs que se van a procesar: *Original ROIs* y *All ROIs* que se pueden apreciar en las Figuras 4.2 y 4.3 respectivamente. Se destaca que, los métodos que tienen **Soporte: All mean**, sólo se evalúan con *Original ROIs*, ya que es redundante que se evalúen con *All ROIs* debido a que cubren las mismas zonas de manera general.

A continuación, se muestran las tablas comparativas entre los métodos Robustos y Tradicionales, en las Tablas 4.2 y 4.3 respectivamente.

Métodos	Soporte		Señales fondo	Obtencion HR	Señales a procesar	
	<i>All mean</i>	<i>sub-ROIs</i>			<i>Original</i>	<i>All</i>
Robustos						
IVA+EEMD	X		X	Peak detection	X	
IVA + Skin reflection	X		X	FFT	X	
PLS + MEMD	X		X	Peak detection	X	
EEMD + MCCA		X		FFT		X
RPCA + EEMD		X		Propio		X
IVA+SSA		X	X	Propio		X
IVA+SSA+ Robust HR detection	X	X	X	Propio	X	X

Cuadro 4.2: Tabla comparativa de características de Métodos Robustos a evaluar.

Se destaca que, IVA+SSA e IVA+SSA+Robust HR detection, son los algoritmo propuestos en la Metodología exploratoria 3 y Metodología exploratoria 4, respectivamente.

Métodos	Soporte		Señales fondo	Obtencion HR	Señales a procesar	
	<i>All mean</i>	<i>sub-ROIs</i>			<i>Original</i>	<i>All</i>
Tradicionales POS [32], ICA [27], PCA [22], LGI [26], CHROM [8]	X			PSD	X	

Cuadro 4.3: Tabla comparativa de características de Métodos Tradicionales a evaluar.

Con los detalles y características planteadas anteriormente sobre los métodos y su evaluación, se procede a mostrar los resultados de la evaluación sobre los Datasets presentados al inicio de este capítulo: UBFC, PBDT-rPPG y NAM-EDGE.

4.1. Análisis cualitativo

Para facilitar el análisis y evidenciar los rangos de funcionamiento de los métodos a evaluar, las mediciones de HR se dividen en tres categorías: *Low HR*, *Normal HR* y *High HR*. Cada categoría corresponde a un rango específico de HR: $[42, 60]$ Beats por Minuto (BPM), $]60, 100]$ BPM y $]100, 240]$ BPM, respectivamente. Al identificar estos rangos, se simplifica el análisis de la capacidad de un método para estimar el HR, ya que no todos los métodos son capaces de identificarlo en todo el rango posible.

Para llevar a cabo este análisis, se presentan dos tipos de gráficos generales que se utilizarán:

- Gráfico HR_{GT} vs $HR_{Estimated}$: Este gráfico muestra una comparación cualitativa entre la GT y el valor estimado por el método para un video de un sujeto en particular. Se construye utilizando el par $(HR_{GT}, HR_{Estimated})$ para cada estimación, lo que genera un nube de puntos que se comparan con una recta identidad que se construye utilizando sólo la GT, usando puntos de la forma (HR_{GT}, HR_{GT}) . Un método es buen estimador cuando la nube de puntos tiende a replicar la forma de la recta de GT. Un ejemplo de este gráfico se puede ver en la Figura 4.4.
- Gráfico Bland-Altman [23]: Es utilizado para comparar dos variables, en este caso HR_{GT} y $HR_{Estimated}$. En el eje vertical muestra las diferencias entre las variables y en el eje horizontal el promedio, por lo que cada punto presente en el gráfico muestra esta interacción de manera individual. Respecto a la forma del gráfico, la línea central y las líneas segmentadas que acompañan al eje, representan el promedio (mean) y el intervalo de confianza, siendo este último ± 1.96 veces la desviación estándar (SD). Para identificar cuándo un método es un buen estimador, *Low HR*, *Normal HR* y *High HR* deben ir agrupados en forma secuencial sin solaparse, ya que, esto indicaría que la estimación de HR no es precisa. Además, la línea de promedio debe tender a 0, al igual que la SD, por lo que las líneas tienden a juntarse.

4.1.1. Todos los Datasets

Para mostrar una comparativa en cuanto al desempeño general de los métodos Robustos y Tradicionales, se muestra en las Figuras 4.4 y 4.5 respectivamente, los resultado de todos los Datasets (UBFC, PBDT-rPPG,NAM-EDGE) para lograr evidenciar a grandes rasgos cuál o cuáles son los métodos que logran estimar de buena manera HR y en cuáles rangos de análisis, *Low HR*, *Normal HR* y *High HR*, funcionan mejor.

Respecto a los métodos Robustos, de la Figura 4.4, se puede apreciar error sistemático en los gráficos A, D y E. Estos métodos no son capaces de estimar de forma adecuada el valor de HR, viéndose una tendencia de estimación a valores concretos, que no son los valores esperados según la GT. Por ejemplo, en el gráfico E, se observa que los valores sólo se encuentran en 60, 120 y 180 BPM, mezclando todas las estimaciones *Low HR*, *Normal HR* y *High HR* en el caso de 120 BPM. Por otro lado, existen métodos capaces de seguir, en líneas generales, la tendencia de la recta de GT en los gráficos B, F y G. Estos métodos muestran una marcada tendencia a seguir la GT en los rangos de *Low HR* y *Normal HR*. Sin embargo, en *High HR*, sólo muestran una tendencia parcial a estimar valores precisos esperados, y en otros valores presentan sub-estimaciones de HR, lo cual puede resultar peligroso dado el contexto de aplicación específico esperado.

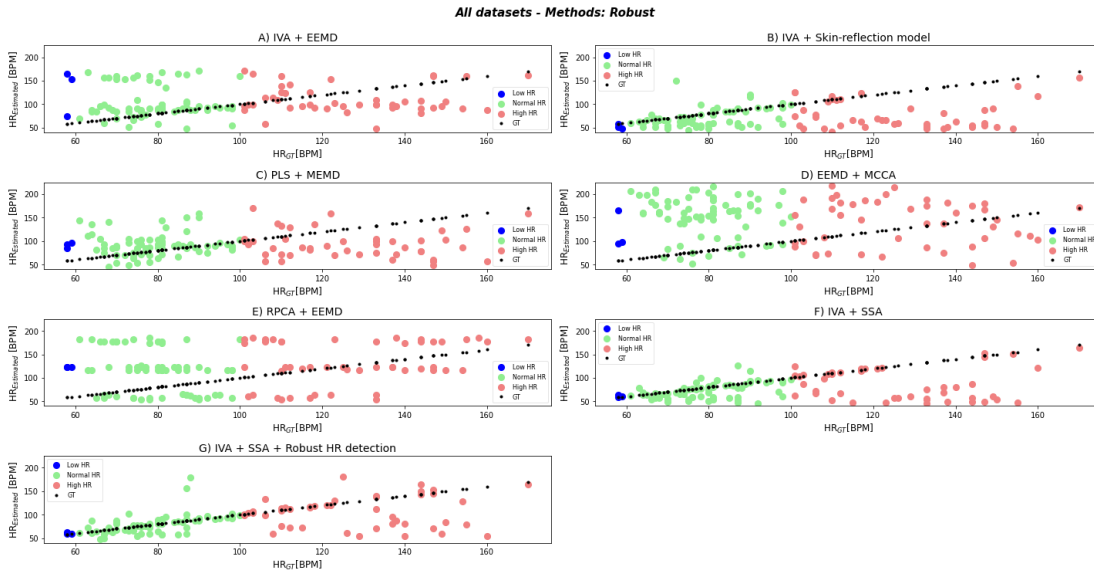


Figura 4.4: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Robustos evaluados en todos los Datasets.

Respecto a los métodos Tradicionales, de la Figura 4.5, se puede apreciar un error generalizado en los gráficos A, B y C. En estos gráficos es posible identificar una tendencia a oscilar entre valores concretos de 80 y 120 BPM aproximadamente, mezclando entre sí los rangos de análisis de HR: *Low HR*, *Normal HR* y *High HR*; Lo que se considera

una mala estimación por parte de estos tres métodos. Por otro lado, en el gráfico D, es posible identificar claramente la tendencia de las estimaciones a seguir la recta de la GT, sobre todo en el rango de *Normal HR*.

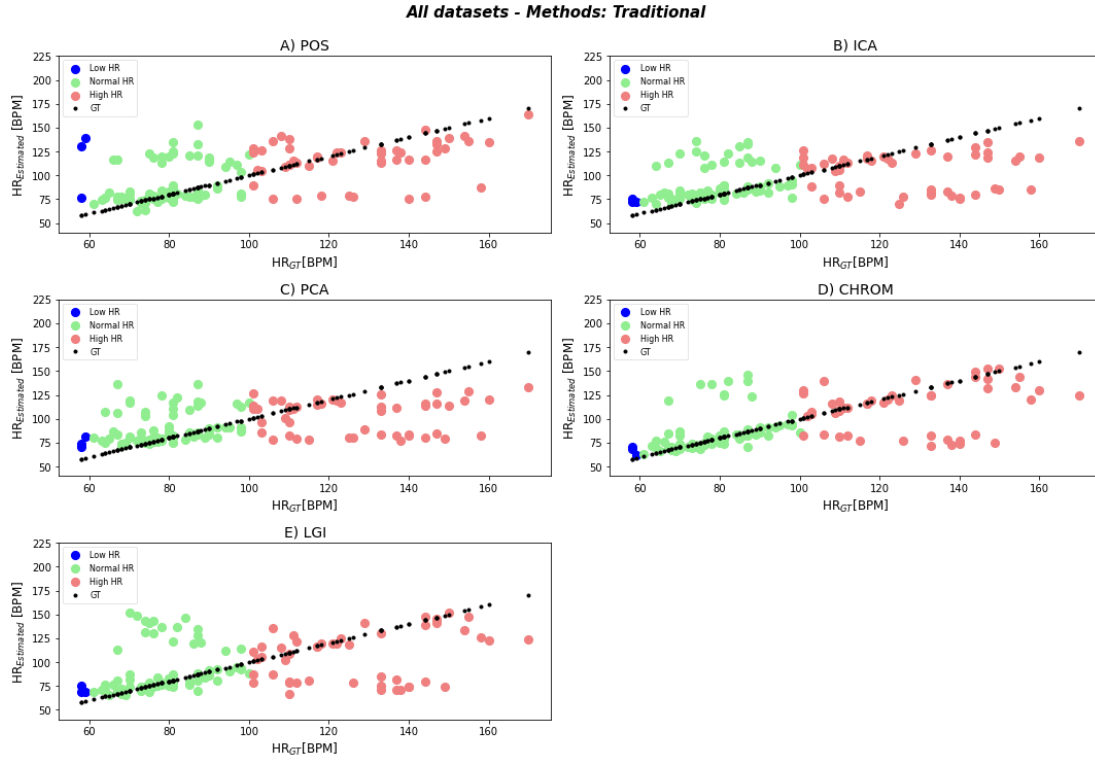


Figura 4.5: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Tradicionales evaluados en todos los Datasets.

Para proporcionar un mejor análisis, en la Figura 4.6 se encuentran los mejores métodos Robustos: **IVA+SSA**, **IVA+Skin-reflection model**, **IVA+SSA+Robust HR detection** y el mejor método Tradicional: **CHROM**.

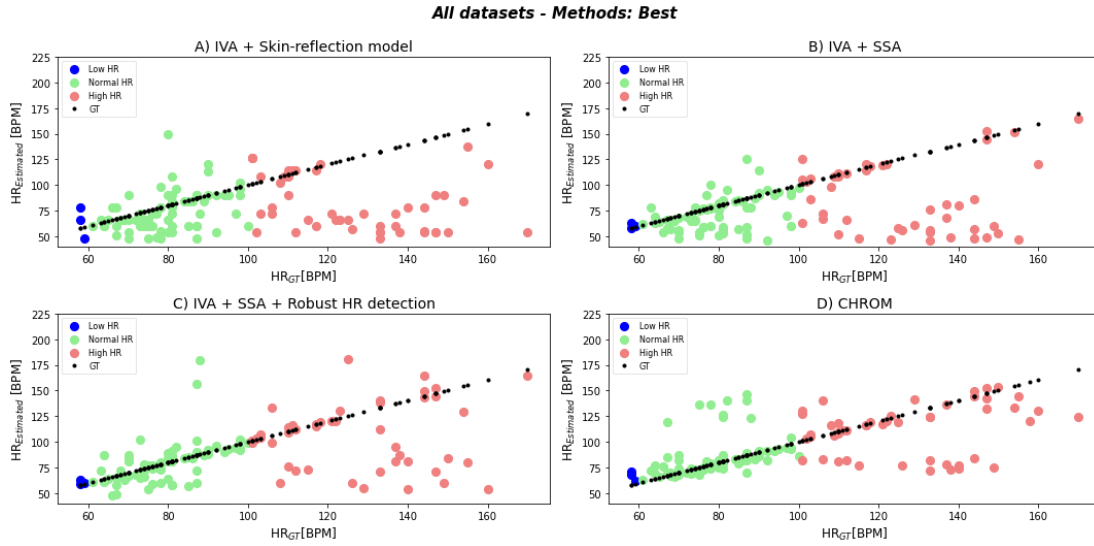


Figura 4.6: Gráficos HR_{GT} vs $HR_{Estimated}$ de los mejores métodos en todos los Data-sets.

Con el anterior análisis de todos los Datasets, de manera general, se procede a realizar una análisis concreto por cada Dataset para un análisis más fino.

4.1.2. Dataset: UBFC

Este Dataset cuenta con valores de HR que sólo se concentran en los rangos *Normal HR* y *High HR*, los sujetos presentan tono de piel claro y las condiciones de iluminación en la escena son estables. Estas condiciones son óptimas para que las estimaciones de HR de los sujetos sean bastante precisas, independientemente del método utilizado para obtener una señal confiable de rPPG y calcular los signos vitales posteriormente. Con lo anterior en consideración, se presentan los resultados sobre métodos Robustos, Tradicionales y resultados del *Framework* pyVHR [4] [5].

Para establecer una base de comparación entre pyVHR y los métodos Robustos y Tradicionales, se adaptó el cálculo de HR del *Framework*. pyVHR utiliza de manera nativa el cálculo de HR instantáneo utilizando ventanas de tiempo de 6 [s] con *stride* de 1 [s]. En cambio, en este trabajo, se utiliza una ventana de tiempo de 20 [s] para el cálculo de HR desde la señal de PPG de GT, lo que implica la pérdida de la variación temporal de HR en la GT. Por lo tanto, los datos de pyVHR se convirtieron al formato utilizado aquí, calculando el número de ventanas que equivalen al tiempo de análisis, promediando los valores de HR y comparándolos gráficamente con los otros métodos.

Para llevar a cabo un análisis comparativo entre los dos tipos de métodos, Robustos y Tradicionales, se han seleccionado los dos mejores métodos de cada tipo.

- **Métodos Robustos:** Los métodos en los gráficos F y G en la Figura 4.7, IVA+SSA e IVA+SSA+Robust HR detection, respectivamente, presentan el mejor desempeño a nivel cualitativo respecto a los demás métodos en los gráficos A, C, D y E. Esto se debe a que la distribución de los resultados tienen una clara tendencia a seguir la pendiente de la recta de GT, en especial el método del gráfico G, IVA+SSA+Robust HR detection, que sigue con una gran precisión los datos en los dos rangos de HR, *Normal HR* y *High HR*.

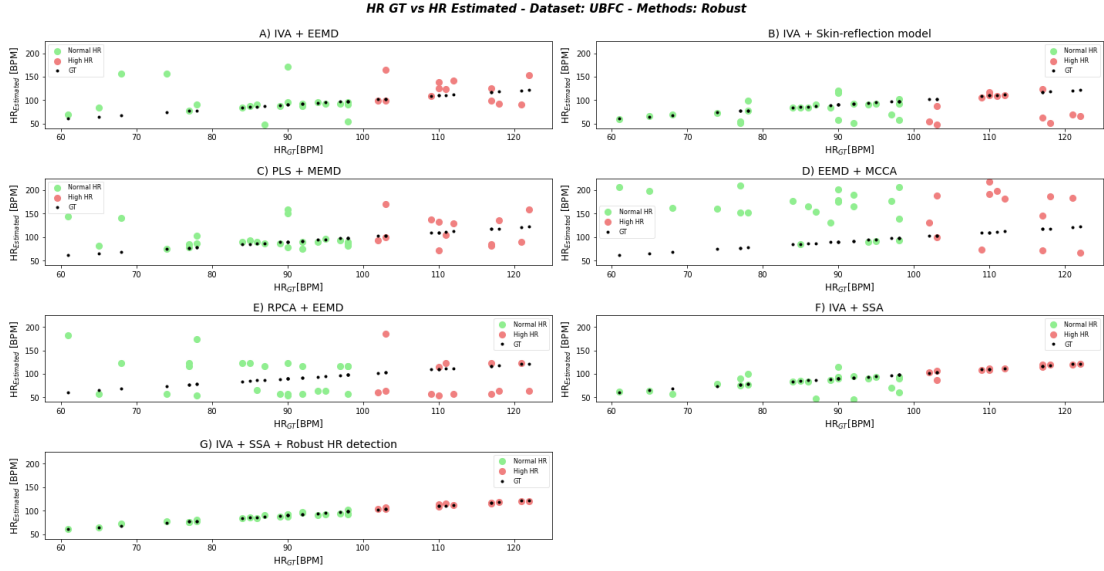


Figura 4.7: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Robustos evaluados en Dataset: UBFC.

- **Métodos Tradicionales:** Los métodos en los gráficos C y D en la Figura 4.8, presentan el mejor desempeño a nivel cualitativo respecto a los demás métodos en los gráficos A, B y E. Esto sucede debido a la clara tendencia de seguir la distribución de la recta de GT del Dataset, sobre todo el método CHROM del gráfico D, que obtiene los mejores resultados a nivel global para este Dataset, comparándolo visualmente con el mejor método de las Figuras 4.7 y 4.9, gráficos F y A, respectivamente.

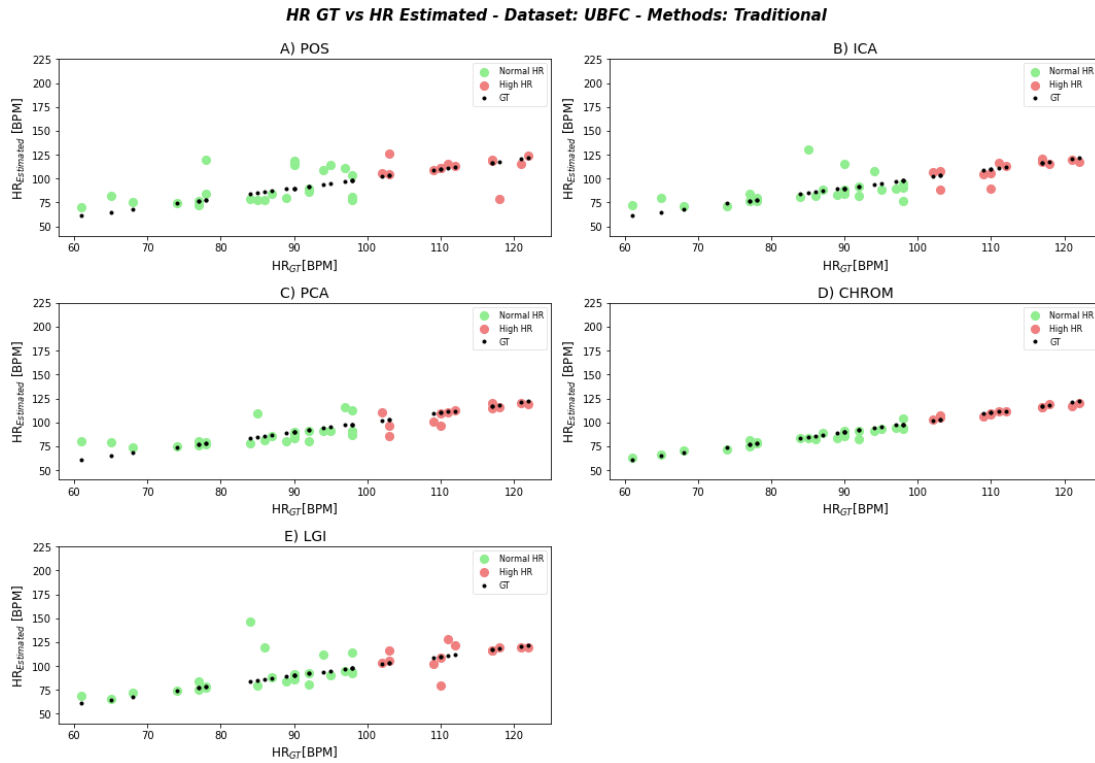


Figura 4.8: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Tradicionales evaluados en Dataset: UBFC.

- Métodos de *framework* pyVHR: Al igual que en los métodos Tradicionales, los mejores métodos en pyVHR son CHROM y PCA, en los gráficos A y E de la Figura 4.9, lo que es esperable debido a que se usan los mismos métodos para la obtención de la señal de rPPG, cambiando sólo en la parte de cálculo de HR.

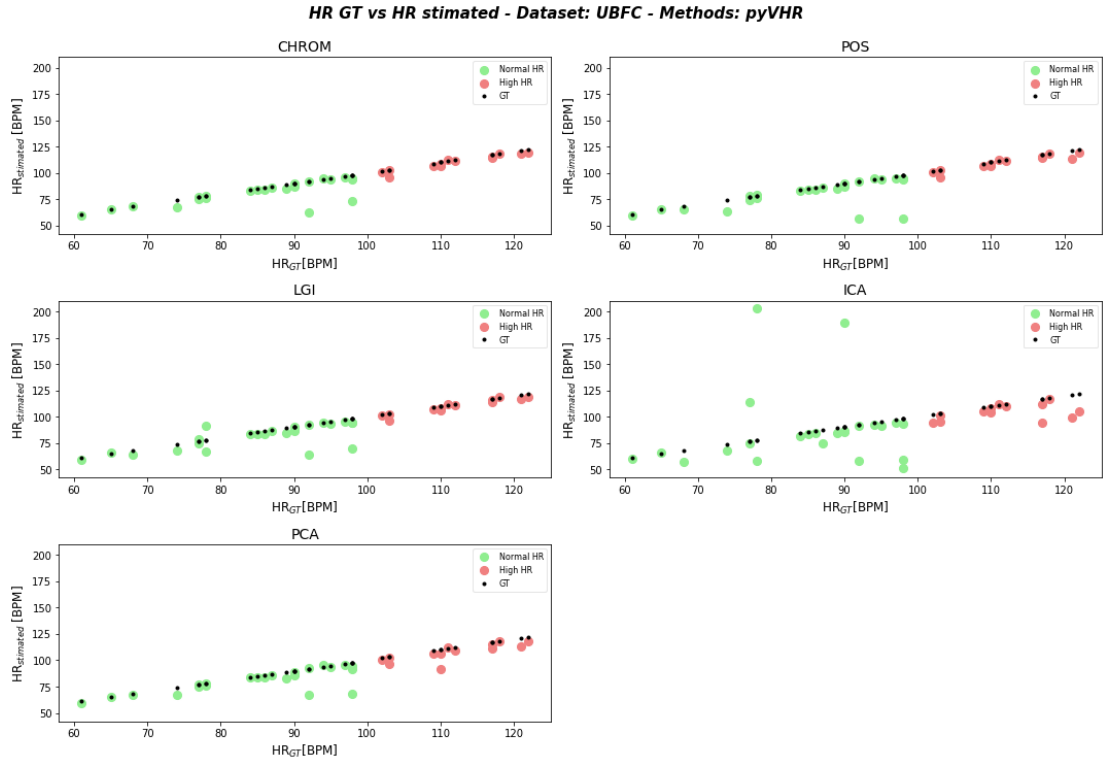


Figura 4.9: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Tradicionales de pyVHR evaluados en Dataset: UBFC.

Para proporcionar un mejor análisis, en la Figura 4.10 se encuentran los mejores métodos, donde destacan **IVA+SSA+Robust HR detection** y **CHROM**, representados por los gráficos C y D, respectivamente, que siguen de cerca la línea de la GT con gran precisión. Se debe señalar que los resultados de Pylvhr no se incluyen en la comparación, ya que se utilizan únicamente para fines de comparación sólo sobre el Dataset UBFC y no se consideran en el análisis general de este estudio.

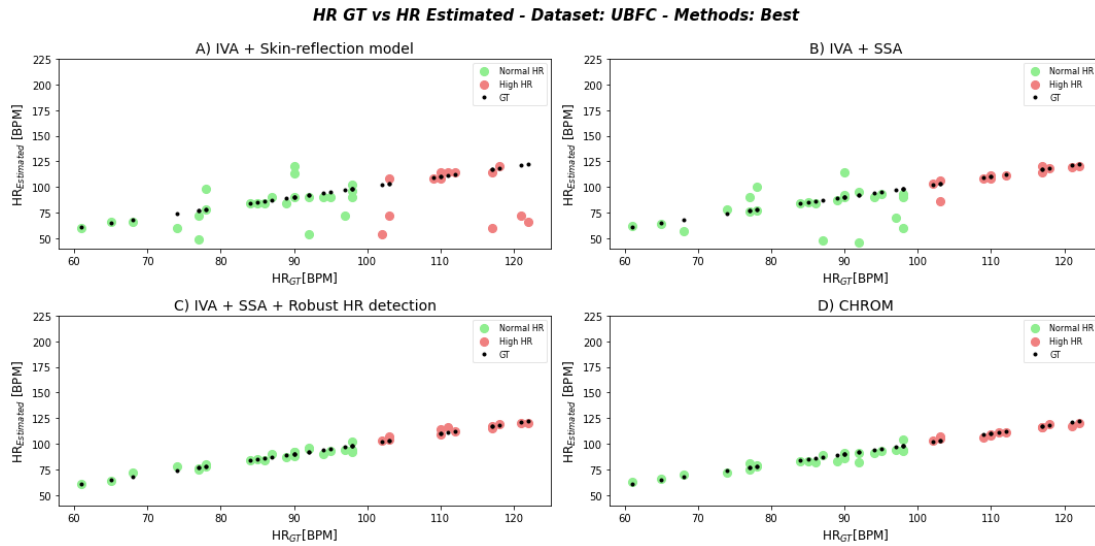


Figura 4.10: Gráficos HR_{GT} vs $HR_{Estimated}$ de los mejores métodos evaluados en Dataset: UBFC

Respecto al análisis del error, que se puede inferir de los gráficos Bland-Altman, se utilizaron los mejores métodos seleccionados del análisis anterior, de los gráficos HR_{GT} vs $HR_{Estimated}$, y se seleccionará el mejor por tipo de método para un posterior análisis a nivel general.

- **Métodos Robustos:** En la Figura 4.11, se observa que IVA+SSA+Robust HR detection puede estimar con precisión el HR en comparación con la distribución de error de los otros métodos robustos. Esto se debe a que IVA+SSA+Robust HR detection mantiene separados los grupos de *Normal HR* y *High HR*, lo que indica un bajo grado de confusión entre los rangos de HR definidos. Además, presenta una menor desviación del error, lo que se traduce en líneas punteadas más cercanas entre sí. Se puede notar en el gráfico G que estas líneas tienden a unirse debido al error mínimo presente en la estimación, donde los valores de error tienden a ser cercanos a cero, lo que se traduce en una muy buena estimación por parte de este método.

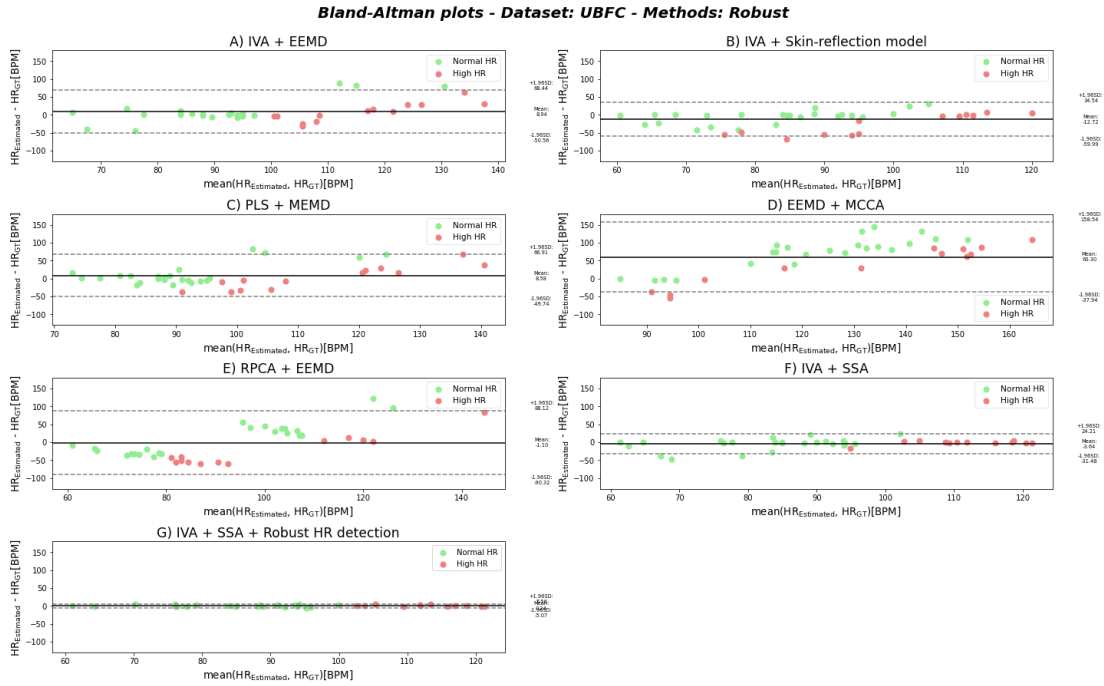


Figura 4.11: Gráficos Bland-Altman de métodos Robustos evaluados en Dataset: UBFC

- Métodos Tradicionales: Comparando los métodos CHROM y PCA, representados por los gráficos C y D respectivamente en la Figura 4.12, se evidencia que CHROM estima los valores de HR de manera altamente precisa y no presenta ninguna confusión al hacerlo, como se puede observar claramente al comparar con el gráfico de PCA. Las líneas segmentadas tienden hacia el valor de cero y se pueden distinguir claramente los valores de *Normal HR* y *High HR*.

Bland-Altman plots - Dataset: UBFC - Methods: Traditional

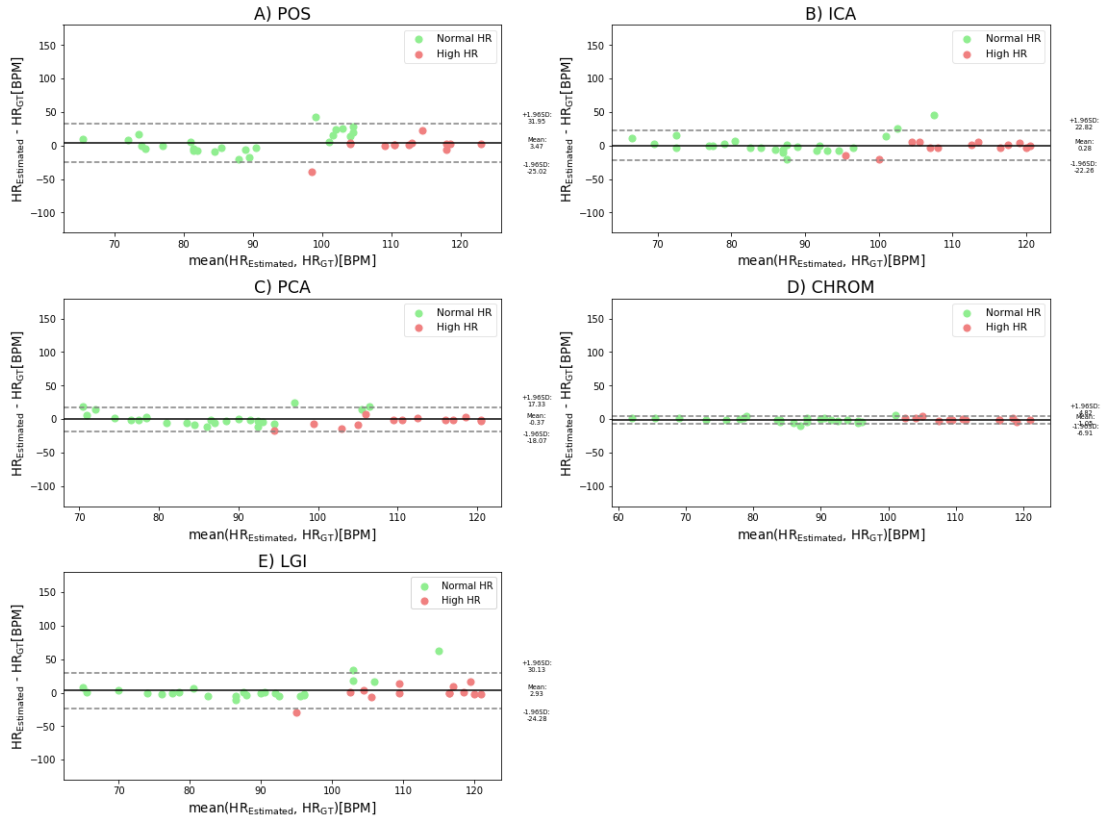


Figura 4.12: Gráficos Bland-Altman de métodos Tradicionales evaluados en Dataset: UBFC

- Métodos de *framework* pyVHR: Al igual que los métodos Tradicionales, CHROM representado en el gráfico A de la Figura 4.13 tiene un menor error que PCA, aunque los dos métodos presenten una distribución similar.

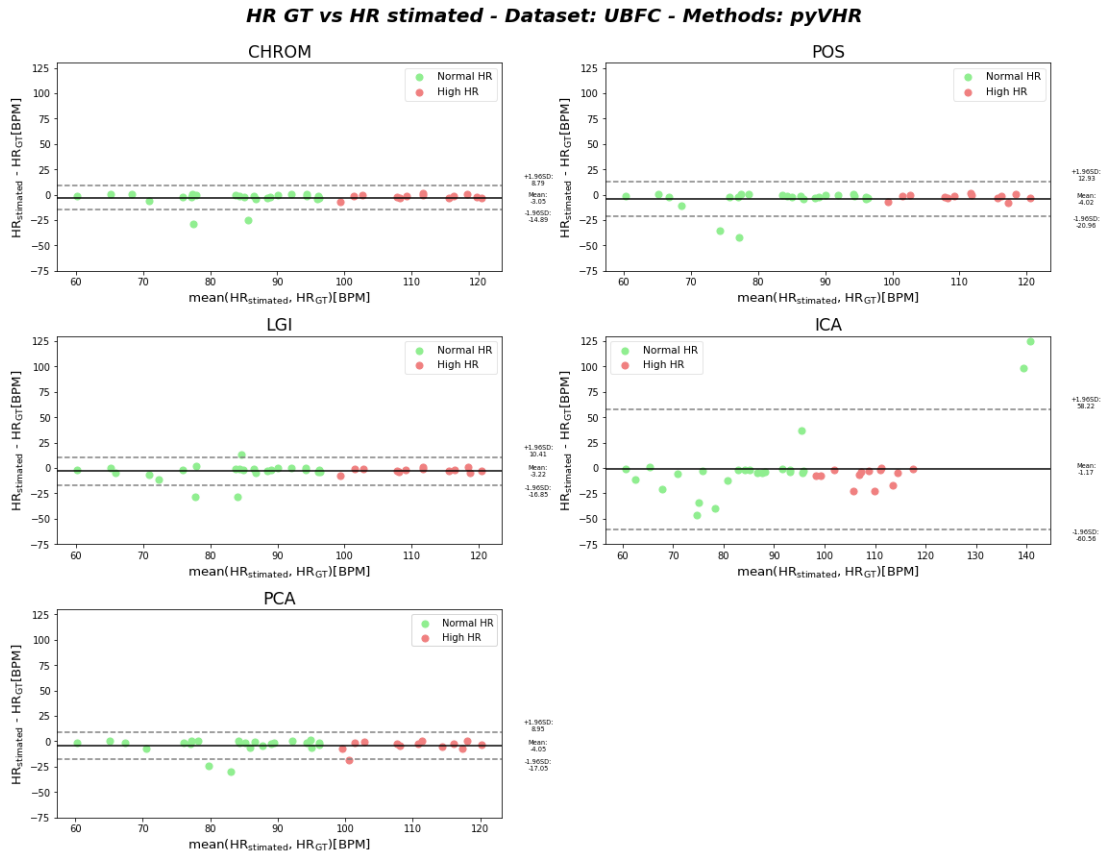


Figura 4.13: Gráficos Bland-Altman de métodos Tradicionales de pyVHR evaluados en Dataset: UBFC

En la Figura 4.14 se presentan los gráficos de error asociados a los mejores métodos. De estos, destacan de manera abismal los métodos IVA+SSA+Robust HR detection y CHROM, gráficos C y D respectivamente, donde el error es mínimo y están bien definidos los grupos de *Normal HR* y *High HR*, por lo que se considera una excelente estimación de HR por parte de ambos métodos.

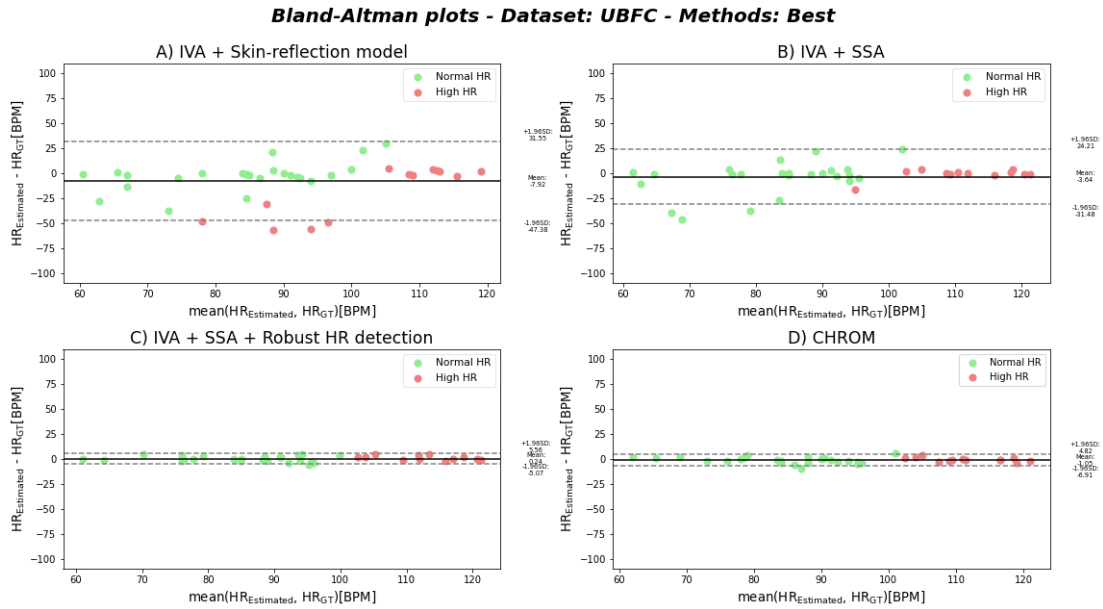


Figura 4.14: Gráficos Bland-Altman de los mejores métodos Tradicionales evaluados en Dataset: UBFC

4.1.3. Dataset: PBDT-rPPG

El análisis de este Dataset en particular resulta bastante interesante debido a que abarca una gran variedad de casos de interés: distintos niveles de iluminación, tonos de piel y movimientos. En concreto, la evaluación del desempeño de los métodos Robustos y Tradicionales en este conjunto de datos se dividirá en dos aspectos: la relación entre el tono de piel vs iluminación y análisis de rendimiento general basándose en los gráficos de HR_{GT} vs $HR_{Estimated}$ y los de bland-altman.

4.1.3.1. Análisis de tono de piel e iluminación

Para analizar la efectividad de los métodos Robustos y Tradicionales frente a distintos niveles de iluminación, se seleccionaron los videos de tres sujetos con distintos tonos de piel: clara, intermedia y oscura, piel tipo 1, 2 y 3 respectivamente. Estos sujetos se sometieron a cinco niveles de iluminación distintos, que están definidos en la Tabla 4.4, para poder evaluar la efectividad de la obtención del HR.

Niveles de iluminación	1	2	3	4	5
Intensidad de Luz [lux]	5.2	36	187	720	2720

Cuadro 4.4: Niveles de iluminación.

A continuación se muestran los resultados por tipos de métodos:

- **Métodos Robustos:** Respecto a estos métodos, queda evidenciado que en los gráficos A, B, C, F y G de la Figura 4.15 la tendencia del error es a disminuir a medida que va subiendo el nivel de iluminación, en al menos dos de tres tipos de piel, a medida que va subiendo el nivel de iluminación, lo que es bastante acertado según las teorías propuestas por varios autores[20][36][37]. Además, es importante destacar que los métodos Robustos representados en los gráficos A, B, C, F y G tienen en cuenta el entorno para determinar si es posible compensar los cambios de iluminación. Mientras que D y E, no lo toman en consideración y sólo se centran en el rostro. Así que, de acuerdo con los gráficos, se observa una tendencia donde, a medida que aumentan los niveles, el error tiende a disminuir. En este sentido, el método más efectivo es el representado en el gráfico G, que corresponde a la combinación de IVA+SSA+Robust HR detection (Ver sección 3.4).

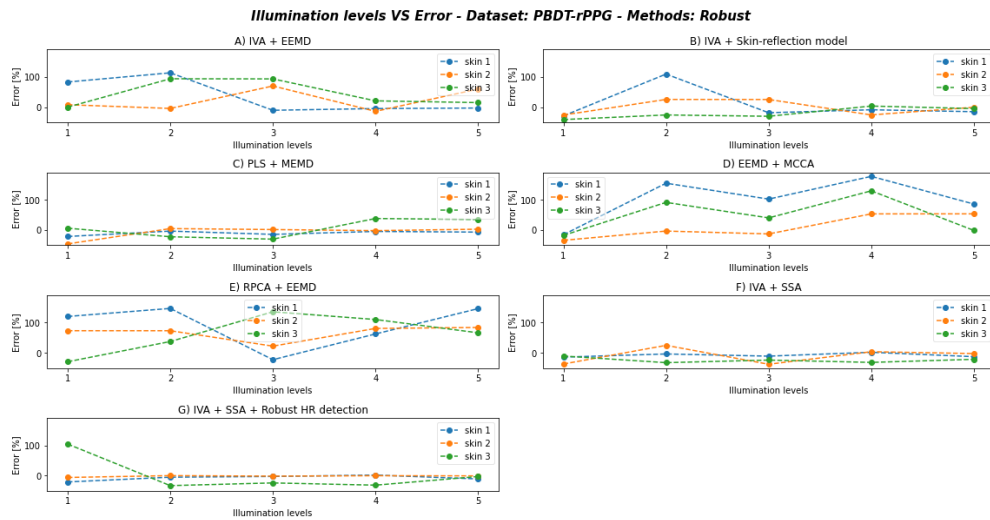


Figura 4.15: Gráfico Niveles de iluminación vs tiempo de métodos Robustos

- **Métodos Tradicionales:** Respecto a estos métodos, se evidencia que sólo en el gráfico D de la Figura 4.16, que corresponde al método CHROM, hay una tendencia decreciente del error a medida que aumenta el nivel de iluminación en los tres tipos de piel. Esto resulta favorable al considerar la diversificación del funcionamiento para una amplia variedad de tonos de piel. Por otro lado, queda evidenciado en el gráfico A, que corresponde al método POS, presenta un buen funcionamiento para sujetos con piel de tipo 1, mientras que en los de tipo 2 y 3, el error se mantiene a medida que aumenta el nivel de iluminación, lo que limita el uso del método considerablemente.

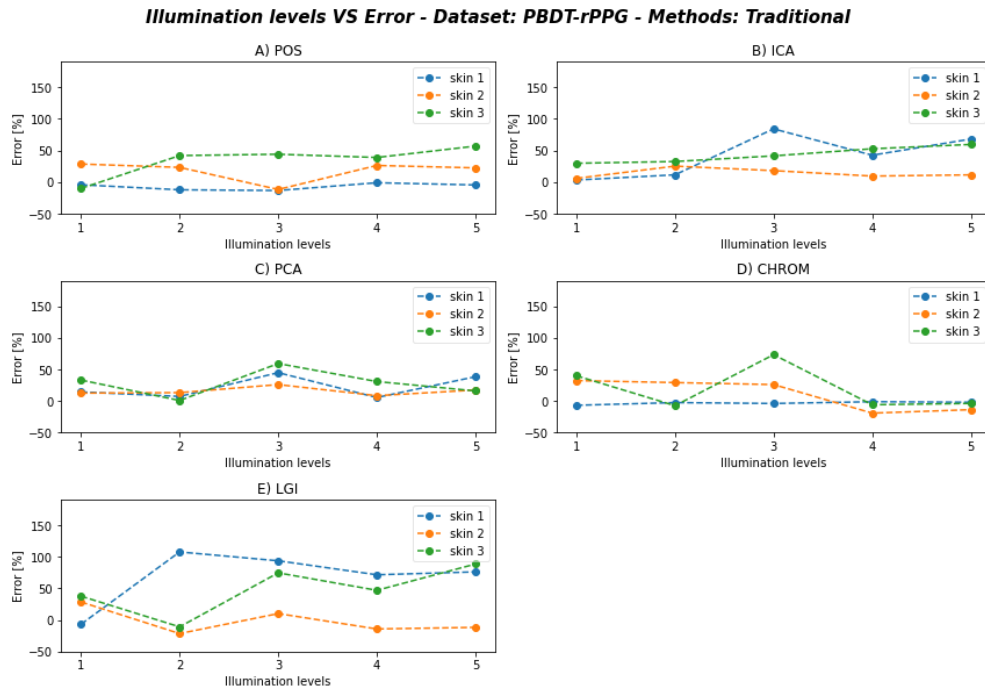


Figura 4.16: Gráfico Niveles de iluminación vs tiempo de métodos Tradicionales

Para evaluar el desempeño de los mejores métodos, la Figura 4.17 exhibe los métodos Robustos destacados junto con el mejor método Tradicional. Entre ellos, IVA+SSA+Robust HR detection se destaca claramente, mostrando una tendencia al error cercano a cero en el tipo de piel 1 y 2, de manera consistente a medida que aumenta la intensidad. En contraste, los demás métodos siguen adecuadamente, en la mayoría de los niveles, sólo el tipo de piel 1, el más claro, lo que facilita una estimación más precisa de los signos vitales. Sin embargo, en los tipos de piel 2 y 3, a medida que aumenta la intensidad, la precisión varía, con algunos oscilando entre la subestimación y la sobreestimación de la frecuencia cardíaca. Mientras tanto, IVA+SSA+Robust HR detection muestra un comportamiento coherente en el tipo de piel 3, ya que a medida que aumenta la intensidad, el error disminuye de manera precisa.

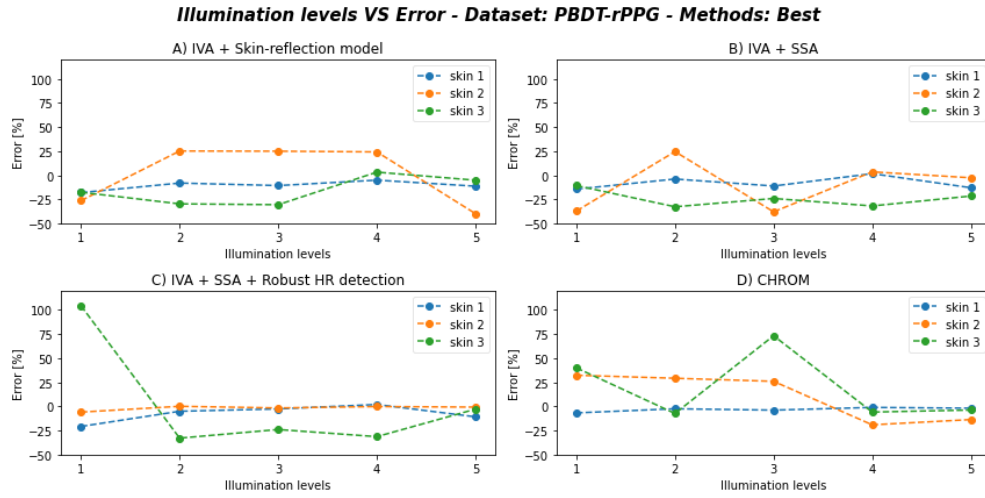


Figura 4.17: Gráfico Niveles de iluminación vs tiempo de los mejores métodos.

Es importante destacar que para lograr una precisión adecuada en el cálculo de HR en cualquier persona, los métodos deben ser robustos frente a los diversos niveles de iluminación presentes en un entorno no controlado, garantizando así una señal de rPPG de calidad para la estimación de los signos vitales, independientemente del tono de piel. Bajo esta premisa, los mejores métodos, tanto Robustos como Tradicionales, IVA+SSA+Robust HR detection y CHROM respectivamente, muestran una tendencia a ser más robustos en los niveles de iluminación 4 y 5 para los tres tonos de piel. Sin embargo, para los niveles inferiores, se observa claramente que el tipo de piel 1, el más claro, es prácticamente invariante ante los cambios de iluminación. Esto respalda la teoría de la distribución difusa y especular de la luz en la piel [37], donde la reflexión difusa lleva la información pulsátil de la señal de rPPG y la reflexión especular la iluminación reflejada sobre la piel. Estos componentes se ven directamente afectados por la cantidad de melanina en la piel, ya que una menor cantidad de melanina provoca que la luz incidente en la piel genere componentes difusos y especulares con mayor intensidad. Por otro lado, las pieles con mayor cantidad de melanina absorben en mayor medida la luz incidente, generando componentes difusos y especulares de menor amplitud. Esto, a su vez, afecta directamente a la calidad de la señal de rPPG como quedó evidenciado en la evaluación de este Dataset.

4.1.3.2. Rendimiento general sobre el Dataset

Para realizar en análisis general de los métodos Robustos y Tradicionales sobre el Dataset, primero se analizan los gráficos HR_{GT} vs $HR_{Estimated}$. Estos gráficos tienen marcadas las estimaciones de HR de los sujetos con tres tonos de piel distintos, loscuáles fueron analizados en el punto anterior. Estas estimaciones están etiquetadas como *skin 1*, *skin 2* y *skin 3*, representadas por un círculo, un rombo y un cuadrado azul, respectivamente. Por otro lado, están las mediciones que no están clasificadas según tono de

piel. Estas constituyen la otra parte del conjunto de datos, donde el sujeto con tono de piel tipo 1 (*skin 1*) realiza una serie de perturbaciones que van desde cambios en la dirección de la iluminación, hasta movimientos voluntarios de la cabeza, e incluso una medición después de ejercicio intenso. Es importante destacar que no se etiquetó con la figura geométrica correspondiente a *skin 1* para resaltar la diferencia entre la parte común del Dataset y la que presenta otro tipo de perturbaciones.

- Los Métodos Robustos, que incluyen IVA+EEMD, EEMD+MCCA y RP-CA+EEMD representados en los gráficos A, D y E respectivamente, presentan los mayores errores en la Figura 4.15 para los cinco niveles de iluminación. Se observa que estos métodos funcionan de manera inconsistente tanto para los diferentes tipos de piel expuestos a distintos niveles de iluminación como para otras perturbaciones. Solo logran una aproximación al valor de la GT para la estimación del video después de realizar ejercicio intenso (170 BPM). Por otro lado, se encuentran los mejores desempeños sobre el Dataset en los gráficos C, F y G en la Figura 4.18 que corresponden a PLS+MEMD, IVA+SSA e IVA+SSA+Robust HR detection, respectivamente. Vinculando estos gráficos con los correspondientes en la Figura 4.15, se puede establecer que, el método IVA+SSA+Robust HR detection del gráfico G presenta los mejores resultados en cuanto a piel tipo 1 y 2 para aproximadamente 4 niveles de iluminación distintos, en comparación a los métodos C y F, PLS+MEMD e IVA+SSA respectivamente, que sólo logran acercarse al valor de la GT en los niveles de iluminación superior. Sin embargo, respecto a la robustez de la otra parte del Dataset, los métodos C y F no son capaces de seguir la distribución de la GT en comparación con el gráfico G, que logra estimar con gran precisión los HR esperados, lo que demuestra en parte la robustez de este algoritmo propuesto en la Sección 3.4, IVA+SSA+Robust HR detection.

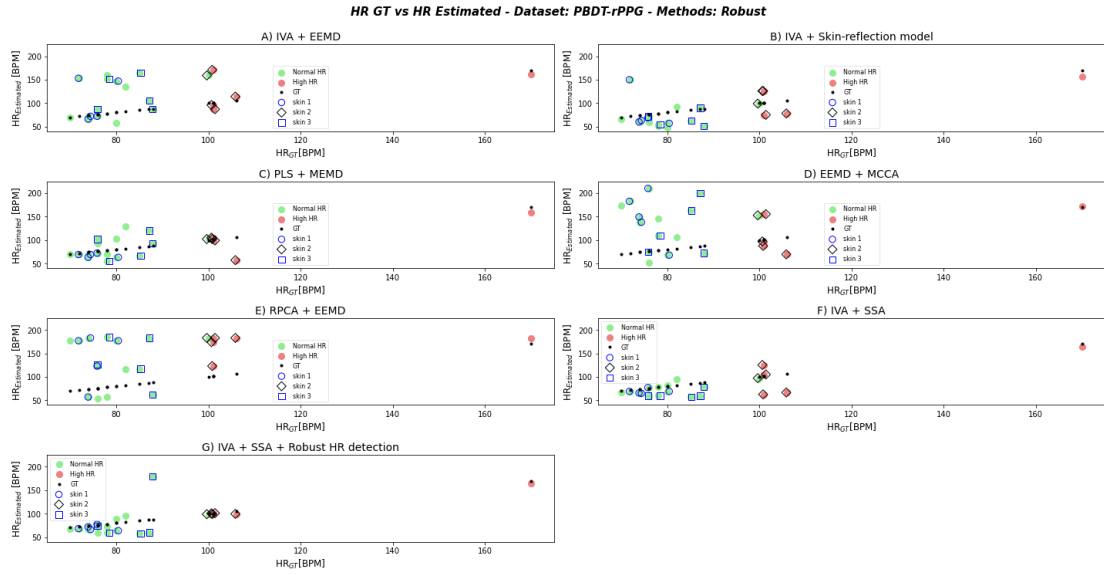


Figura 4.18: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Robustos evaluados en Dataset: PBDT-rPPG

- Los Métodos Tradicionales, que incluyen ICA, PCA y LGI representados en los gráficos B, C y E respectivamente, muestran los mayores errores en la Figura 4.16 para los cinco niveles de iluminación. Se observa que estos métodos funcionan de manera deficiente de forma sistemática, tanto para los diferentes tipos de piel expuestos a distintos niveles de iluminación, como para otras perturbaciones.

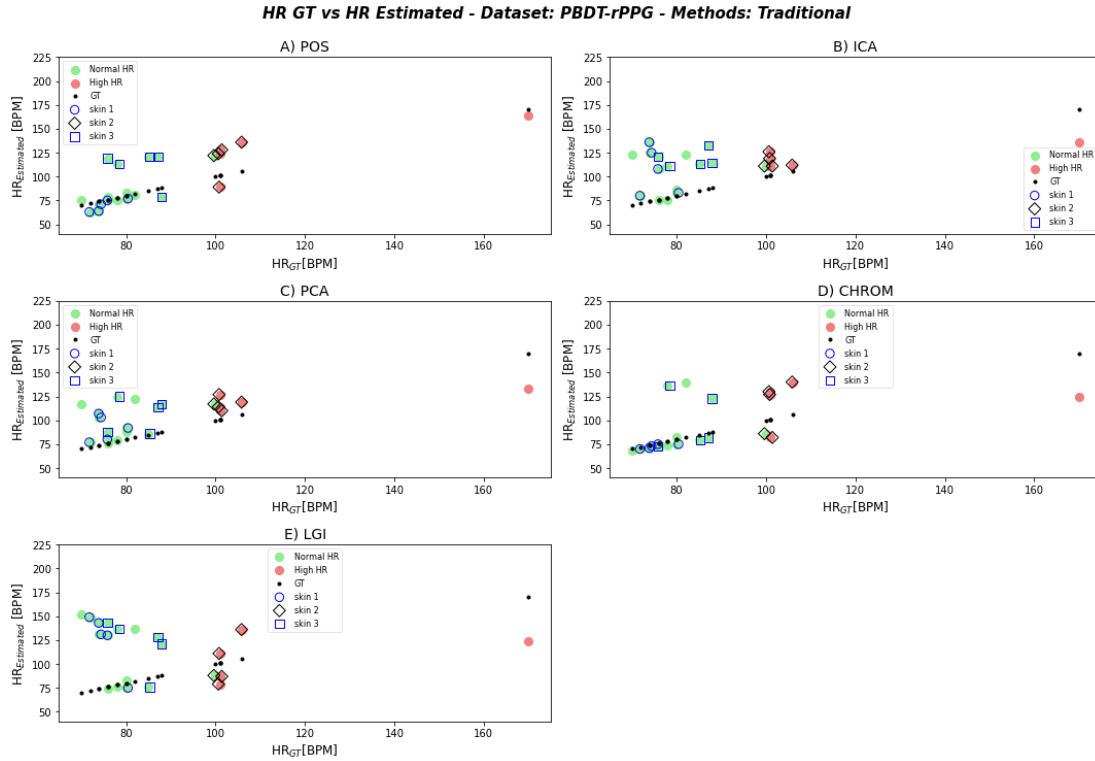


Figura 4.19: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Tradicionales evaluados en Dataset: PBDT-rPPG

Por otro lado, se encuentran los mejores desempeños sobre el Dataset en los gráficos A y D, métodos POS y CHROM respectivamente, en la Figura 4.19. Vinculando estos gráficos con los correspondientes en la Figura 4.16, es posible realizar un análisis más exhaustivo:

- **POS:** Este método estima con gran precisión los valores de HR para la persona con tono de piel tipo 1, tanto para los distintos niveles de iluminación, así como para la otra parte del Dataset que presenta distintos tipos de perturbaciones. Por otro lado, las personas con tono de piel tipo 2 y 3 no experimentan un aumento en la precisión de la estimación de HR a medida que aumenta el nivel de iluminación. Más bien, el error tiende a mantenerse constante a medida que aumenta la iluminación, lo que limita la versatilidad de este método para el uso previsto, especialmente en relación con el tono de piel.
- **CHROM:** Este método estima con gran precisión los valores de HR para la persona con tono de piel tipo 1, igual que POS, salvo en un valor estimado que pertenece a la parte del Dataset que presenta distintas perturbaciones. Sin embargo, en comparación con POS, la precisión en la estimación y cálculo de HR aumenta para los tonos de piel tipo 2 y 3 a medida que

incrementa el nivel de iluminación. Esto es bastante importante considerando que se busca un método altamente versátil que funcione en distintos tonos de piel. Por otro lado, es relevante señalar que el valor de HR estimado para el video tomado después de un ejercicio intenso, no alcanza una precisión adecuada en comparación con los demás métodos, lo que limita su capacidad en términos de detección HR muy elevado.

En la Figura 4.20, se muestran los resultados de los métodos más destacados. Tanto CHROM como IVA+SSA+Robust HR detection exhiben un buen desempeño en general, sin embargo, IVA+SSA+Robust HR detection muestra resultados más precisos en los casos de interés, especialmente en el caso de HR alto.

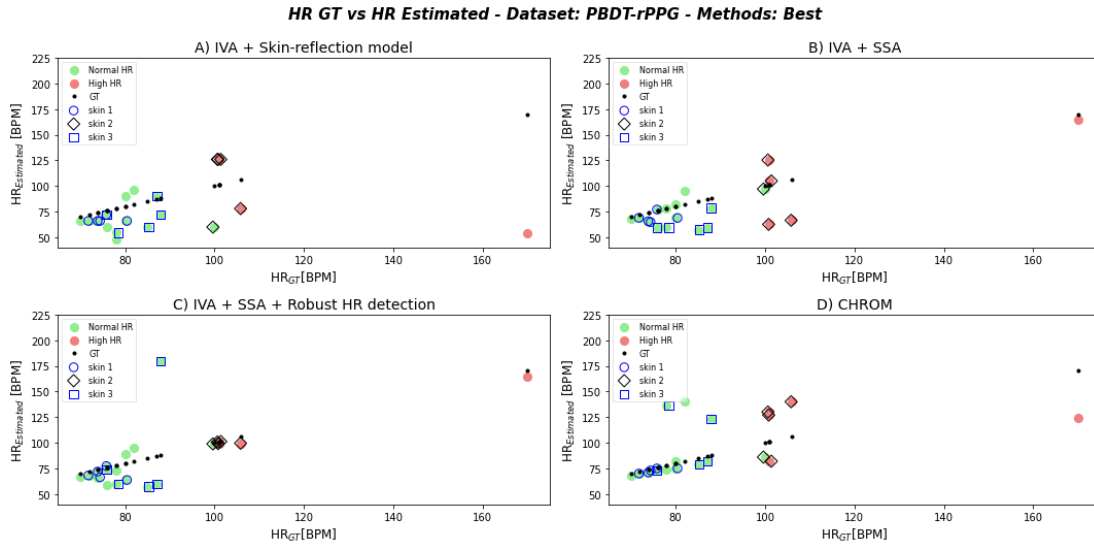


Figura 4.20: Gráficos HR_{GT} vs $HR_{Estimated}$ de los mejores métodos evaluados en Dataset: PBDT-rPPG

4.1.4. Dataset: NAM-EDGE

El análisis de este Dataset es bastante interesante debido a las características realistas que se pueden encontrar dentro de los videos que lo componen. Estas características se consideran perturbaciones directas que afectan a la obtención de la señal de rPPG, por lo que es de sumo interés analizarlas y evaluar la efectividad de los métodos Robustos y Tradicionales. Además, es importante destacar que aproximadamente la mitad de los sujetos que contribuyeron a la generación de los videos utilizan lentes ópticos, lo que incide negativamente en la calidad de la señal de rPPG. Esto se debe a que la posición espacial de los lentes en el rostro coincide en algún porcentaje con la Región de Interés (ROI) objetivo, y al hacer las segmentaciones de sub-ROIs, *all ROIs* y *original ROIs*, es posible que los lentes formen parte de alguna segmentación. Si las segmentaciones

son grandes, como las *original ROIs*, el efecto de los anteojos puede verse atenuado, ya que en comparación con la superficie total de la sub-ROI, la área que ocupan los lentes es mínima. Sin embargo, esto no ocurre cuando las segmentaciones son más pequeñas como en *all ROIs*, ya que es posible que ocupe un porcentaje de área no despreciable y que al ser seleccionada, afecte de manera global la selección de HR a través de la señal de rPPG estimada.

Con estas consideraciones iniciales, se procede a analizar los gráficos HR_{GT} vs $HR_{Estimated}$ para los métodos Robustos y Tradicionales. Se destaca que, los sujetos con lentes, están marcados en el gráfico con un rombo de color azul para hacer un análisis más preciso de robustez de los algoritmos.

- Métodos Robustos: En los gráficos A, C, D, y E de la Figura 4.21, los métodos IVA+EEMD, PLS+MEMD, EEMD+MCCA y RPCA+EEMD, respectivamente, se puede apreciar un error sistemático para las estimaciones de HR. Esto se debe a que ningún método es capaz de seguir una distribución similar a la recta de GT en ningún rango de HR, con o sin lentes. Por lo que quedan totalmente descartados.

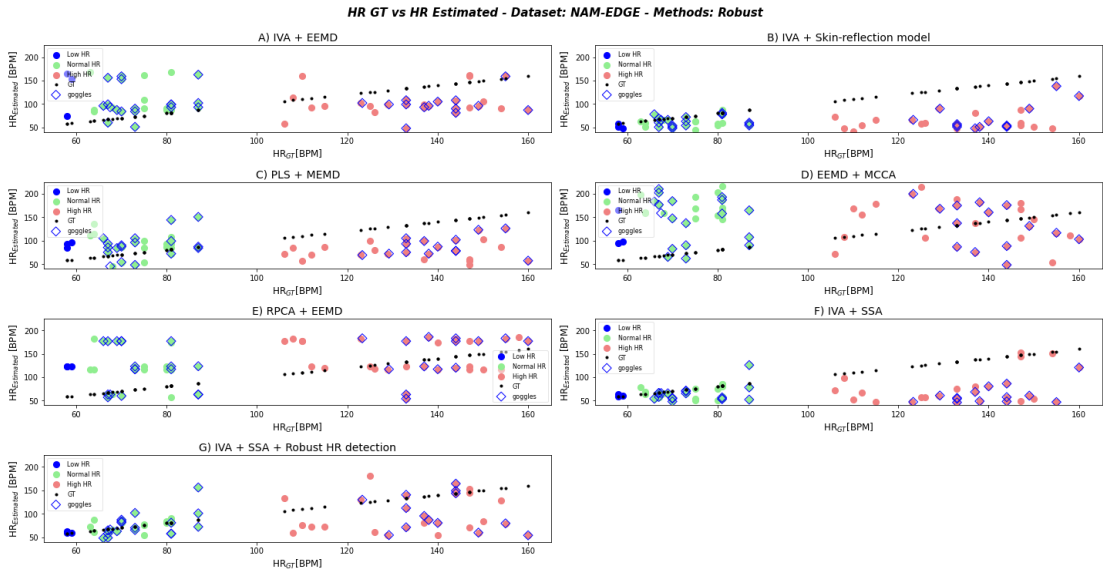


Figura 4.21: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Robustos evaluados en Dataset: NAM-EDGE

Por otro lado, se encuentran los gráficos B, F y G, IVA+*Skin reflection*, IVA+SSA e IVA+SSA+Robust HR detection, respectivamente, que presentan una tendencia a seguir la recta de GT en algunos rangos. Para evidenciar el comportamiento de estos métodos, se procede a mostrar un análisis por rango de HR:

- *Low HR*: Los tres métodos presentan una estimación precisa en este rango, especialmente el método IVA+SSA e IVA+SSA+*Robust HR detection* en comparación con IVA+*Skin reflection*.
 - *Normal HR*: En este rango, los métodos presentan un comportamiento similar tanto para los sujetos con anteojos como para los que no los tienen. Esto se puede apreciar al observar la tendencia a seguir la línea GT para los valores en general. De igual manera, es importante destacar que este Dataset, al contener una gran variedad de perturbaciones y considerando que la señal de rPPG es significativamente pequeña en comparación con estas perturbaciones, es posible que ya presente un error base de estimación.
 - *High HR*: En este rango, los tres métodos muestran un error sistemático en la estimación de HR que impacta tanto en los sujetos con anteojos como en aquellos sin ellos. No obstante, IVA+SSA+*Robust HR detection* logra estimar algunos valores de HR alto de manera correcta, pero subestima la mayoría del rango en cuestión. Es importante resaltar que debido a las grandes perturbaciones no controladas, los algoritmos tienden a fallar con frecuencia, especialmente en este rango.
- Métodos tradicionales: En general en todos los gráficos de la Figura 4.22, se puede apreciar la tendencia de seguir la GT en el rango *Normal HR*, sobre todo en los métodos de los gráficos D y E. Estos gráficos también siguen la tendencia de la GT en el rango de *Low HR*. Sin embargo, en el rango de *High HR*, los métodos CHROM y LGI presentan una gran cantidad de errores, aunque en menor medida que los demás métodos. Además, es posible notar que son capaces de estimar con precisión los valores de HR para algunos sujetos con anteojos. Esto se debe a que la morfología facial y el tamaño de los lentes varían para cada persona, lo que provoca una perturbación superficial en mayor o menor medida para el cálculo de la señal de rPPG. Además, es importante destacar que todos los métodos tradicionales utilizan *original ROIs* y soporte *All mean*, lo que minimiza el efecto de la superficie contaminada de los lentes al emplear sub-ROIs más grandes y promediar todas las señales entre sí.

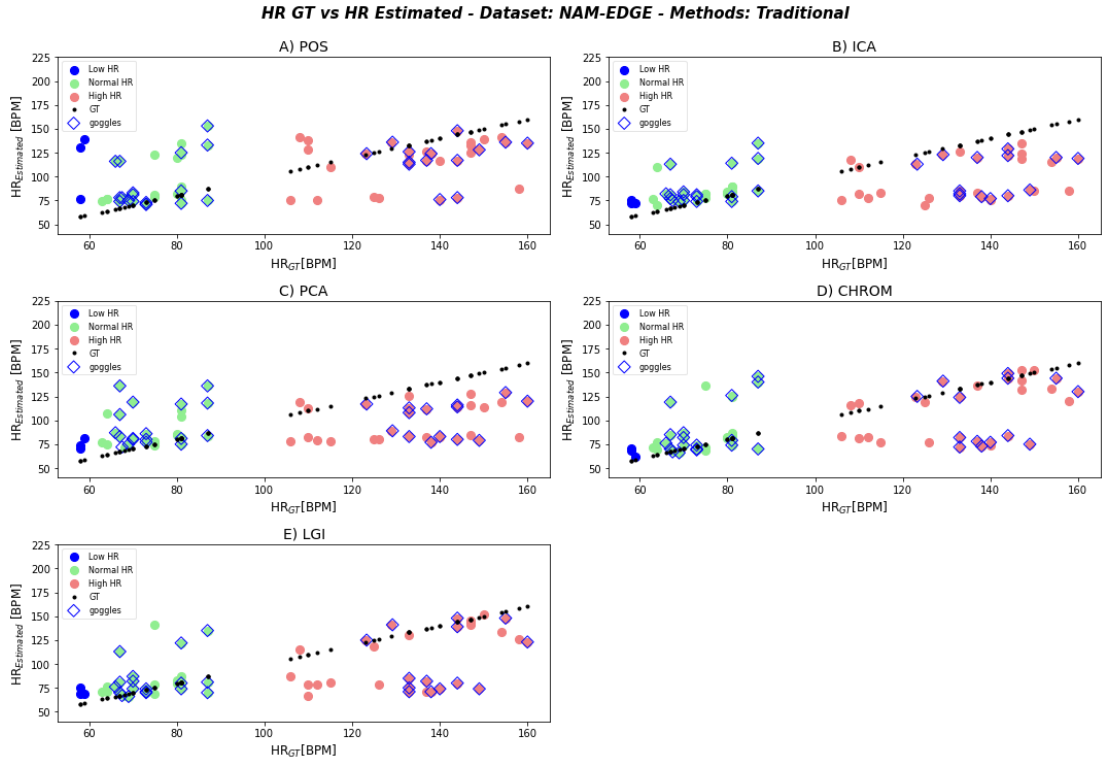


Figura 4.22: Gráficos HR_{GT} vs $HR_{Estimated}$ de métodos Tradicionales evaluados en Dataset: NAM-EDGE

En la Figura 4.23, los métodos IVA+SSA+*Robust HR detection* y CHROM presentan el mejor desempeño de manera generalizada, logrando seguir en parte la tendencia de la GT en los dos rangos de HR de análisis, *Normal HR* y *High HR*. Sin embargo, es posible identificar que ocurren errores sistemáticos en la inferencia de los signos en algunos casos particulares, ya sea con o sin lentes. Esto se debe a que no sólo las perturbaciones de iluminación afectan directamente a la señal de rPPG, sino que también perturban agresivamente todo lo relacionado con el movimiento facial, generando cambios en la iluminación local que no pueden ser compensados de antemano y que afectan la calidad de la señal de rPPG. Por lo tanto, es importante tener un protocolo de uso donde sea posible minimizar el efecto de las perturbaciones que pueden ser controladas por el sujeto de prueba para obtener señales de buena calidad.

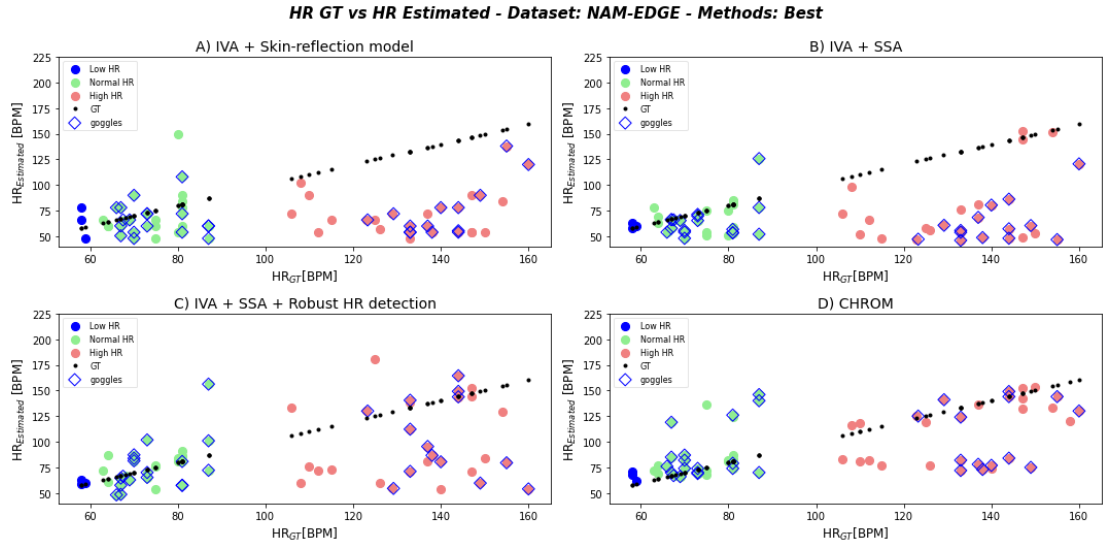


Figura 4.23: Gráficos HR_{GT} vs $HR_{Estimated}$ de los mejores métodos evaluados en Dataset: NAM-EDGE

Respecto al error en la inferencia de HR de los mejores métodos seleccionados, se observa claramente en la Figura 4.24 que un grupo específico de sujetos presenta un error sistemático independiente del método utilizado, como se evidencia en el centro de cada gráfico en la Figura. Esta peculiaridad podría deberse a un error de arrastre debido a una falta de protocolización adecuada en la toma de datos, lo que afecta directamente a la señal de rPPG. Por lo tanto, es necesario enfatizar la importancia de contar con un protocolo claro para la toma de datos con el fin de preservar de manera efectiva la señal de rPPG, incluso en entornos difíciles.

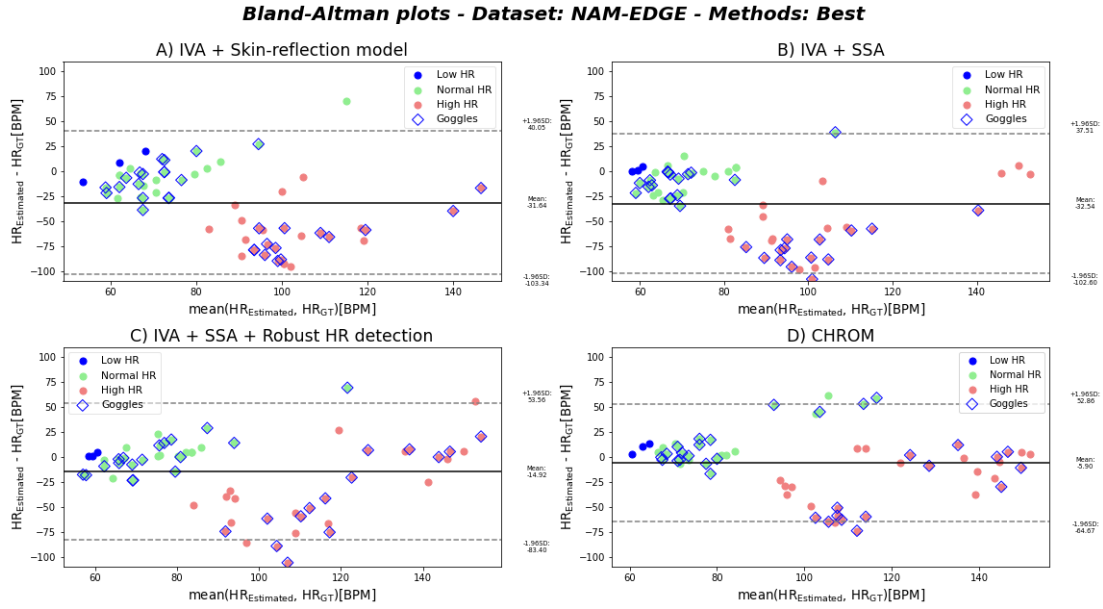


Figura 4.24: Gráficos Bland-Altman de los mejores métodos evaluados en Dataset: NAM-EDGE

4.2. Análisis cuantitativo

Desde el análisis cualitativo, fue posible identificar los mejores métodos en términos generales y para cada Dataset en particular. Específicamente, se identificaron dos: IVA+SSA+Robust HR detection y CHROM, que fueron los que mostraron los mejores resultados. También se agregaron para efectos de análisis, dos métodos más que presentaron buenos resultados en los rangos *Low HR* y *High HR*, IVA+SSA e IVA+Skin-reflection model. En esta sección, se evaluarán de manera cuantitativa los mejores métodos seleccionados del análisis anterior, comparando tanto el desempeño general como el específico, así como también la comparación entre los tiempos de ejecución de cada uno para lograr tales resultados.

4.2.1. Desempeño por Dataset

En particular en esta sección, se utilizarán métricas para cuantificar el error de estimación de cada método sobre todos los datasets de análisis. Estas métricas son:

- *Mean Absolut Error* (MAE): Es una medida estadística para cuantificar la precisión de un modelo basado en la diferencia promedio entre los datos estimados, que en este caso son los valores de Heart Rate (HR) estimados por los algoritmos,

y los valores reales, o sea, los valores de la Ground Truth (GT). Entonces, como se basa en el error absoluto, se debe minimizar para conseguir buenos resultados.

La fórmula general para el cálculo de MAE es la siguiente:

$$MAE = \frac{1}{N} \sum_{i=1}^N |HR_{GT}(i) - HR_{Estimated}(i)| \quad (4.1)$$

Donde:

- N es el número total de sujetos.
 - $HR_{GT}(i)$ es el valor de la GT para el sujeto i .
 - $HR_{Estimated}(i)$ es el valor de la Estimación del método seleccionado para el sujeto i .
- *Mean Absolute Percentage Error* (MAPE): Es una medida estadística similar a MAE pero que utiliza el error relativo en vez del error absoluto. Entonces, como se basa en error relativo, se debe minimizar para conseguir buenos resultados.

La fórmula general para el cálculo de MAPE es la siguiente:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{HR_{GT}(i) - HR_{Estimated}(i)}{HR_{GT}(i)} \right| \quad (4.2)$$

Donde las variables usadas son las mismas que la ecuación anterior 4.1 .

Con las métricas presentadas, se muestran en las Tablas 4.5, 4.6 y 4.7 la información sobre el desempeño de los mejores métodos evaluados sobre cada uno de los Datasets mostrando las métricas correspondientes según rango de HR: *Low HR*, *Normal HR* y *High HR*.

El orden de análisis y presentación de los resultados de la evaluación sobre los Datasets parte desde el que contiene mínimas perturbaciones, que corresponde UBFC, hasta llegar al Dataset que contiene gran cantidad de perturbaciones de iluminación y movimiento, que corresponde a NAM-EDGE.

4.2.1.1. Dataset: UBFC

En la Tabla 4.5 se presentan los resultados de la evaluación de los mejores métodos seleccionados en la sección anterior. Es importante destacar que este Dataset sólo presenta valores de HR en los rangos *Normal HR* y *High HR*, en donde los mejores resultados por rango se encuentran marcados de color azul.

Methods	UBFC					
	Low HR		Normal HR		High HR	
	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]
IVA + Skin-reflection model	-	-	10.84	9.38	17.84	20.21
IVA + SSA	-	-	12.11	10.71	2.46	2.64
IVA + SSA + Robust HR detection	-	-	2.86	2.47	1.65	1.81
CHROM	-	-	3.31	2.88	1.51	1.68

Cuadro 4.5: Métricas Dataset: UBFC.

Se puede ver claramente que los métodos IVA+SSA+*Robust HR detection* y CHROM, presentan los errores más bajos, con un grado de error similar. Sin embargo, IVA+SSA+*Robust HR detection* logra la mejor estimación de HR en el rango de *Normal HR* y CHROM en el rango de *High HR*. Esta información concuerda gratamente con el análisis cualitativo correspondiente a este Dataset.

4.2.1.2. Dataset: PBDT-rPPG

En la Tabla 4.6 se presentan los resultados de la evaluación de los mejores métodos seleccionados en la sección anterior. Es importante destacar que este Dataset sólo presenta valores de HR en los rangos *Normal HR* y *High HR*, igual que UBFC, en donde los mejores resultados por rango se encuentran marcados de color azul.

Methods	PBDT-rPPG					
	Low HR		Normal HR		High HR	
	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]
IVA + Skin-reflection model	-	-	17.07	14.05	33.8	43.81
IVA + SSA	-	-	13.22	10.66	21.23	22.19
IVA + SSA + Robust HR detection	-	-	18.62	15.49	2.18	2.68
CHROM	-	-	15.29	12.58	26.73	31.03

Cuadro 4.6: Métricas Dataset: PBDT-rPPG.

En el rango de *Normal HR*, se puede apreciar que el mejor método es IVA+SSA, en comparación a los demás métodos, debido a que contiene el valor mínimo de error. Es importante destacar que, la mayoría de las perturbaciones del Dataset, tanto de movimiento como distintos niveles de iluminación, están presentes en este rango. Por lo que presentan, de manera general, todos los métodos valores similares de error, MAE y MAPE. Se destaca IVA+SSA sólo por un 3 o 4 % respecto a lo demás métodos.

Por otro lado, en el rango de *High HR*, el método IVA+SSA+*Robust HR detection* presenta el mejor desempeño, destacándose significativamente por obtener un error mínimo considerable respecto a los demás métodos, presentando un error aproximadamente diez veces menor que los demás. Es importante destacar que en este rango se encuentran los valores de HR estimados para el sujeto de piel tipo 2 y el valor estimado de HR del sujeto que fue grabado después de una sesión intensa de ejercicio. Mostrar un error

tan bajo en la estimación es un buen indicio de la robustez del algoritmo para las personas de piel tipo 2 sometidas a distintos niveles de iluminación y a estimaciones robustas después de HR después de actividad física intensa. Sin embargo, estos resultados se dan en condiciones ambientales estables, o sea, sin grandes perturbaciones de movimiento ni de iluminación, por lo que falta un análisis de estos factores en ambientes reales.

4.2.1.3. Dataset: NAM-EDGE

En la Tabla 4.7 se presentan los resultados de la evaluación de los mejores métodos seleccionados en la sección anterior. Es importante destacar que este Dataset contiene valores en los tres rangos de interés : *Low HR*, *Normal HR* y *High HR*. Los mejores resultados por rango se encuentran marcados de color azul.

Methods	NAM-EDGE					
	Low HR		Normal HR		High HR	
	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]
IVA + Skin-reflection model	21.93	12.77	20.4	15.76	46.78	63.7
IVA + SSA	3.8	2.21	16.55	12.76	46.43	62.58
IVA + SSA + Robust HR detection	3.74	2.17	16.15	12.18	31.99	43.31
CHROM	14.95	8.7	20.26	15.47	20.49	27.61

Cuadro 4.7: Métricas Dataset: NAM-EDGE.

En el rango *Low HR*, se observa que IVA+SSA e IVA+SSA+*Robust HR detection* muestran los valores más bajos en términos de error. Estos métodos tienen en común el módulo de compensación de iluminación y sumado a que en este rango se presentan los sujetos con mayores perturbaciones, se puede establecer que el módulo de compensación realiza efectivamente el trabajo de compensar, presentando errores hasta cuatro veces menores que el estimado por CHROM e IVA+Skin-reflection model. Para más detalles sobre el módulo de compensación, revisar Sección 3.3.

En el rango *Normal HR*, se observa que IVA+SSA e IVA+SSA+*Robust HR detection* muestran los valores más bajos en términos de error, lo mismo que ocurre en *Low HR*; sin embargo, en comparación con los demás métodos, son solo un 20 % mejores aproximadamente. Es relevante señalar que en este rango se encuentran la mayoría de sujetos en reposo y con lentes, lo que lleva a un aumento en el error de las estimaciones. Por lo tanto, los errores presentan un rango similar en esta situación.

En el rango *High HR*, se observa que CHROM presenta el valor más bajo de error, un 30 % mejor, aproximadamente, que IVA+SSA+*Robust HR detection* el segundo mejor método. Es relevante señalar que en este rango se encuentran los mismos sujetos con lentes, pero ahora después de un ejercicio intenso. Durante la toma de datos, los sujetos presentan grandes movimientos rígidos de cabeza junto con movimientos involuntarios del rostro debido a la agitación después de esta actividad. Estas perturbaciones son de gran energía en el espectro en frecuencia y también se aprecian movimientos bruscos

que se evidencian en cambios de intensidad de color en cada canal que afectan directamente a la señal de rPPG para la obtención de los signos vitales.

4.2.1.4. Todos los Datasets

En la Tabla 4.8 se presentan los resultados de la evaluación de los mejores métodos seleccionados en la sección anterior. Es importante destacar que este Dataset contiene valores en los tres rangos de interés : *Low HR*, *Normal HR* y *High HR*. Los mejores resultados por rango se encuentran marcados de color azul.

Se puede observar que, en términos generales, el método IVA+SSA+Robust HR detection exhibe el mejor rendimiento en los rangos de *Low HR* y *Normal HR* en todos los datasets, excepto en *High HR*, donde muestra el segundo mejor desempeño. Por otro lado, CHROM presenta el segundo mejor rendimiento en el rango de *Normal HR* y el mejor rendimiento en *High HR*. De esta manera, se puede inferir que IVA+SSA+Robust HR detection tiende a ser más estable en condiciones generales que CHROM.

Methods	All Datasets					
	Low HR		Normal HR		High HR	
	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]	MAE[BPM]	MAPE[%]
IVA + Skin-reflection model	21.93	12.77	16.36	13.18	37.19	49.25
IVA + SSA	3.8	2.21	14.27	11.58	31.91	42.11
IVA + SSA + Robust HR detection	3.74	2.17	12.16	9.61	20.84	28.07
CHROM	14.95	8.7	13.31	10.49	16.09	21.08

Cuadro 4.8: Métricas de todos los Datasets.

Durante el desarrollo de las distintas metodologías, se ha observado que en ambientes no controlados con diferentes tipos de perturbaciones, no es suficiente contar únicamente con un algoritmo simple de identificación de HR, como el algoritmo detector de peaks en el tiempo o la selección del mayor valor en frecuencia de la señal obtenida. Es necesario establecer metodologías robustas de selección de HR según las condiciones estipuladas, que pueden incluso ser determinadas empíricamente. Un ejemplo de ello es la metodología presentada en el trabajo previo [39], que contiene a CHROM. Esta metodología emplea un enfoque de detección de HR basado en un *threshold* simple. Si se identifican tres valores de HR mayores a 100 BPM, se establece que la tendencia de HR en la señal se encuentra en dicho rango, y se promedian todos los valores que se encuentren en ese rango. En caso contrario, se promedian los valores que estén por debajo de 100 BPM. Como se mencionó, esta metodología incluye a CHROM. Por lo tanto, al utilizar el mismo algoritmo pero con una metodología de estimación de HR diferente, se obtendrían valores estimados distintos, incluso partiendo de la misma señal generada por CHROM.

Este último punto es crucial, dado que en la metodología exploratoria 4 (Sección 3.4) se hizo hincapié en utilizar únicamente el canal G de cada sub-ROI de análisis, sin realizar modificaciones, y se desarrollaron varios algoritmos robustos para estimar la

HR con precisión. Sin embargo, dependiendo de las condiciones de selección de los mejores candidatos, la estimación de la HR siempre variaba principalmente en el Dataset NAM-EDGE, que presenta condiciones realistas de entorno. Por lo anterior, se puede establecer que es de suma importancia contar con un algoritmo de detección robusto para casos reales, como en sesiones de telemedicina. No basta únicamente con la obtención de señales de rPPG obtenidas a partir de algoritmos Tradicionales para la estimación de HR, sino que deben ir acompañadas de un estimador de HR robusto para que funcione en todos los casos posibles.

4.2.2. Tiempo de ejecución

Esta sección, corresponde al análisis del tiempo necesario para dar una respuesta, considerando que el objetivo de este trabajo es identificar si es factible cumplir con las restricciones de tiempo real. Se debe tener completamente identificadas las partes de la implementación que ocupan la mayor cantidad de tiempo.

En primer lugar, se debe caracterizar de manera general la aplicación real en el ambiente de una sesión de telemedicina. En la Figura 4.25, se presenta un esquema simplificado de obtención y procesamiento de datos para estimación de signos vitales. Este esquema se divide en cinco partes importantes que se definen como:

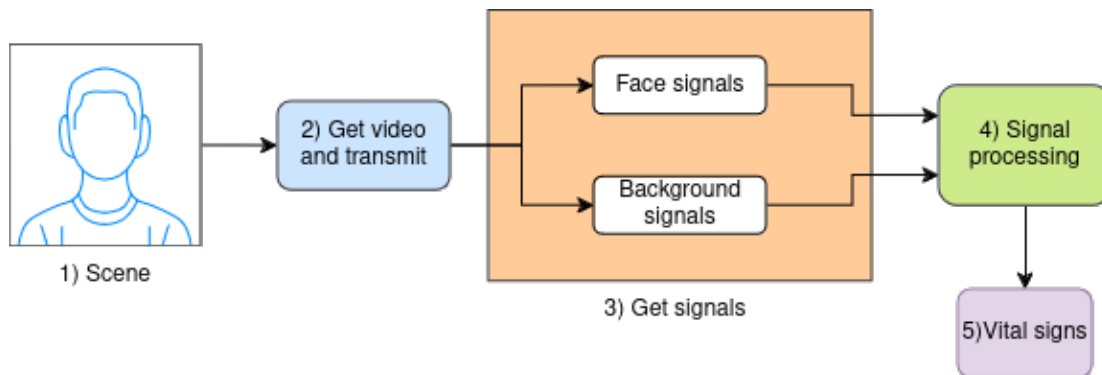


Figura 4.25: Esquema básico de obtención y procesamiento de señales para estimar signos vitales.

1. **Scene:** Este bloque es para caracterizar la escena en donde se desarrolla la sesión de telemedicina, o sea, donde se obtiene el video para ser procesado. Es de suma importancia tener un protocolo para la grabación del sujeto en la escena para minimizar las perturbaciones que afectan directamente a la señal de rPPG del rostro.
2. **Get video and transmit:** Este bloque es crucial en términos de tiempo de respuesta, dado que dependiendo del procesamiento del vídeo, ya sea localmente en el

teléfono o de forma remota en un servidor, implicará un tiempo considerable de procesamiento o de transmisión, respectivamente.

En el primer caso, esto se debe a que la mayoría de las personas actualmente no cuenta con un teléfono que disponga de la capacidad de procesamiento necesaria para proporcionar una respuesta adecuada en un tiempo acorde a la duración de la sesión de telemedicina. Sin embargo, dependiendo de la estrategia de estimación de los signos vitales, ya sea utilizando la señal completa de un video de 20 [s] o segmentos de 1 [s] que se van construyendo a medida que avanza la toma de datos, se puede reducir el tiempo de procesamiento y estimación de los signos vitales.

En el segundo caso, implicará un tiempo significativo de transmisión que depende de cuatro factores: velocidad de transmisión, ancho de banda disponible, tamaño del video y tipo de conexión.

3. **Get signals:** Este bloque hace referencia al tiempo necesario para la obtención de las señales de rPPG del rostro y del fondo. En primer lugar, se debe establecer cuáles son las señales que se pueden obtener del video, que en este caso son: *Original ROIs*, *All ROIs* y *Background*.

Como ya se identificaron los dos mejores métodos, IVA+SSA+*Robust HR detection* y CHROM, se calculará un estimado de los tiempos necesarios para la obtención de las señales dadas las distintas configuraciones de sub-ROIs. Estos tiempo están detallados en la Tabla 4.9, donde se puede ver el tiempo de obtención de señales de cada tipo de sub-ROIs por Dataset.

	UBFC	PBDT	NAM-EDGE
Resolution	(640, 480)	(1080, 608)	(2160, 3840)
Original ROIs time [s]	37	49	146
All ROIs time [s]	64	84	535
SEEDS time [s]	4	10	112
Background time [s]	27	46	488

Cuadro 4.9: Tiempos promedio de obtención de señales dadas las distintas configuraciones de sub-ROIs por Dataset.

- CHROM: Este método utiliza únicamente *Original ROIs*, lo que resulta en un tiempo de obtención de señales más bajo en comparación con IVA+SSA+*Robust HR detection*, ya que utiliza menos señales. Como se muestra en la Tabla 4.10, el tiempo de obtención de señales tiende a aumentar con la resolución, pero sigue siendo significativamente menor en comparación con el otro método, ya que sólo trabaja con 3 señales en el rostro. Aunque el tiempo de obtención es reducido, se pierde la caracterización general de la iluminación del rostro al tener una mayor área de información condensada por conjunto de señales.

- **IVA+SSA+Robust HR detection:** Esta metodología utiliza las señales *All ROIs* y *Background*, lo que incrementa significativamente el tiempo requerido para obtener las señales, lo que a su vez retrasa la estimación de los signos vitales en el paso siguiente. Como se puede observar en la Tabla 4.10, el tiempo necesario para la obtención de las señales aumenta considerablemente a medida que se incrementa la resolución del video de análisis, casi de manera exponencial. Esto se debe principalmente a la cantidad de conjuntos de señales que se obtienen, siendo 31 en el caso del rostro y 6 en el caso del fondo. Además, es importante destacar que los tiempos se suman porque la obtención se realiza de manera secuencial.

	UBFC	PBDT	NAM-EDGE
Resolution	(640, 480)	(1080, 608)	(2160, 3840)
CHROM time [s]	37	49	146
IVA+SSA+Robust HR detection time [s]	91	130	1.023

Cuadro 4.10: Tiempo promedio de obtención de señales de los mejores métodos.

Es relevante destacar que el tiempo requerido para obtener las señales depende exclusivamente de la estrategia utilizada para su construcción, tal como la posibilidad de paralelizar dicho proceso. Por otro lado, también influye el poder de procesamiento del dispositivo utilizado para la obtención de las señales, ya sea localmente en el teléfono del paciente o de forma remota en un servidor. Este último aspecto constituye un punto importante a tener en cuenta.

4. **Signal processing:** Este bloque es para caracterizar el tiempo necesario para procesar las señales obtenidas del rostro y el fondo para lograr una estimación de los signos vitales.

Como los métodos trabajan de forma distinta, se tiene que IVA+SSA+Robust HR detection ocupa aproximadamente 43 [s] en promedio para realizar una estimación a partir de las señales obtenidas. Mientras que CHROM se demora aproximadamente 2 [s] en promedio para lograr una estimación.

5. **Vital signs:** Este bloque es para caracterizar la entrega de las estimación de los signos vitales del paso anterior.

Según este análisis, el tiempo total de cada método para una estimación, considerando la toma de datos, la obtención de señales y el procesamiento de datos (partes 2, 3 y 4, respectivamente), es el siguiente:

	UBFC	PBDT	NAM-EDGE
Resolution	(640, 480)	(1080, 608)	(2160, 3840)
CHROM time [s]	39	52	148
IVA+SSA+Robust HR detection time [s]	154	193	1086

Cuadro 4.11: Tiempo promedio estimado de estimación de HR de los mejores métodos.

En términos de tiempo de ejecución, CHROM es superior a IVA+SSA+*Robust HR detection*, siendo aproximadamente hasta 10 veces más rápido en UHD y 4 veces más rápido en baja resolución, en el Dataset UBFC específicamente, como se puede apreciar en la Tabla 4.11.

4.3. Evaluación

Considerando el análisis anterior, análisis cualitativo y lo presentado en las secciones de las metodologías exploratorias, es posible confirmar que los mejores métodos dentro de este trabajo son CHROM e IVA+SSA+*Robust HR detection*. Debido a esto, hay que identificar cuáles son las condiciones de uso de cada metodología para maximizar su precisión y si es posible sólo utilizar una.

En primer lugar, es necesario identificar los principales mecanismos de funcionamiento de cada método, evaluando tanto su precisión en el cálculo como el tiempo de ejecución. Para lograr un análisis detallado, se presenta un breve resumen del funcionamiento general de las metodologías para luego pasar a un análisis más detallado, respecto al funcionamiento específico de cada una, para identificar los límites de funcionamiento.

- **IVA+SSA+*Robust HR detection*:** La idea principal en la que se basa este método es que una vez compensadas las perturbaciones de iluminación de todas las sub-ROIs, se debe encontrar un valor robusto de HR entre todos los segmentos de todas las señales de análisis compensadas. Para lograr esta identificación robusta entre todas las señales, se trabaja con ventanas de 10[s] que aportan una resolución espectral que permite visualizar los *peaks* espectrales con una resolución adecuada para una correcta identificación de continuidad. Con los caminos de continuidad contruidos a partir de las ventanas de cada señal, se procede a identificar mediante puntajes cuáles son los mejores candidatos a HR para su posterior unificación utilizando un histograma para agrupar la información conjunta recopilada de todas las señales de análisis seleccionadas. Construido el histograma, se selecciona el valor HR que tiene mayor densidad de valores en un entorno que depende de una holgura variable según la resolución espectral, eligiendo el que tiene mayor valor de acumulación.
- **CHROM:** la idea principal de este método es crear una señal de *Blood Volume Pulse* (BVP), que contiene la información de HR y los demás signos vitales, y lograr que sea robusta a las perturbaciones de iluminación. Para conseguir esto, asumen que la reflexión especular, que es el componente reflejado de la piel y que puede contener perturbaciones de iluminación, afecta a todos los canales de color, por lo que para eliminar este componente, se deben utilizar señales de diferencia entre canales, es decir crominancia, para eliminar este componente de iluminación especular no deseada quedando sólo la información de los cambios de volumen de la sangre en las venas del rostro.

Con este pequeño resumen del funcionamiento de los mejores métodos, es necesario identificar cuáles son sus límites de funcionamiento junto con las condiciones que permiten maximizar la precisión al obtener los signos vitales.

Según lo desarrollado en este trabajo, se logró identificar cuáles eran las funcionalidades principales de cada método en cuestión, CHROM e IVA+SSA+*Robust HR detection*, para lograr identificar los límites de funcionamiento y las condiciones en las que se puede maximizar el uso tanto individual como combinado.

Una de las principales diferencias en cuanto a las funcionalidades de cada método, radica en cómo se compensan las perturbaciones de iluminación. CHROM utiliza únicamente la información contenida en el rostro, aprovechando el hecho de que los canales RGB contienen información necesaria sobre la reflexión especular, que contiene las perturbaciones, y la señal objetivo que contiene los signos vitales. Al emplear la diferencia entre canales, CHROM puede eliminar gran parte de las perturbaciones de iluminación. Sin embargo, bajo ciertas condiciones particulares, en donde la reflexión especular presenta componentes de distinta magnitud en los canales RGB, es posible que también se atenúen los componentes objetivo y se resalten componentes no deseados. Esto se debe a las combinaciones lineales de los canales RGB para generar las señales de crominancia, las cuáles se construyen utilizando coeficientes numéricos que multiplican a cada canal según requiera la señal de crominancia en cuestión. Es importante destacar que estos coeficientes son estáticos, o sea, siempre se aplican los mismos coeficientes sin importar la escena o entorno de desarrollo de la obtención de las señales de rPPG.

A continuación se presenta un pequeño ejemplo de comparación sobre el Caso 1: Perturbaciones de iluminación. En la Figura 4.26, se observa el espectro de las ventanas de análisis de la señal obtenida mediante el método CHROM. Es evidente que el espectro muestra un alto nivel de ruido en las primeras ventanas, aunque también se aprecia una ligera tendencia hacia el valor de GT (representado por la barra de color verde oscuro) en algunas de ellas. Sin embargo, los valores máximos dentro de estas ventanas, en la mayoría de los casos, no coinciden con el valor de GT. Aunque la tendencia está presente, el método de selección del máximo no elige los valores esperados. Esto se debe principalmente al ruido presente en la escena, que no fue compensado por el método, que genera componentes en frecuencia que atenúan el valor de la GT y resalta componentes no deseados.

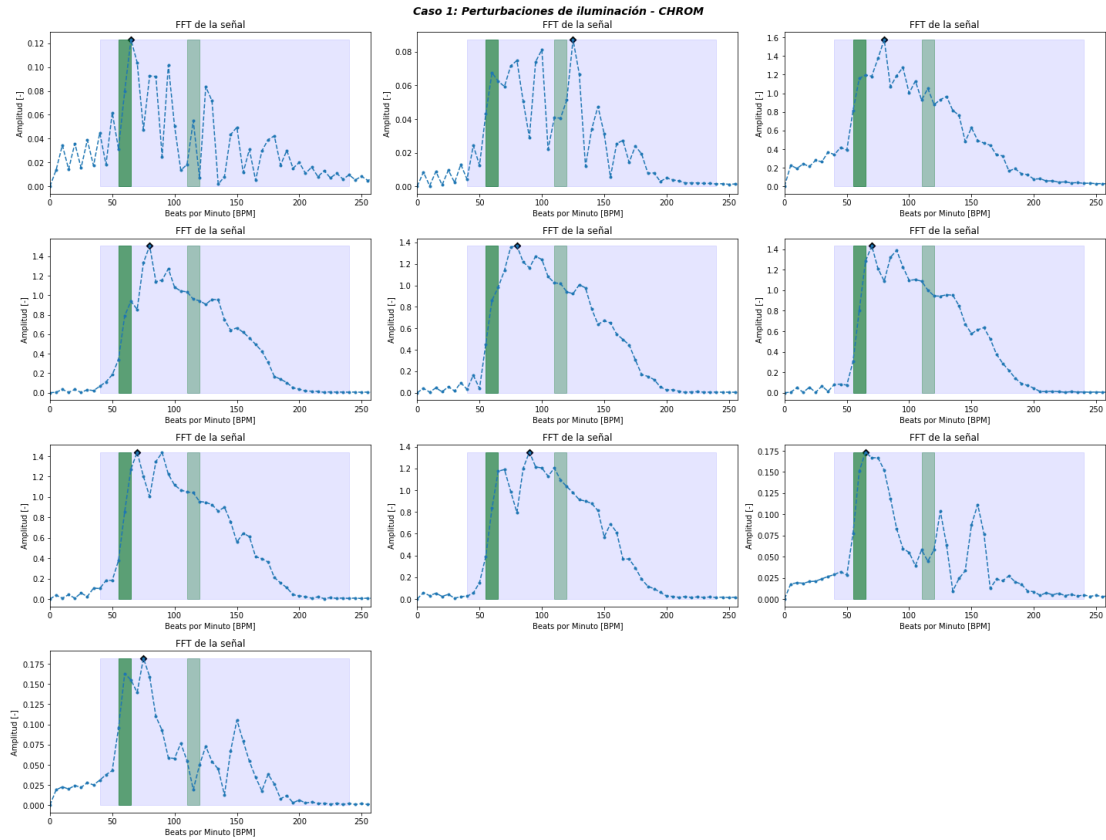


Figura 4.26: Espectro de todas las ventanas de análisis del método CHROM para el caso 1.

Por otro lado, se encuentra el método IVA+SSA+*Robust HR detection* en la Figura 4.27, donde se observa la clara tendencia del espectro de las ventanas a seguir el valor de la GT esperado. Se puede inferir que el método logra compensar la iluminación agresiva (Revisar Figura 3.5) con gran precisión en comparación con el método CHROM, en la mayoría de las ventanas, excepto en las últimas tres, donde se puede apreciar que el valor de la GT es atenuado en relación con las demás ventanas, resaltando componentes no deseados.

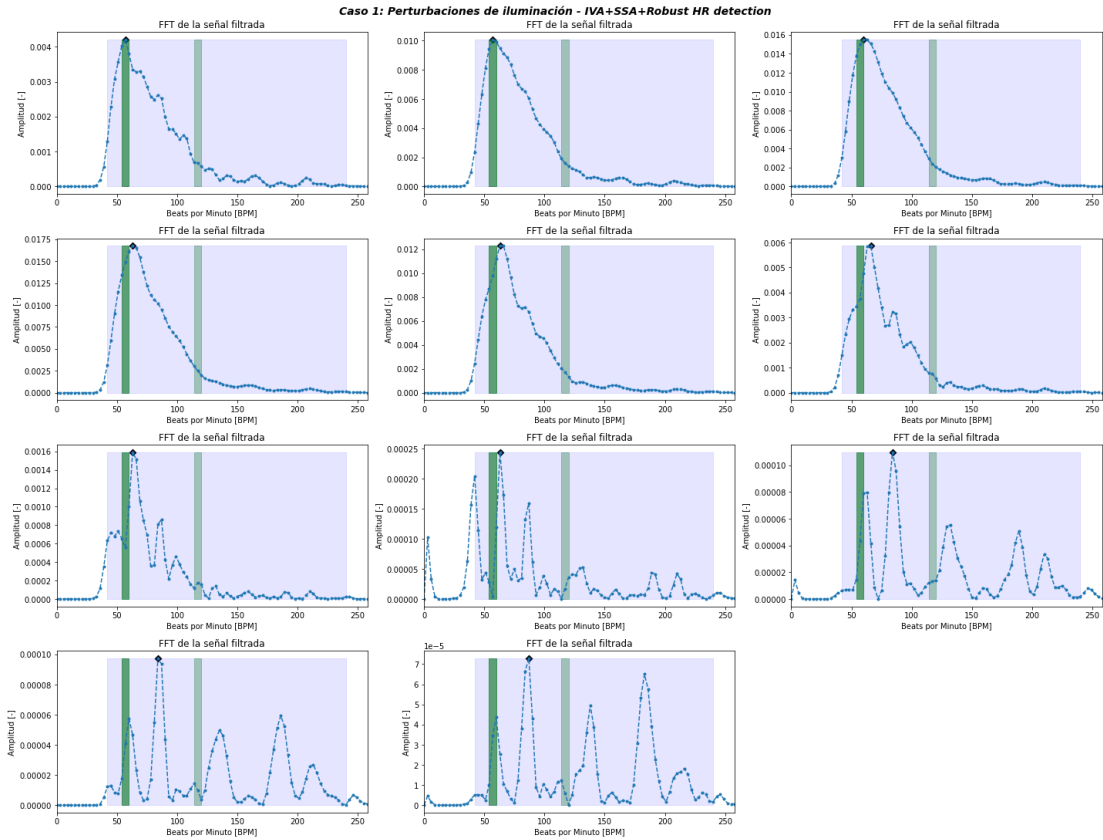


Figura 4.27: Espectro de todas las ventanas de análisis del método IVA+SSA+*Robust HR detection* para el caso 1.

Es relevante señalar que lograr una compensación precisa de las perturbaciones contribuye significativamente a mejorar la precisión de HR en entornos no controlados. Por lo tanto, el método IVA+SSA+*Robust HR detection* sobresale en esta situación. Sin embargo, no siempre existen perturbaciones en la escena y es posible que se presenten otros casos de interés, como estimar HR alto. A continuación se presenta un pequeño ejemplo de comparación sobre el Caso 2: Heart Rate alto mayor a 120[BPM].

En la Figura 4.28, se puede ver claramente que el método CHROM atenuó el valor real (representado por la barra de color verde oscuro) y resaltó el componente no deseado en 120 BPM aproximadamente. Comparando este método con IVA+SSA+*Robust HR detection* en la Figura 4.29, es evidente que CHROM resaltó un componente no deseado al realizar las operaciones de suma y resta de componentes. En cambio, IVA+SSA+*Robust HR detection* resalta completamente el valor esperado. Si bien el valor de 120 [BPM] aparece en el espectro, es prácticamente despreciable en términos de magnitud en comparación con el *peak* principal, que corresponde a la GT.

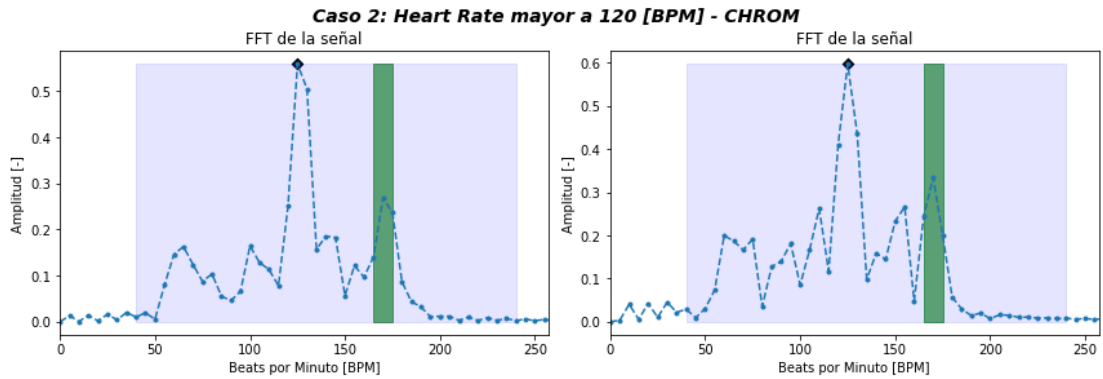


Figura 4.28: Espectro de las primeras dos ventanas de análisis del método CHROM para el caso 2.

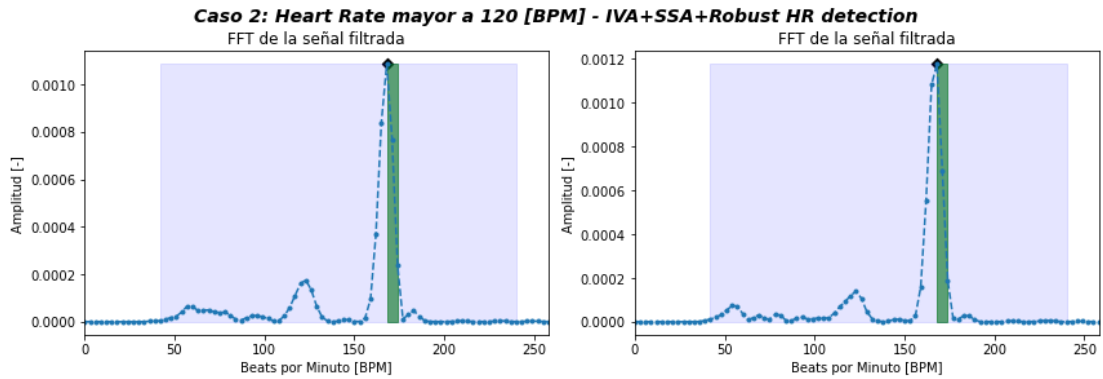


Figura 4.29: Espectro de las primeras dos ventanas de análisis del método IVA+SSA+Robust HR detection para el caso 2.

Por otro lado, IVA+SSA+Robust HR detection utiliza la información del fondo de la escena y la información contenida en el rostro, para identificar algún grado de correlación entre las señales y en caso de existir, compensarlas en el rostro. La principal debilidad que presenta esta metodología radica en el hecho de que algunos movimientos de la cabeza o el rostro generan variaciones de iluminación que no pueden ser compensadas, debido a que no tienen correlación con la iluminación del fondo. Estas perturbaciones, afectan en gran medida la calidad de la estimación de HR debido a que generan componentes de baja frecuencia y gran energía que se mantienen en las ventanas de análisis de las señales correspondientes, por lo que la metodología robusta de identificación de HR falla.

Para visualizar el problema descrito anteriormente, se muestra en la Figura 4.30 el espectro de las ventanas generadas a partir de una de las señales utilizadas para el análisis del método IVA+SSA+Robust HR detection que corresponde al canal G de *All mean*. En esta figura, se ve claramente que en ninguna de las ventanas destaca como tal

el *peak* que corresponde a la GT, más bien, destaca su entorno, que posiblemente sean movimientos del sujeto ya que a medida que se avanza en las ventanas, la magnitud de estos *peaks* va variando de manera no homogénea desplazándose levemente en el rango posible de HR. Es relevante destacar que *HR Robust detection* se ha desarrollado en esta metodología para utilizarlo en casos que involucran una gran cantidad de ruido. Sin embargo, cuando el sujeto realiza grandes movimientos que provocan variaciones no controladas en la iluminación del rostro, no es posible identificar el HR con gran precisión.

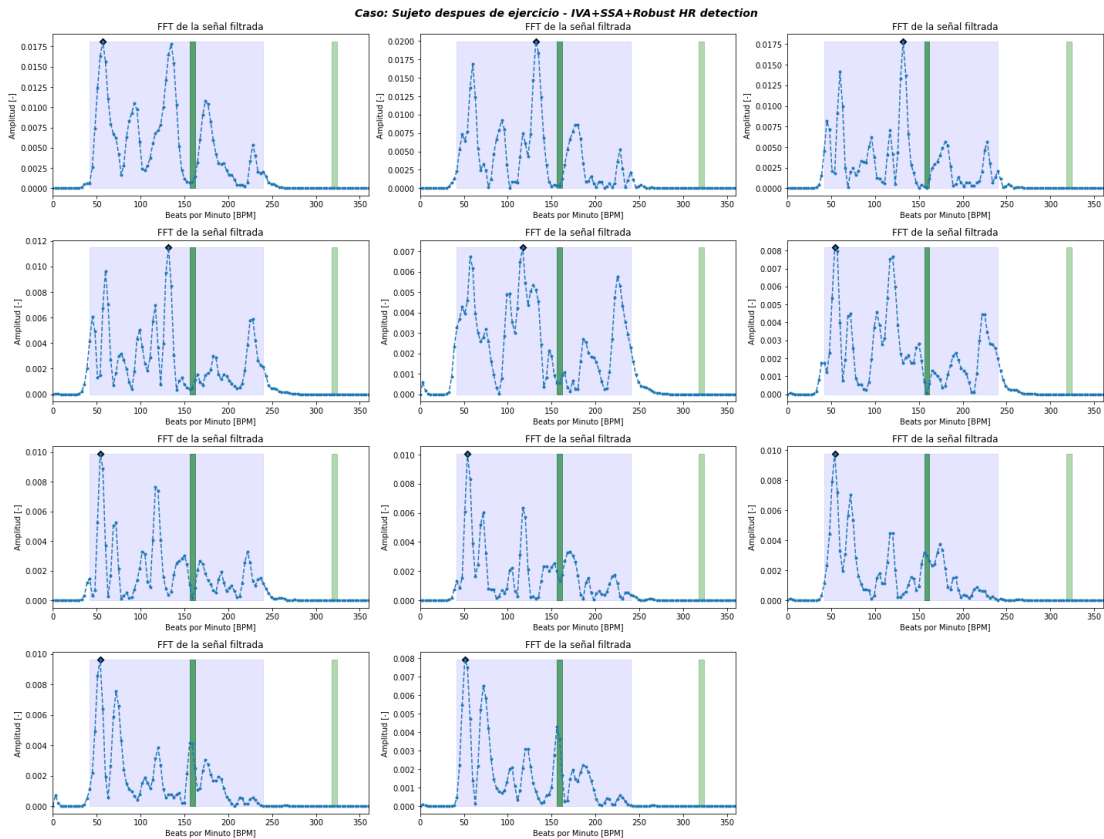


Figura 4.30: Espectro de todas las ventanas de análisis del método IVA+SSA+*Robust HR detection* para el caso donde un sujeto presenta perturbaciones que afectan a la estimación de HR.

Para realizar una comparación adecuada, en la Figura 4.31 se presenta el mismo caso. Dado que se trata de un sujeto que presenta un gran movimiento, se observa una gran cantidad de *peaks* en el espectro de las ventanas asociadas a la señal, similar al caso anterior, loscuáles corresponden a perturbaciones no deseadas. Aunque CHROM logra resaltar el *peak* que corresponde a la GT en algunos casos, en términos generales, no logra atenuar con precisión las perturbaciones. Por lo tanto, la estimación precisa de HR depende exclusivamente de la metodología diseñada para identificar la tendencia

del comportamiento de los *peaks* para una correcta estimación de los signos vitales. Sin una metodología robusta de estimación, el método CHROM por sí sólo no lograría una alta precisión.

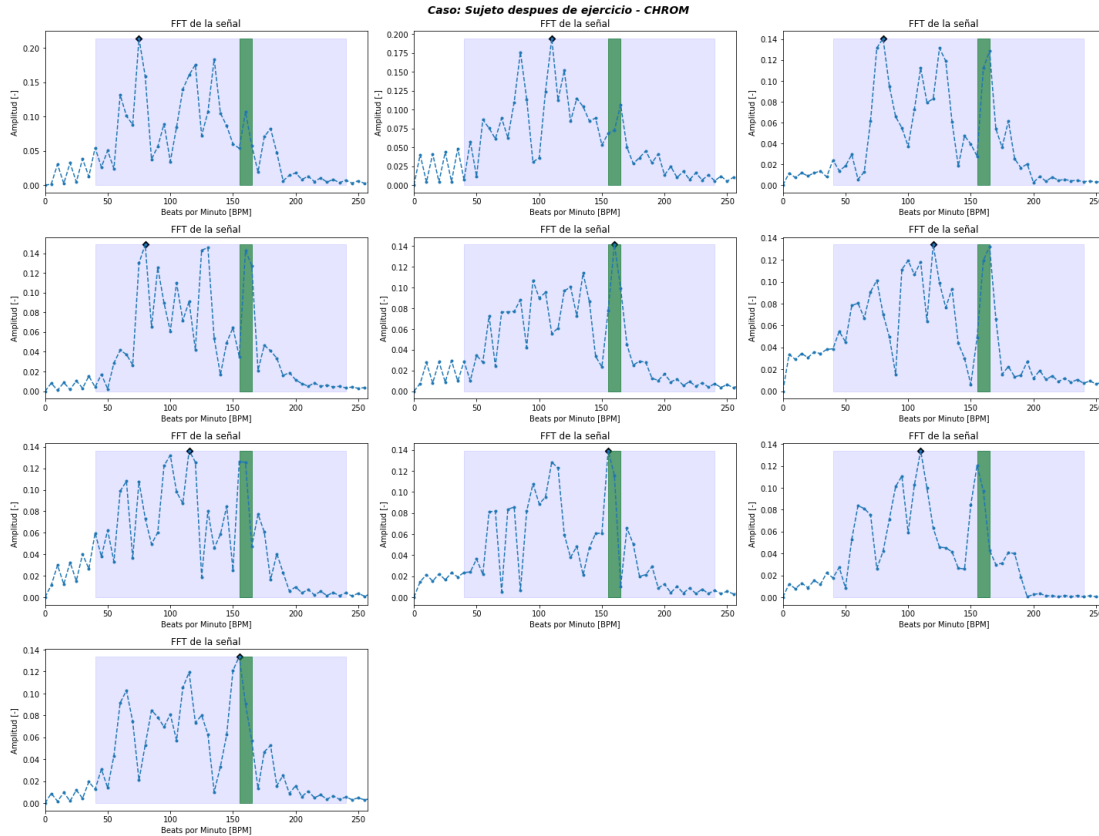


Figura 4.31: Espectro de todas las ventanas de análisis del método CHROM para el caso donde un sujeto presenta perturbaciones que afectan a la estimación de HR.

Es importante destacar que las dos metodologías fallan en ambientes no controlados. Una falla por falta de compensación de perturbaciones del sujeto y la otra falla por ignorar las condiciones del entorno. Es por esto que se propone una metodología conjunta que abarque la compensación global, IVA+SSA+Robust HR detection, junto con la compensación local, CHROM. Esta metodología conjunta permite abarcar, de forma teórica, un gran abanico de perturbaciones y estimar con una gran precisión los signos vitales. Sin embargo, al implementar esta propuesta no se logran buenos resultados respecto al desempeño individual de cada método por separado, si no que, crece el error de bastantes estimaciones en los rangos *Normal HR* y *High HR*, como se puede ver en la Figura 4.32.

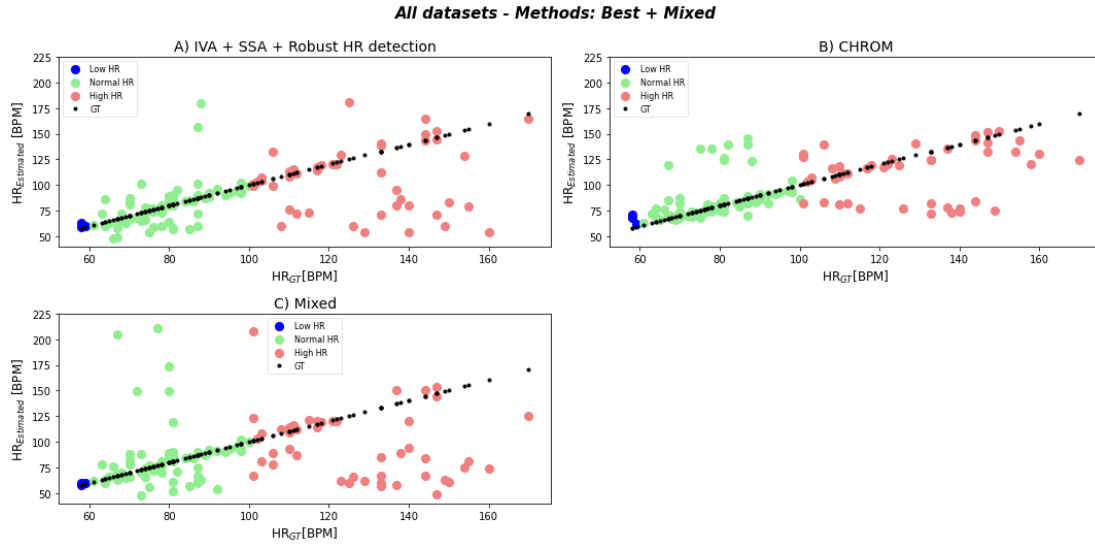


Figura 4.32: Gráficos HR_{GT} vs $HR_{Estimated}$ de los dos mejores métodos junto con su combinación evaluados sobre todos los Datasets.

Los errores en las estimaciones, de los métodos en conjunto, son claramente por el uso de la metodología combinada, por lo que las fallas individuales de cada método se amplifican. Por un lado, CHROM depende de las combinaciones lineales de los canales RGB, donde las mezclas están determinadas por valores estáticos que, en entornos no controlados, resaltan los componentes no deseados al no ser valores dinámicos que dependen de la escena. Por otro lado, IVA+SSA+Robust HR detection al compensar los componentes de alta correlación entre el fondo y el rostro, las señales RGB compensadas de las sub-ROIs presentan un desbalance en la proporción en la que se compensa cada sub-ROI, ya que se toma como una zona individual. Cada zona compensada puede mostrar variaciones en las señales RGB reconstruidas en términos de perturbaciones. Esto se debe a que los componentes fuente compensados utilizados para reconstruir la sub-ROI son diferentes en cada zona, lo que genera un desequilibrio en la reconstrucción de los canales RGB, aunque con las perturbaciones compensadas. Esto impacta directamente a CHROM, ya que las combinaciones lineales utilizadas para generar las señales de crominancia, necesarias para calcular la señal de rPPG, presentan valores estáticos de mezcla de señales en vez de ser dinámicos respecto a la escena. Por lo tanto, si las señales de las sub-ROIs ya han sido compensadas, el desequilibrio en la reconstrucción de IVA+SSA+Robust HR detection puede introducir un error considerable en algunas mediciones debido a los componentes estáticos de mezcla de CHROM, loscuáles resaltan componentes no deseados en este caso.

Para que la combinación de los mejores métodos sea comparable respecto a las implementaciones individuales, se presentan los casos de análisis en los que se probaron los mejores métodos para ver el efecto de la compensación y el calculo de HR con los métodos combinados.

Respecto al caso 1, se puede apreciar en la Figura 4.33 que en todas las ventanas se resalta el valor esperado de GT. Esto es posible debido a que las dos metodologías se complementaron gracias a los alcances de cada método. La compensación de CHROM es limitada respecto a la perturbación agresiva de este caso, como se puede ver en la Figura 4.26. Mientras que IVA+SSA+*Robust HR detection*, por sí sólo compensa la perturbación con gran precisión, como se puede ver en la Figura 4.27. Al combinar estos dos métodos, se observa que, para este caso específico, se complementan de manera precisa. En este sentido, se mejora la precisión en la estimación del valor de GT en el espectro en comparación con el método IVA+SSA+*Robust HR detection*, que realiza una compensación ligeramente menos precisa pero siempre mostrando una tendencia hacia GT.

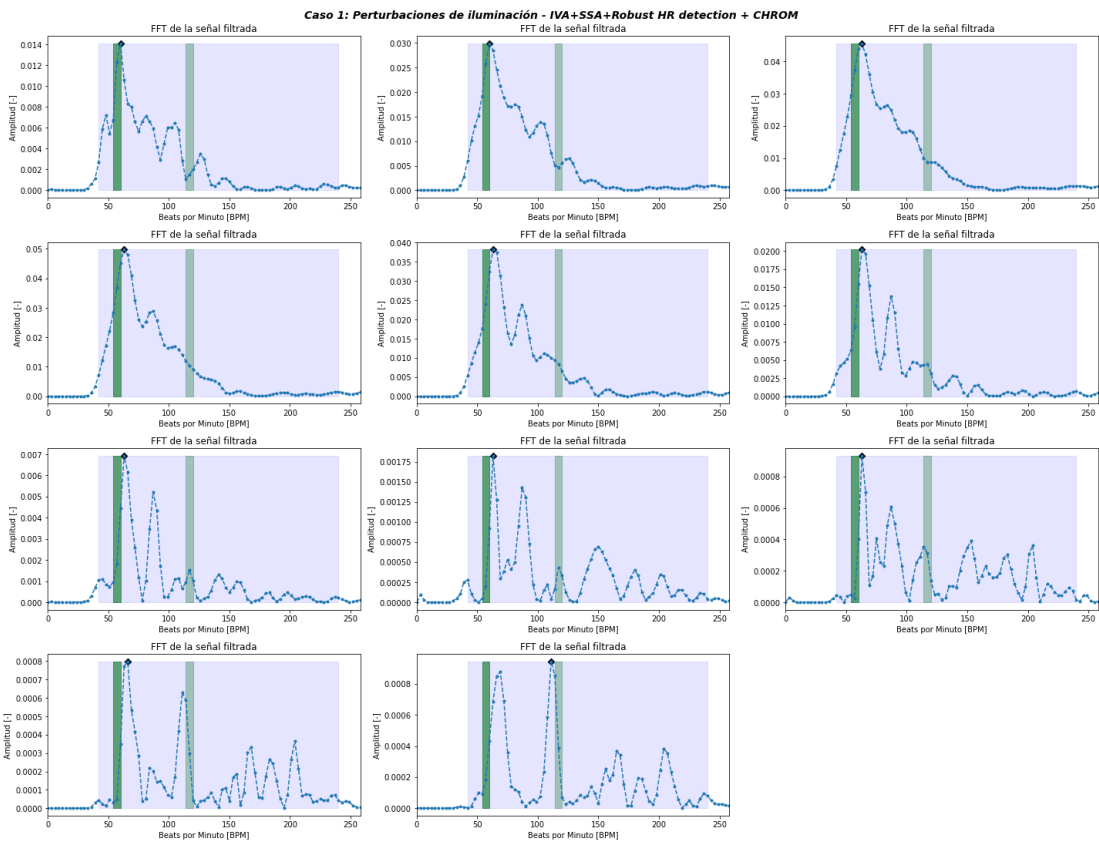


Figura 4.33: Espectro de todas las ventanas de análisis del método CHROM combinado con IVA+SSA+*Robust HR detection* para el caso 1.

En relación al caso 2, aunque el espectro de las ventanas muestra menos ruido, gracias al uso de IVA+SSA+*Robust HR detection*, en comparación con el método CHROM por sí solo, no logra estimar el valor esperado de la GT. Esto se debe a que el método CHROM falla al calcular la señal de rPPG, lo que resalta un componente no deseado en aproximadamente 120 BPM y atenúa el valor esperado de manera agresiva, como

se puede observar en la Figura 4.34. En este caso en particular, independientemente de la contribución de IVA+SSA+Robust HR detection, el método CHROM falla en la estimación de HR, error que también ocurre al utilizar el método solo, como se demuestra en la Figura 4.34.

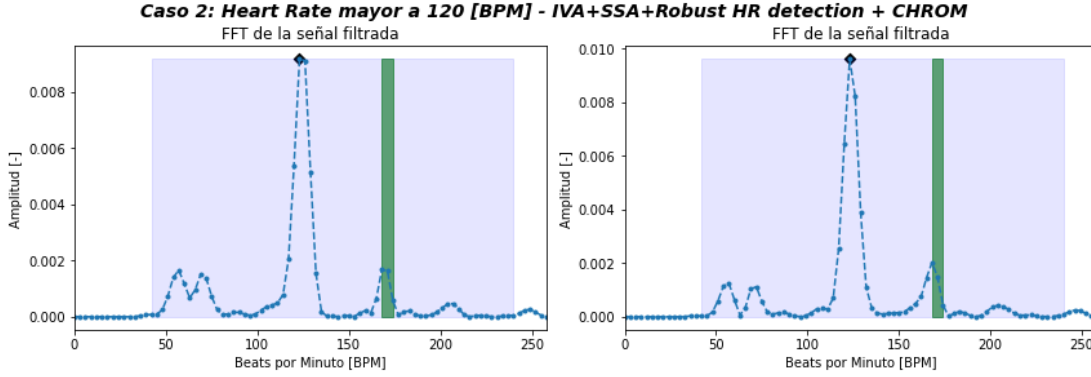


Figura 4.34: Espectro de las primeras dos ventanas de análisis del método CHROM combinado con IVA+SSA+Robust HR detection para el caso 2.

En relación al caso del sujeto después de ejercicio, comparando con los gráficos en las Figuras 4.31 y 4.30, que corresponden a la evaluación individual de los métodos CHROM e IVA+SSA+Robust HR detection, respectivamente. Se observa claramente que el espectro de todas las ventanas presenta menos ruido, lo que conlleva a la compensación de algunos componentes no deseados. Sin embargo, al compararlo con la metodología CHROM, si bien en el gráfico individual resalta en las últimas ventanas el *peak* correspondiente a la GT, en este caso destacan los componentes no deseados, especialmente en HR bajo, que se atribuye al movimiento asociado a la persona y que no es posible compensar en un primer paso por IVA+SSA+Robust HR detection. Esto genera un desequilibrio en la compensación y, en conjunto, ambos métodos producen una subestimación del valor esperado, resaltando los componentes de ruido por parte de ambos métodos.

Mediante este análisis específico de la metodología combinada, se ha logrado establecer los límites de funcionamiento de los métodos individuales, los cuáles no se pudieron determinar antes de aplicar la metodología conjunta. Este proceso ha permitido determinar que los dos métodos seleccionados son los mejores en términos del análisis global, de manera individual. Sin embargo, en conjunto, no alcanzan la precisión esperada en comparación con los alcances técnicos de los métodos individuales. Debido a lo anterior, no es posible establecer cuál de los métodos, CHROM o IVA+SSA+Robust HR detection, es el mejor en términos generales, debido a la naturaleza y alcance de cada método.

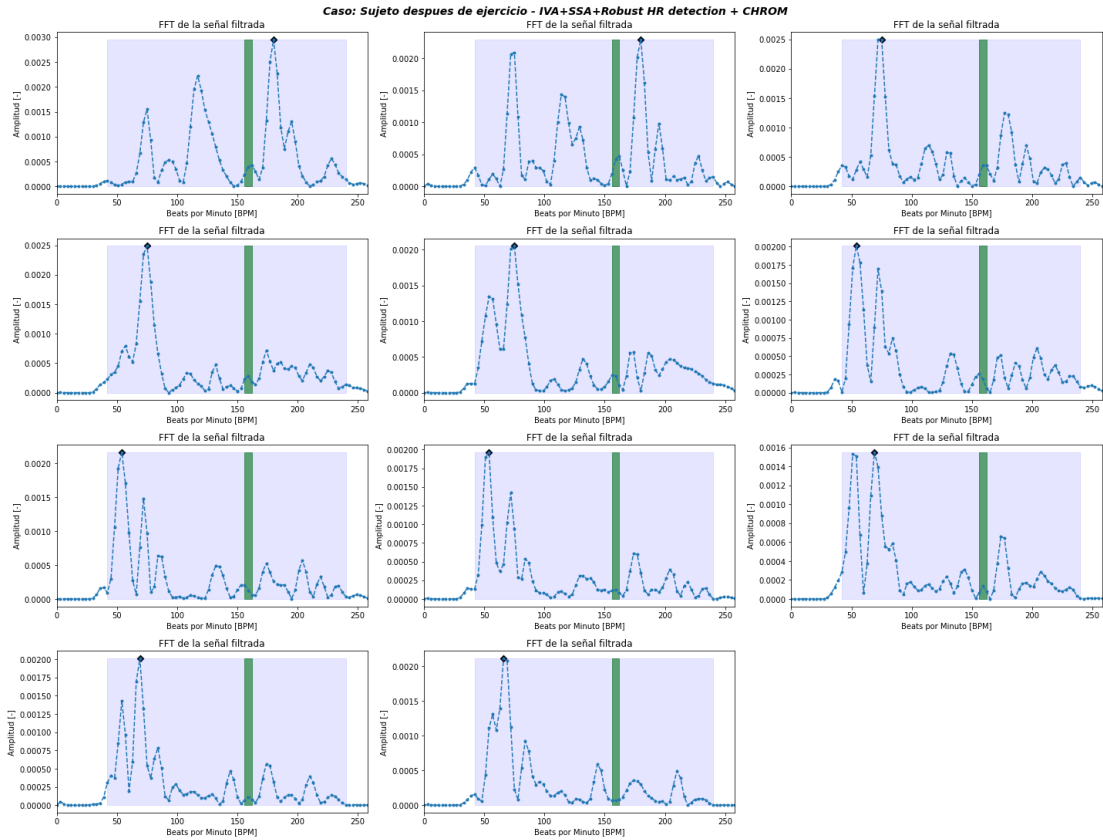


Figura 4.35: Espectro de todas las ventanas de análisis del método CHROM combinado con IVA+SSA+Robust HR detection para el caso donde un sujeto presenta perturbaciones que afectan a la estimación de HR.

Se procede a responder las preguntas de trabajo planteadas al inicio de este documento en la Sección 1.3.

- ¿Este proceso de compensación se puede hacer de forma rápida?
La rapidez de la compensación de iluminación depende exclusivamente del entorno y del tipo de perturbación presente. Si la compensación no afecta al entorno, aunque sea en un grado mínimo, y sólo afecta al rostro, es muy probable que no se pueda realizar la compensación. Esto se debe a que el algoritmo depende exclusivamente de la interacción de la perturbación en el rostro y el fondo. Por lo tanto, si no hay compensación, la rapidez es alta, como sucedió en el Dataset UBFC, que contiene un fondo con una mínima relación entre la iluminación del fondo y del rostro. En este caso, las compensaciones eran mínimas y rápidas. Además, la rapidez de la compensación depende exclusivamente de la agresividad de la perturbación. En el caso de una perturbación compleja, como el caso 1 analizado en las metodologías exploratorias 2, 3 y 4, en las secciones 3.2, 3.3 y 3.4, respectivamente, es posible que el proceso tome mucho tiempo debido a

la naturaleza compleja de la situación y a las múltiples correlaciones entre los componentes del fondo y el rostro, que superan el umbral de compensación. Por ende, cuanto mayor sea la correlación entre la iluminación del fondo y el rostro, más componentes en común se generarán para compensar, lo que implicaría una mayor demora en el proceso.

- ¿Se puede obtener una alta precisión en la inferencia de signos vitales con esta nueva metodología?

Para garantizar una precisión óptima en el cálculo de los signos vitales mediante una nueva metodología, resulta crucial emplear una estrategia robusta para la identificación de HR, tal como se propone en la metodología exploratoria 4 (Sección 3.4). Al utilizar esta metodología robusta para obtener la frecuencia cardíaca, se asegura la correcta identificación del rango posible de HR para el sujeto en el momento específico de la toma de datos.

Para recuperar la señal de rPPG, que contiene los demás signos vitales, es necesario realizar una retroproyección de los parámetros y buscar las señales, o bien construirlas directamente, mediante la aplicación de un filtro pasabanda en la frecuencia correspondiente al HR final, rango determinado por el método robusto. Con esta retroproyección y la aplicación del filtro pasabanda, se reconstruye la señal en el tiempo y es posible calcular los demás signos vitales. Con la aplicación de esta estrategia, se obtiene una señal libre de perturbaciones que sólo contiene la información de HR, junto con HRV y SpO₂.

- ¿Cuáles son los escenarios en donde se maximiza la precisión?

Respecto a los análisis anteriores, es posible identificar los escenarios donde se puede maximizar la precisión. Por lo general, se deben cumplir ciertos criterios en la escena:

- Respecto a la iluminación del sujeto, se debe emplear iluminación frontal siempre que sea posible, o asegurarse de que la iluminación dentro del rostro sea lo suficientemente uniforme para que la señal pulsátil mantenga una magnitud coherente en frecuencia en todas las sub-ROIs. Por lo tanto, se deben evitar aquellas posiciones de la fuente de iluminación que proyecten sombras en el rostro. En caso de que no se pueda controlar, es necesario ubicar al sujeto de manera que la iluminación sea lo más uniforme posible en el rostro, garantizando así la estabilidad de la señal objetivo.
- En relación con la iluminación de la escena, ya sea en un entorno controlado con iluminación artificial o en un ambiente con luz natural, es fundamental garantizar la homogeneidad de la iluminación en el rostro del sujeto. Esto es crucial, dado que el tipo de iluminación también incide en la calidad de la señal de rPPG.

En cuanto al comportamiento de los sujetos, es fundamental que se mantengan lo más quietos posible y que miren directamente a la cámara, cumpliendo con las

condiciones mencionadas anteriormente. Además, la persona debe evitar el uso de crema o maquillaje, la presencia de sudor u otra sustancia brillante u opaca en la piel que pueda afectar la incidencia de la iluminación ambiental hacia las venas. Esto se debe a que la estimación de los signos vitales depende exclusivamente de la señal pulsátil obtenida mediante rPPG, que se genera mediante la reflexión especular de la luz en la piel. La luz debe atravesar la piel y debe reflejar la información pulsátil de los cambios de volumen en la sangre.

Es importante recordar las limitaciones de funcionamiento de los mejores métodos, debido a que si no se logra cumplir los requerimientos de la escena y de los sujetos, es posible que las estimaciones fallen en algún punto.

Conclusión y trabajo futuro

Conclusiones

Durante el desarrollo de este trabajo, se analizaron las diferentes metodologías en el estado del arte para la compensación de iluminación de manera espacio-temporal. Estas compensaciones, en su mayoría, se centran únicamente en las perturbaciones temporales de la iluminación, sin abordar la característica espacial de la obtención de la señal de rPPG tanto en el rostro como en el fondo. Se identificaron las mejores metodologías con combinaciones de herramientas interesantes para la compensación temporal y se desarrollaron desde los artículos en los cuales se propusieron. Como no se cumplió de manera general resolver la problemática de compensación temporal, se diseñó una metodología para la compensación de perturbaciones, adoptando un enfoque espacio-temporal acorde a la problemática en cuestión. Esta metodología desarrollada, si bien logra ajustar los cambios bruscos de iluminación y corregir las perturbaciones en general que sean comunes con el fondo, debe complementarse con un algoritmo de detección e identificación robusta de *peaks* para una estimación precisa del Heart Rate(HR) y así lograr obtener la señal pulsátil que contiene los demás signos vitales. Por otro lado, se analizaron los métodos temporales Tradicionales para la obtención de la señal pulsátil que contiene los signos vitales, destacándose el método CHROM.

Durante la evaluación de la metodología propuesta en este trabajo se pudo establecer que, IVA+SSA+*Robust HR detection*, y el método Tradicional CHROM, son competitivos a nivel general en términos de precisión dependiendo del rango de HR de análisis. En especial, el método propuesto en este trabajo logra un gran desempeño a la hora de compensar perturbaciones en la escena, alcanzando una gran precisión de estimación en la mayoría de los casos. En ciertos casos, cuando no hay una gran correlación con la información del fondo, entraba en juego la metodología robusta de detección de *peaks* e identificación de HR, logrando alta precisión en Datasets donde la escena es estable. Por otro lado, el método CHROM logra un desempeño general competitivo tanto en precisión como en tiempo de ejecución, permitiendo su implementación, si se logra optimizar, para sesiones de telemedicina.

Dado que ambos métodos son competitivos por sí mismos, podría suponerse que al combinarlos se mejora la precisión. Sin embargo, al combinarlos, si bien en casos donde

las perturbaciones de iluminación son agresivas se logra mejorar la precisión levemente respecto a IVA+SSA+*Robust HR detection*, en los demás casos no, ya que interfieren entre sí al ser implementados de manera secuencial, primero IVA+SSA+*Robust HR detection* y luego CHROM. Esta interferencia en el funcionamiento pone de manifiesto la falencia principal de CHROM debido a cómo se construyen las señales de crominancia, que dependen de valores estáticos de mezcla para su construcción. Por lo tanto, en ambientes no controlados, CHROM no necesariamente estimará con buena precisión los signos vitales debido a la falta de flexibilidad y adaptabilidad en la construcción de las señales de crominancia en entornos no controlados. Por otro lado, IVA+SSA+*Robust HR detection* cumple con el propósito de compensar las perturbaciones de iluminación que afectan a la escena completa. Sin embargo, presenta debilidades frente a las perturbaciones de iluminación generadas por movimientos del rostro. Para maximizar la precisión de HR, es necesario aplicar directivas que aseguren una iluminación homogénea en el rostro y minimicen el impacto de las perturbaciones locales. De esta manera, la parte de detección robusta de este método sería capaz de calcular con gran precisión los signos vitales, gracias a su adaptabilidad respecto a la información proporcionada por la caracterización espacio-temporal de las señales obtenidas de todas las sub-ROIs identificadas en el rostro.

Trabajo Futuro

Los trabajos futuros contemplan detalles de implementación necesarios para una respuesta en tiempo real, como también el añadir funcionalidades extra a la metodología desarrollada en este trabajo para robustecer su funcionamiento. Las propuestas son:

- Implementación de todo el sistema de manera local en un teléfono, para la estimación robusta de signos vitales en tiempo real.
- Optimización de la metodología de compensación: La metodología IVA+SSA+*Robust HR detection* logra compensar las perturbaciones del entorno con gran precisión, lo que la hace útil para casos de aplicación real en entornos no controlados, como sesiones de telemedicina. Sin embargo, debido a su costo computacional, es necesario optimizar su implementación para cumplir con la restricción de tiempo real asociada a la sesión de telemedicina.
- Robustecer la metodología IVA+SSA+*Robust HR detection* para compensar las perturbaciones locales del rostro para maximizar la precisión cuando los sujetos presenten movimientos.
- Creación de Dataset con perturbaciones controladas: Considerando los tres Datasets a los que se tuvo acceso en este trabajo, sólo en NAM-EDGE existían casos de perturbación de iluminación agresiva en el entorno, los cuales fueron adecuadamente identificados debido a la naturaleza de estos al obtener las señales de

rPPG. Se debe investigar con mayor profundidad el efecto de las perturbaciones de la escena en conjunto con las perturbaciones locales en el rostro. Para realizar este análisis, se debe crear un Dataset que contenga perturbaciones sincrónicas con la escena y que permitan establecer la robustez de la metodología desarrollada en este trabajo.

- Evaluación de metodología propuesta combinada con métodos tradicionales: La metodología propuesta en este trabajo puede implementarse como bloques de pre y post procesamiento de metodologías ya existentes. Esto añadiría robustez a las perturbaciones y a la identificación de los valores de HR a cualquier método al que se le agreguen estos bloques. Es necesario estudiar el efecto e identificar si la mejora es considerable en relación con el costo/beneficio.

Bibliografía

- [1] Sujan Adhikari and Sanjeeb Prasad Panday. Image enhancement using successive mean quantization transform and homomorphic filtering. In *2019 Artificial Intelligence for Transforming Business and Society (AITB)*, volume 1, pages 1–5, 2019.
- [2] Matthew Anderson, Tülay Adalı, and Xi-Lin Li. Joint blind source separation with multivariate gaussian model: Algorithms and performance analysis. *IEEE Transactions on Signal Processing*, 60(4):1672–1683, 2012.
- [3] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019. Award Winning Papers from the 23rd International Conference on Pattern Recognition (ICPR).
- [4] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro D’Amelio, Giuliano Grossi, and Raffaella Lanzarotti. An open framework for remote-PPG methods and their assessment. *IEEE Access*, pages 1–1, 2020.
- [5] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro D’Amelio, Giuliano Grossi, Raffaella Lanzarotti, and Edoardo Mortara. pyvhr: a python framework for remote photoplethysmography. *PeerJ Computer Science*, 8:e929, 2022.
- [6] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *CoRR*, abs/0912.3599, 2009.
- [7] Juan Cheng, Xun Chen, Lingxi Xu, and Z. Jane Wang. Illumination variation-resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition. *IEEE Journal of Biomedical and Health Informatics*, 21(5):1422–1433, 2017.
- [8] Gerard de Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.
- [9] Michael Van den Bergh, Xavier Boix, Gemma Roig, and Luc Van Gool. Seeds: Superpixels extracted via energy-driven sampling, 2013.

- [10] Nina Golyandina and Anatoly Zhigljavsky. *Basic SSA*, pages 11–70. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [11] Jiawei Guo, Jieming Ma, Ángel F. García-Fernández, Yungang Zhang, and Haining Liang. A survey on image enhancement for low-light images. *Heliyon*, 9(4):e14558, 2023.
- [12] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017.
- [13] Wouter Hoffman and Daniel Lakens. Public benchmark dataset for testing rppg algorithm performance, 2020.
- [14] Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. Real-time facial surface geometry from monocular video on mobile gpus, 2019.
- [15] J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.
- [16] Taesu Kim, Intae Lee, and Te-Won Lee. Independent vector analysis: Definition and algorithms. In *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, pages 1393–1396, 2006.
- [17] Antony Lam and Yoshinori Kuno. Robust heart rate measurement from video using select random patches. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 3640–3648, 2015.
- [18] Edwin H. Land and John J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, Jan 1971.
- [19] Isabell Lehmann, Evrim Acar, Tanuj Hasija, M.A.B.S. Akhonda, Vince D. Calhoun, Peter J. Schreier, and Tülay Adalı. Multi-task fmri data fusion using iva and parafac2. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1466–1470. IEEE, 2022.
- [20] Xuenan Liu, Xuezhi Yang, Dingliang Wang, and Alexander Wong. Detecting pulse rates from facial videos recorded in unstable lighting conditions: An adaptive spatiotemporal homomorphic filtering algorithm. *IEEE Transactions on Instrumentation and Measurement*, 70:1–15, 2021.
- [21] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Ubowaja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines, 2019.

- [22] Magdalena Madej and Jędrzej Nowak. Measuring pulse rate with a webcam. *Journal of Medical Imaging and Health Informatics*, 2:87–92, 03 2012.
- [23] J. Martin Bland and Douglas G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986. Originally published as Volume 1, Issue 8476.
- [24] Ewa Magdalena Nowara, Tim K. Marks, Hassan Mansour, and Ashok Veeraraghavan. Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1353–135309, 2018.
- [25] Madhab Pal, Rajib Roy, Joyanta Basu, and Milton S. Bepari. Blind source separation: A review and analysis. In *2013 International Conference Oriental COCODA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*, pages 1–5, 2013.
- [26] Christian S. Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. Local group invariance for heart rate estimation from face videos in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1335–13358, 2018.
- [27] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express*, 18(10):10762–10774, May 2010.
- [28] Rencheng Song, Jiji Li, Minda Wang, Juan Cheng, Chang Li, and Xun Chen. Remote photoplethysmography with an eemd-mcca method robust against spatially uneven illuminations. *IEEE Sensors Journal*, 21(12):13484–13494, 2021.
- [29] Arthur Tenenhaus and Michel Tenenhaus. Regularized generalized canonical correlation analysis. *Psychometrika*, 76:257–284, 04 2011.
- [30] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.
- [31] Wencheng Wang, Xiaojin Wu, Xiaohui Yuan, and Zairui Gao. An experiment-based review of low-light image enhancement methods. *IEEE Access*, 8:87884–87917, 2020.
- [32] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.
- [33] ZHAOHUA WU and NORDEN E. HUANG. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 01(01):1–41, 2009.

- [34] Lingxi Xu, Juan Cheng, and Xun Chen. Illumination variation interference suppression in remote ppg using pls and memd. *Electronics Letters*, 53, 12 2016.
- [35] Shuchang Xu, Lingyun Sun, and Gustavo Kunde Rohde. Robust efficient estimation of heart rate pulse from video. *Biomed. Opt. Express*, 5(4):1124–1135, Apr 2014.
- [36] Ruo-Nan Yin, Rui-Sheng Jia, Zhe Cui, Jin-Tao Yu, Yan-Bin Du, Li Gao, and Hong-Mei Sun. Heart rate estimation based on face video under unstable illumination. *Applied Intelligence*, 51, 08 2021.
- [37] Boyuan Zhang, Hengkang Li, Lisheng Xu, Lin Qi, Yudong Yao, and Stephen Greenwald. Noncontact heart rate measurement using a webcam, based on joint blind source separation and a skin reflection model: For a wide range of imaging conditions. *Journal of Sensors*, 2021:1–18, 07 2021.
- [38] Qing Zhang, Yongwei Nie, and Wei-Shi Zheng. Dual illumination estimation for robust exposure correction. *Computer Graphics Forum*, 38:243–252, 10 2019.
- [39] Marcos Zúñiga. Integral telehealth solution for elderly care, integrating videoconferencing and automatic detection of vital signs through artificial vision, 2022.