

E.T.S. de Ingeniería Industrial, Informática
y de Telecomunicación

Comparativa de algoritmos para Fotopletismografía
Remota e implementación de una arquitectura en
tiempo real para cámaras web de bajo coste



Grado en Ingeniería
en Tecnologías de Telecomunicación

Trabajo Fin de Grado

Daniel Adín Ferrer

Rafael Cabeza Laguna

Pamplona, 26/10/2020

Resumen

A lo largo de este trabajo se busca la implementación de las técnicas de Fotoplethismografía Remota (rPPG) para estimar la frecuencia cardíaca (FC) de un individuo a partir de un vídeo de su cara capturado por una cámara de bajo coste.

Esta implementación se divide principalmente en dos vertientes: modo *offline* y modo *online*.

En el modo *offline*, se analizan los vídeos de individuos recogidos en dos bases de datos distintas y las estimaciones obtenidas de FC como resultado se comparan con los valores de FC *groundtruth* calculados a partir de los archivos generados por los sensores fisiológicos durante las pruebas.

Y en el modo *online*, se implementa este análisis en tiempo real para estimar la FC de un individuo mientras es grabado por una webcam.

Palabras Clave: Fotoplethismografía Remota, rPPG, Estimación Frecuencia Cardíaca, FC.

Abstract

Throughout this work, the implementation of the Remote Photoplethysmography techniques (rPPG) is searched to estimate the heart rate (HR) of an individual from a video of their face captured by a low-cost camera.

This implementation is mainly divided in two paths: offline mode and online mode.

In offline mode, the videos of individuals collected in two different databases are analysed and the estimates of HR obtained as results are compared with the groundtruth HR values calculated from the files generated by the physiological sensors during the tests.

And in online mode, this analysis is implemented in real time to estimate de HR of an individual while is recording by a webcam.

Keywords: Remote Photoplethysmography, rPPG, Heart Rate Estimation, HR.

Índice

| | |
|---|----|
| Introducción | 5 |
| 1.1 Objetivos | 6 |
| 1.2 Marco teórico..... | 7 |
| 1.2.1 Frecuencia cardíaca..... | 7 |
| 1.2.2 Electrocardiograma (ECG) | 8 |
| 1.2.3 Fotopletismografía (PPG) | 10 |
| 1.3 Estado del arte | 11 |
| 1.3.1 Soluciones propuestas a las variaciones de iluminación..... | 13 |
| 1.3.2 Soluciones propuestas al movimiento | 16 |
| Materiales y Métodos | 22 |
| 2.1 Bases de datos..... | 23 |
| 2.1.1 MAHNOB HCI..... | 24 |
| 2.1.2 COHFACE | 28 |
| 2.2 Herramientas y algoritmos..... | 32 |
| 2.2.1 Obtención de la FC <i>groundtruth</i> (modo offline) | 32 |
| Algoritmo Pan-Tompkins | 33 |
| 2.2.2 Obtención y tratamiento de la ROI facial..... | 35 |
| Algoritmo iCCR-master | 35 |
| Máscaras: filtrado espacial. | 37 |
| Rectificación de iluminación | 39 |
| Filtrado temporal | 41 |
| 2.2.3 Extracción de la señal BVP..... | 43 |
| Green Verkruysse (GVERK) | 44 |
| POS..... | 45 |
| CHROM..... | 46 |
| ICA..... | 47 |
| 2.3 Arquitectura del modo online (GUI)..... | 48 |
| Resultados | 50 |
| 3.1 Resultados <i>Kooij-Naber (MAHNOB HCI)</i> | 51 |
| 3.2 Resultados <i>iPhys-toolbox (MAHNOB HCI)</i> | 54 |
| 3.3 Resultados <i>iPhys-toolbox (COHFACE)</i> | 64 |
| Conclusiones y líneas futuras | 72 |
| Bibliografía | 74 |

Apartado 1

Introducción

La frecuencia cardíaca de un individuo es fundamental en la determinación de su estado de salud, tanto a nivel fisiológico como mental. Así pues, su interés y relevancia en los ámbitos médico y social, centra el punto de mira de numerosas líneas de investigación.

En la actualidad, existen diversos métodos y técnicas para la monitorización de este parámetro fisiológico. Desde técnicas que emplean sensores dispuestos directamente sobre la piel (métodos con contacto) hasta técnicas más avanzadas basadas en sistemas radar, luz estructurada o iluminación ambiental (métodos sin contacto).

Este trabajo de fin de estudios se centra en esta última vertiente: método sin contacto basado en iluminación ambiental. Esta técnica aprovecha las variaciones de color cutáneas, concretamente de la tez, o los movimientos inducidos en el sujeto asociados a la actividad cardiovascular, para estimar la frecuencia cardíaca del individuo.

Estas variaciones, a pesar de ser imperceptibles para el ojo humano, no escapan al detalle de una cámara digital, y pueden ser tratadas con rigurosidad mediante software. Esta praxis, recogida en numerosas investigaciones recientes, recibe el nombre de fotopletiografía remota o rPPG (acrónimo del nombre en inglés, *Remote Photoplethysmography*).

En el presente documento, se implementa el rPPG para la estimación de la frecuencia cardíaca en dos bases de datos con diferentes sujetos (modo *offline*), así como en directo, mediante la grabación del usuario a tiempo real por webcam (modo *online*).

Por tanto, la estructura a seguir en las páginas posteriores es la siguiente:

En el primer apartado, se plasma el objeto del proyecto y se introduce el rPPG bajo un marco de fundamentos teóricos esenciales para su comprensión. A su vez, se recoge el estado del arte con los últimos avances de investigación hasta la fecha, en lo que a rPPG se refiere.

En el segundo apartado, se detalla la metodología a seguir, la cual reúne la descripción tanto de las bases de datos analizadas, como la información relativa a las etapas del procesado y los algoritmos empleados.

En el tercer apartado, se expone un análisis comparativo de los resultados obtenidos mediante rPPG con las dos bases de datos, sin perder la referencia de los resultados predecesores recogidos en publicaciones anteriores.

Y en el cuarto y último apartado, se esbozan unas conclusiones y líneas futuras sobre esta investigación.

1.1 Objetivos

El objetivo principal del proyecto es realizar un análisis comparativo a detalle sobre el comportamiento de las técnicas existentes de fotopleletismografía remota (rPPG) bajo diferentes condiciones físicas y ambientales.

Para ello, se subdivide el proyecto en dos vertientes: modo *offline* y modo *online*.

- En el modo *offline*, se ejecuta el análisis rPPG en cascada sobre distintos individuos, cuyos videos están recopilados en dos bases de datos de diferente origen: *MAHNOB* ^[1] y *COHFACE* ^[2].
- En el modo *online*, se adquiere la imagen del usuario mediante una webcam convencional, y se procesa en directo, obteniendo resultados a tiempo real.

Los resultados del modo *offline* brindan una visión global y objetiva sobre comportamiento del rPPG bajo diferentes condiciones de laboratorio, dependiendo del algoritmo empleado, el propio sujeto, tipo de iluminación ambiental, equipo disponible, etc.

Por otra parte, los resultados a tiempo real del modo *online* permiten evaluar de manera más laxa y subjetiva el comportamiento de esta praxis, pudiendo testear *in situ* distintas condiciones de análisis, como por ejemplo el tipo de iluminación o distintos sujetos.

1.2 Marco teórico

1.2.1 Frecuencia cardíaca

La frecuencia cardíaca (FC) es el número de veces que se contrae el corazón durante un minuto (latidos por minuto). Una FC normal, en reposo, oscila entre 50 y 100 latidos por minuto en las personas adultas. En los niños, esta FC es más elevada debido a las necesidades de su crecimiento.

El ejercicio físico o las situaciones de estrés provocan un aumento de la FC (taquicardia sinusal), que evidentemente se considera normal. La FC máxima que una persona puede alcanzar durante un ejercicio físico intenso es individual, siendo más baja conforme se avanza en edad. Existen varias fórmulas para calcularla, siendo la más conocida la de: $FC \text{ máxima} = 220 - \text{edad}$.

La frecuencia cardíaca puede ser medida directamente desde la muñeca, u otras zonas del cuerpo como la sien o el cuello. Además, existen muchos dispositivos orientados a esta medición e inclusive aquellos que miden la presión arterial también indican la FC. Es importante tener en cuenta que la FC varía directamente con la actividad desarrollada por el individuo en ese momento, por lo que ésta debería medirse en reposo, sentado y en un ambiente templado [3].

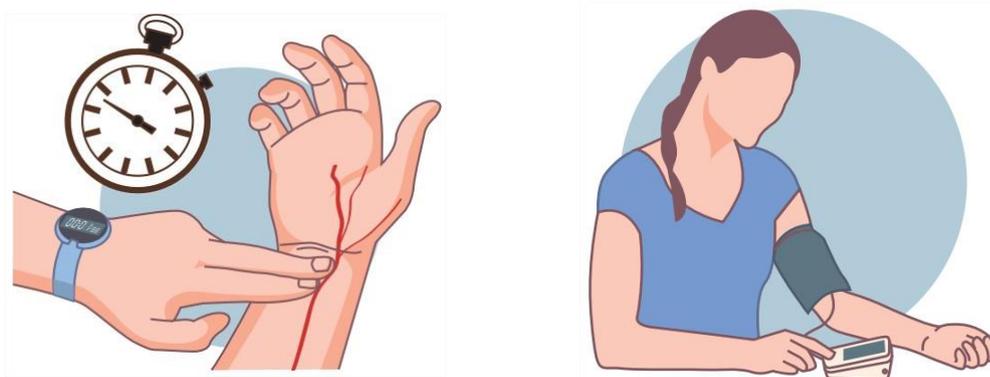


Fig. 1.1: Ejemplo sobre 2 técnicas de medición de la FC.

Una conversión rápida de unidades, muy útil para la comprensión del rango de frecuencias utilizado en las páginas posteriores es la siguiente:

45 pulsaciones por minuto (bpm), freq. mín. corazón → 0.75 variaciones por segundo (Hz).

220 pulsaciones por minuto (bpm), freq. máx. corazón → 3.67 variaciones por segundo (Hz).

1.2.2 Electrocardiograma (ECG)

El electrocardiograma o ECG es un gráfico en el que se estudian las variaciones de voltaje en función del tiempo para registrar la actividad de la corriente eléctrica que se está desarrollando en el corazón durante un tiempo determinado (un ECG normal no suele exceder de los 30 segundos).

La actividad eléctrica del corazón recogida en el ECG se observa en forma de un trazado que presenta diferentes deflexiones (ondas de ECG) que se corresponden con el recorrido de los impulsos eléctricos a través de las diferentes estructuras del corazón.

Descripción de un ECG

El ECG presenta como línea guía la denominada línea isoelectrónica o línea basal, que puede identificarse fácilmente como la línea horizontal existente entre cada latido. Los latidos cardíacos quedan representados en el ECG *normal* por las diferentes oscilaciones de la línea basal en forma de ángulos, segmentos, ondas e intervalos, constituyendo una imagen característica que se repite con una frecuencia regular. El recorrido en sentido horizontal hace referencia al tiempo transcurrido, mientras que la distancia en sentido vertical (altura de picos) al voltaje que se está produciendo.

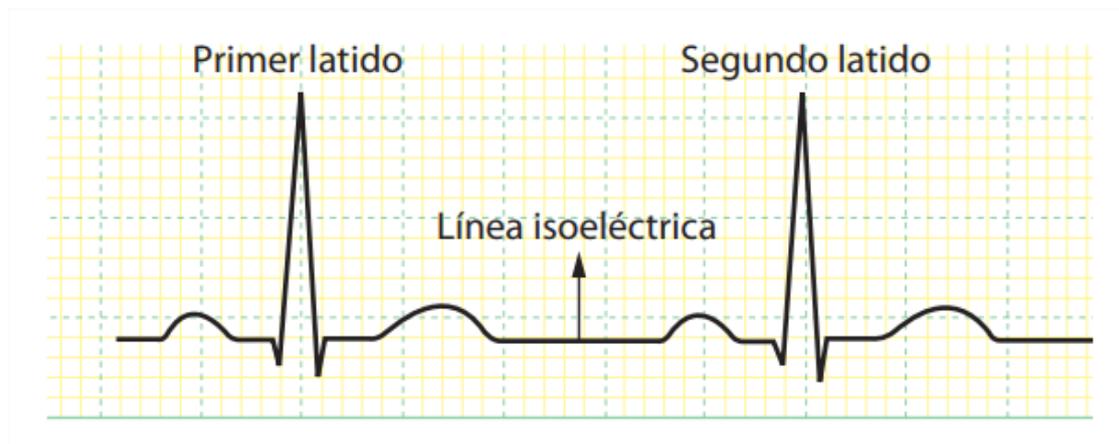


Fig. 1.2: Representación de dos latidos consecutivos en el ECG.

En un ECG *normal*, como se observa en el esquema de la fig. 1.3, cada complejo consta de una serie de deflexiones (ondas ECG) que alternan la línea basal. Realizando la lectura de izquierda a derecha, se distinguen la onda P, el segmento P-R, el complejo QRS, el segmento ST y finalmente la onda T ^[73].

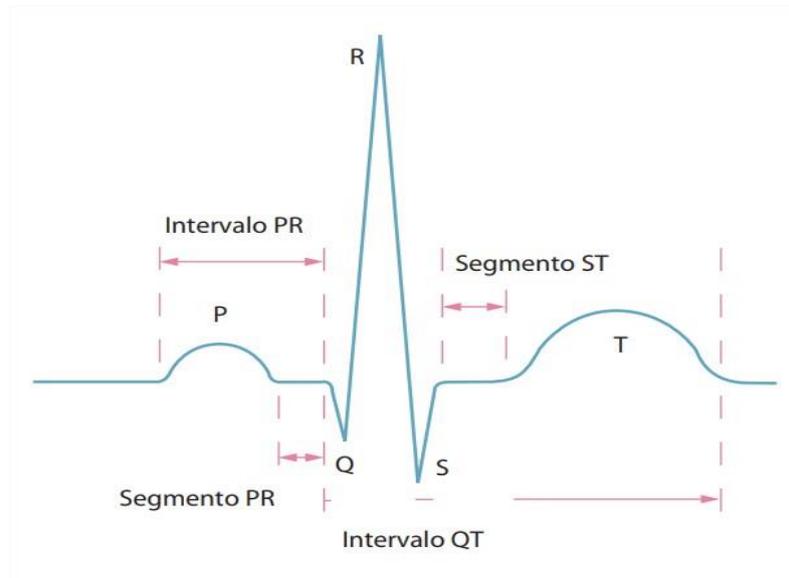


Fig. 1.3: Esquema descriptivo de un ECG.

Cuando se mide el pulso cardíaco, se trata de localizar temporalmente los complejos QRS que se corresponden con los momentos en los que los ventrículos se contraen y expulsan su contenido sanguíneo (despolarización ventricular). Si se localizan los complejos QRS en un tiempo suficientemente largo (unos 30 segundos), la deducción de la FC es trivial. Considerando los picos R como marcadores temporales, por ser este pico el más significativo del complejo QRS, y conociendo la duración de la medición (distancia horizontal entre el primer y último pico R), aplicando la siguiente fórmula (Ecuación 1) se obtienen las pulsaciones por minuto:

$$\text{Pulsaciones por minuto (BPM)} = \frac{n_R \cdot 60}{t} \quad [\text{Ec. 1}]$$

Siendo n_R el número de picos R del complejo QRS localizados en un tiempo t de observación.

1.2.3 Fotopletismografía (PPG)

La Pletismografía es una técnica por la cual se registra de manera no invasiva las variaciones de volumen sanguíneo en las diferentes partes del cuerpo de una persona, comúnmente las extremidades.

La Fotopletismografía o PPG, como se puede intuir del propio nombre, no es más que una vertiente de Pletismografía cuya medida está basada en los cambios de absorción de la luz incidente en la piel provocados por el bombeo de sangre en cada pulsación. Normalmente, en este tipo de mediciones se emplea luz estructurada o infrarroja por contacto [4].

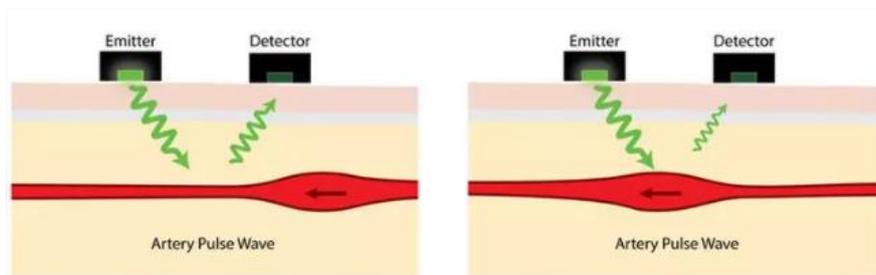


Fig. 1.4: Ejemplo detección de los cambios de absorción de luz mediante sensores PPG.

En este trabajo se presenta la Fotopletismografía Remota o rPPG, la cual es una mera adaptación de la Fotopletismografía tradicional a los métodos *sin contacto* (de ahí la palabra *Remota* en el nombre). Los fundamentos teórico-prácticos son los mismos que la técnica original, solo que esta vez, la captura de datos para la medición se realiza a distancia, mediante una cámara de vídeo orientada al rostro del individuo en cuestión. La estimación en la medida, se realiza a posteriori mediante procesado de imagen (Véase el apartado 2, *Materiales y Métodos*, pág. 23).

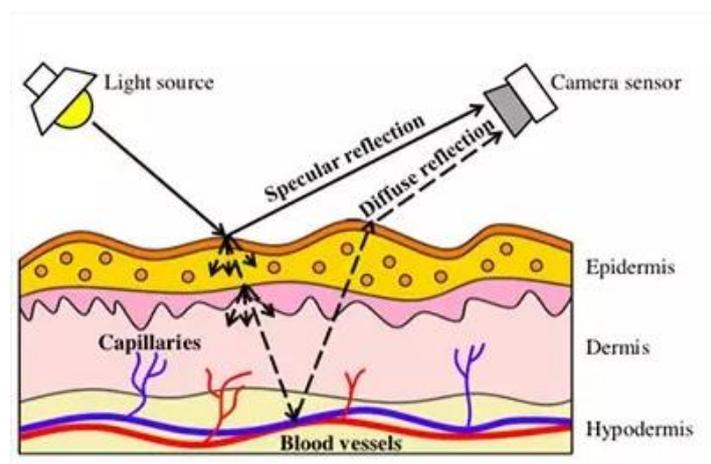


Fig. 1.5: Ejemplo detección de los cambios de absorción de luz mediante cámara de vídeo (rPPG)

1.3 Estado del arte

En esta sección, se exponen los últimos avances de investigación hasta la fecha sobre los métodos de estimación de la frecuencia cardíaca (FC) mediante fotopletismografía remota (rPPG).

En la última década, se han publicado numerosos artículos sobre distintas técnicas de rPPG debido al creciente interés en la investigación de esta praxis. Esta tendencia se puede apreciar con claridad en la siguiente figura, fig. 1.6, donde se clasifican cronológicamente los 111 artículos de revisión publicados más relevantes.

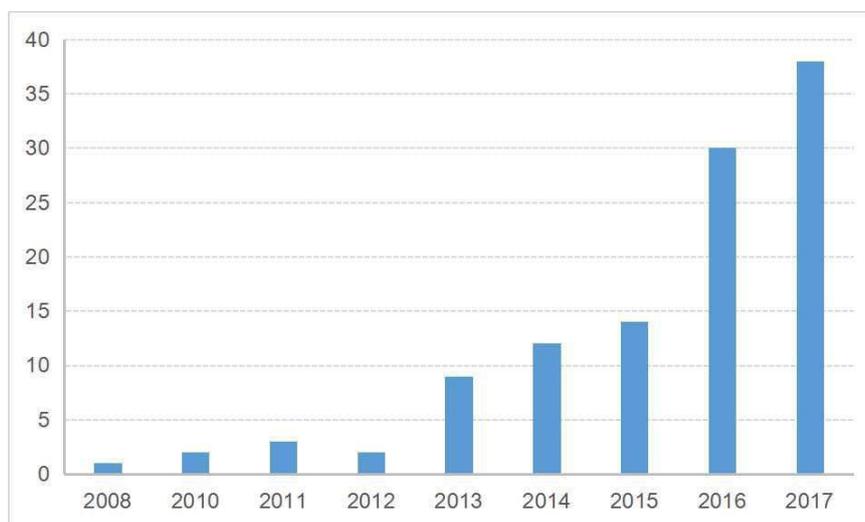


Fig. 1.6: Número de artículos de revisión sobre rPPG publicados por año.

Existe una reciente publicación ^[5] (octubre, 2019) sobre este estudio, la cual recoge detalladamente un sumario de los métodos actualizados de rPPG, junto a los pros y contras de cada uno de ellos bajo diferentes condiciones de análisis. Por lo tanto, el contenido posteriormente expuesto se apoya en gran medida sobre el texto de dicha publicación.

La medición de la FC mediante rPPG se basa en el mismo principio que la fotopletismografía tradicional (PPG), donde el pulso de la sangre propagada por el sistema cardiovascular provoca variaciones periódicas en el volumen sanguíneo del tejido microvascular bajo la piel.

El análisis rPPG mediante cámaras de bajo coste fue propuesto por primera vez en 2008 por Verkruysse *et al.* ^[6]. Ellos demostraron la posibilidad de medir la FC a partir videos grabados de rostros de sujetos bajo luz ambiental empleando cámaras digitales de bajo coste.

Más tarde, en 2010, Poh *et al.* ^[7] propuso la combinación lineal de los canales RGB para estimar la FC empleando métodos de separación ciega de fuentes o BSS (acrónimo del inglés *Blind Source Separation*).

Como alternativa, Sun *et al.* ^[8], en 2012, propuso un marco de medición remota de la FC en situaciones con luz ambiental mediante el análisis conjunto de tiempo-frecuencia. Desde entonces, se han realizado un número creciente de estudios basados en modelos ópticos realistas y técnicas avanzadas de procesamiento de señal, para medir de forma remota las señales PPG a partir de vídeos faciales ^{[9]-[13]}.

Este progreso ha sido recogido en varios artículos de revisión relevantes desde varios aspectos. Sun y Thakor ^[14] describieron las técnicas de medición de PPG de métodos con contacto a métodos sin contacto, y de punto a imagen. Al-Naji y Chahl ^[15] proporcionaron una amplia revisión de la literatura para el monitoreo cardiorrespiratorio remoto, incluyendo el efecto Doppler, tratamiento de imágenes térmicas y por video cámara.

Sikdat *et al.* ^[16] hizo una revisión metodológica para la estimación del pulso cardíaco guiada por visión sin contacto, actualizado al año 2014, momento en el que la mayoría de los estudios todavía se realizaban en un entorno relativamente estable.

Hassan *et al.* ^[17] investigó la estimación basada en cámaras digitales, tanto por rPPG como por balistocardiografía (BCG), mientras que el estudio más reciente de rPPG en la revisión se informó en el año 2015.

Más recientemente, el foco de investigación hasta ahora centrado en demostrar la viabilidad de medir la FC bajo condiciones de laboratorio bien controladas, se ha reorientado para contemplar condiciones más realistas y complejas (incluyendo variaciones de iluminación dinámica y artefactos en movimiento) ^{[18] [19]}.

Se han realizado un gran número de estudios para reducir o eliminar el impacto de los artefactos de ruido resultantes de los movimientos del sujeto, expresiones faciales, tono de piel y variaciones de iluminación ^{[20]-[24]}. Sin embargo, hasta hace muy poco no se había realizado una revisión exhaustiva sobre el desarrollo del rPPG que aborde esta clase de problemas más realistas sobre las variaciones dinámicas de iluminación y artefactos en movimiento.

Por esta razón, Chen *et al.* ha publicado recientemente (finales del 2019) un documento que proporciona una oportuna y sistemática revisión de los recientes avances en rPPG. Dicho documento es el mismo que sigue esta sección y en el cual se resume, compara y discute los avances metodológicos de rPPG en detalle, centrados en soluciones para los artefactos de iluminación y movimiento inducidos desde la perspectiva del procesado de señales.

1.3.1 Soluciones propuestas a las variaciones de iluminación

Trabajo Relacionado

Para suprimir la influencia de las variaciones de iluminación, una posible forma es emplear cámaras infrarrojas. Por ejemplo, Jeanne *et al.* [25] aprovechó las ventajas de las cámaras infrarrojas para estimar la FC en condiciones de luz altamente dinámica.

En cuanto a las soluciones para cámaras RGB, Xu *et al.* [26] propuso la extracción de las señales rPPG usando la ley de Lambert-Beer. Ellos probaron la viabilidad de estimar el pulso cardíaco bajo diferentes niveles de iluminación y notificaron una mejoría satisfactoria.

Al capturar videos faciales de sujetos en situaciones de iluminación variable, tanto la variación periódica en los niveles de reflectancia correspondientes a la información pulsátil y los cambios en la iluminación son grabados en las señales RGB de la cámara.

Chen *et al.* [27] [28] aplicaron un método tolerante a cambios en la iluminación basado en la descomposición empírica en modos por conjuntos o EEMD (acrónimo del inglés, *Ensemble Empirical Mode Decomposition*) al canal verde para separar las señales reales de FC del ruido provocado por la iluminación ambiental.

El procedimiento, usando EEMD seguido de un modelo de regresión lineal múltiple fue más tarde empleado en la evaluación de la FC para reducir el efecto de los cambios de iluminación ambiental [29].

Lam y Kuno [30] asumieron que la extracción de la FC a partir de subregiones de la cara podía ser tratada como un problema lineal de BSS. En ciertos casos, las variaciones de iluminación tanto en el rostro del individuo como en la región del fondo contienen patrones de variación similares. Varios métodos de medida de la FC emplean la región del fondo como referencia de ruido para rectificar las interferencias inducidas por las variaciones en la iluminación.

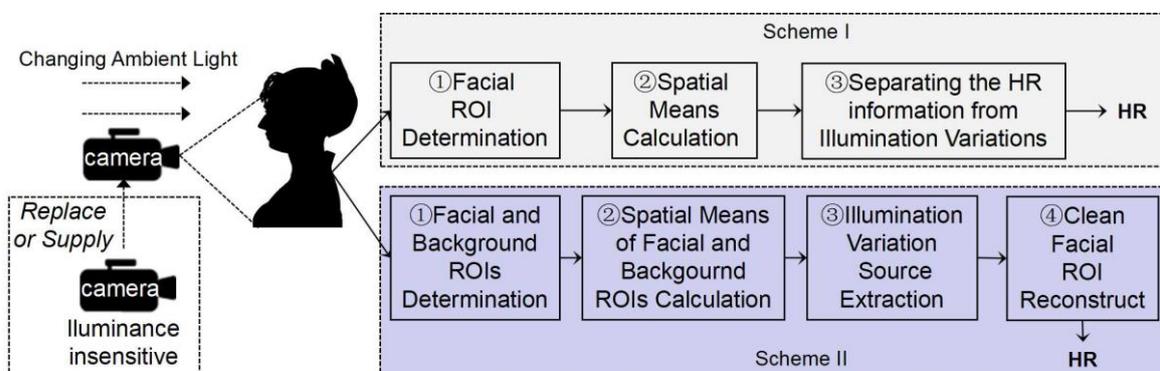


Fig. 1.7: Dos esquemas para abordar las variaciones de iluminación en la estimación de la FC con cámara RGB.

Li *et al.* ^[31] propuso un método de rectificación de iluminación basado en un filtro adaptativo de mínimos cuadrados medios normalizados o NLMS (*Normalized Least Mean Square*), con la suposición que tanto la región de interés (ROI) facial como el fondo eran modelos Lambertianos y compartían las mismas fuentes de luz.

Por lo tanto, el fondo puede tratarse como referencia para las variaciones de iluminación y podría ser filtrado en la ROI facial para rectificar la interferencia de las variaciones de iluminación cuando los sujetos ven los videos del test.

Lee *et al.* ^[32] también trataba el parpadeo provocado por la luz artificial mediante un modelado autorregresivo (AR). Los polos, asociados a las componentes frecuenciales del espectro en frecuencia derivado tras la aplicación del modelo AR sobre la región del fondo, se presentaban de igual manera en el modelo AR de la cara.

Gracias a esta similitud, estos polos podrían ser cancelados en la ROI de la cara para encontrar la componente asociada a la FC. Resultados experimentales mostraron que el modelado AR con la cancelación de polos era muy adecuado para los casos con iluminación fluorescente muy alta. Sin embargo, debido a que el modelado AR es un método basado en el análisis espectral, éste podría ser comprometido por las variaciones periódicas en la iluminación.

Recientemente, basándose en la misma suposición de que tanto la ROI facial como la ROI del fondo tienen los mismos patrones de variación de iluminación, Cheng *et al.* ^[33] han propuesto un marco de trabajo (*framework*) para el tratamiento robusto a la iluminación usando la separación conjunta de fuentes ciegas (JBSS), y la aplicación de la técnica EEMD (*Ensemble Empirical Mode Decomposition*) para la descomposición de la señal del pulso cardíaco en sus funciones de modo intrínseco (IMFs).

Este método propuesto demostró ser efectivo en una serie de situaciones con cambios dinámicos de iluminación. Además, Xu *et al.* ^[34] propuso un *framework* novedoso basado en mínimos cuadrados parciales (PLS) y EMD multivariante (MEMD) para evaluar eficazmente la FC a partir de las señales rPPG capturadas de los rostros en situaciones con iluminación variable.

La principal función del PLS es extraer las fuentes comunes de variación de iluminación subyacentes en las ROI facial y del fondo, mientras que el MEMD es capaz de extraer modos comunes a partir de varios canales de señal al considerar la información dependiente entre los canales RGB ^[34] ^[35].

Marco básico y Resumen

Las soluciones propuestas para manejar variaciones de iluminación presentes en la imagen del sujeto, al margen del empleo de cámaras insensibles a la luz tales como cámaras infrarrojas, se pueden agrupar principalmente en dos esquemas, mostrados en la fig. 1.7 bajo el supuesto de cámaras RGB convencionales,.

El primer esquema está basado en los métodos de procesamiento de señal para separar las señales asociadas a las variaciones de iluminación de las señales propias del pulso cardíaco.

Una solución tolerante a la iluminación es el algoritmo EEMD, del cual ya se ha demostrado su efectividad en la eliminación del ruido ^[36] ^[37]. Chen *et al.* ^[27] aplicó el algoritmo EEMD al canal verde para separar las señales reales del pulso cardíaco del ruido provocado por la iluminación ambiental. La mayoría de pasos están descritos en el primer esquema (Scheme I) en la fig. 1.7. Primero, se detecta y realiza un seguimiento de cara del individuo mediante técnicas de *face-tracking* para obtener la ROI facial con los píxeles de interés. Después, se calcula el promediado espacial de los tres canales RGB o únicamente del canal verde (G) de la ROI facial obtenida. En tercer lugar, la información de la FC se extrae bien mediante EEMD para derivar el IMF objetivo que representan las señales cardíacas o bien, aplicando BSS para seleccionar al azar regiones locales óptimas que contengan señales cardíacas. De este modo, la FC se puede estimar desde el IMF objetivo, o desde múltiples regiones locales combinadas con un esquema de votación por mayoría.

Sin embargo, el EEMD podría ser comprometido en casos con una variación periódica de la iluminación, especialmente si la frecuencia de ésta es muy próxima al rango de la FC (típicamente de 0.75Hz a 4Hz).

El segundo esquema está basado principalmente en asumir que las trazas originales de las señales rPPG (p. ej., región facial) contienen tanto las variaciones del volumen de sangre causadas por el pulso cardíaco como las variaciones de iluminación temporales. Tales variaciones de iluminación pueden ser consideradas como una referencia de ruido derivado de las regiones del fondo (sin el rostro) para eliminar el ruido de las señales rPPG extraídas de las regiones propias del rostro. Los procedimientos detallados se muestran en el Scheme II de la fig. 1.7.

En primer lugar, se determinan las ROI facial y del fondo mediante detección y seguimiento (*tracking*). En segundo lugar, se calculan los promedios espaciales de los canales de color (RGB) tanto de las regiones de la cara como del fondo. En tercer lugar, el fondo se trata como referencia del ruido para extraer la fuente de variación de iluminación mediante AR, el filtrado adaptativo NLMS, ajuste de curva de mínimos cuadrados, JBSS o PLS. Seguidamente, se resta la fuente de variación de iluminación al

ROI facial para reconstruir una nueva ROI facial libre de variaciones de iluminación. Y, por último, se mide la FC a partir de esta región facial depurada.

Por otra parte, un estudio ha demostrado que la selección aleatoria de patches ^[30] es mejor que los métodos basados en ICA ^[38] y NLMS ^[31]. El método binario JBSS-EEMD funciona mejor que los métodos basados en ICA ^[38], NLMS ^[31], MOCF (*Multi-Order Curve Fitting*) ^[32] y EEMD ^[27].

El rendimiento de estos métodos depende del grado de similitud en la variación de iluminación entre las regiones del rostro y del fondo. Chen *et al.* propuso varias medidas novedosas de similitud que pueden emplearse para un filtrado y regresión más robustos. Cabe señalar que cuando la variación del ROI facial es diferente a la de otras regiones distintas de la piel, los métodos del segundo esquema (*Scheme II*) pueden ser ineficaces.

1.3.2 Soluciones propuestas al movimiento

Un número creciente de estudios han sido realizados para suprimir el impacto del movimiento ^{[39]-[43]}. Como se puede ver en la fig. 1.8, los cambios de distancia y ángulo entre la cámara y la cara causados por el movimiento del sujeto, pueden describirse mediante un modelo óptico ^{[13] [24]}.

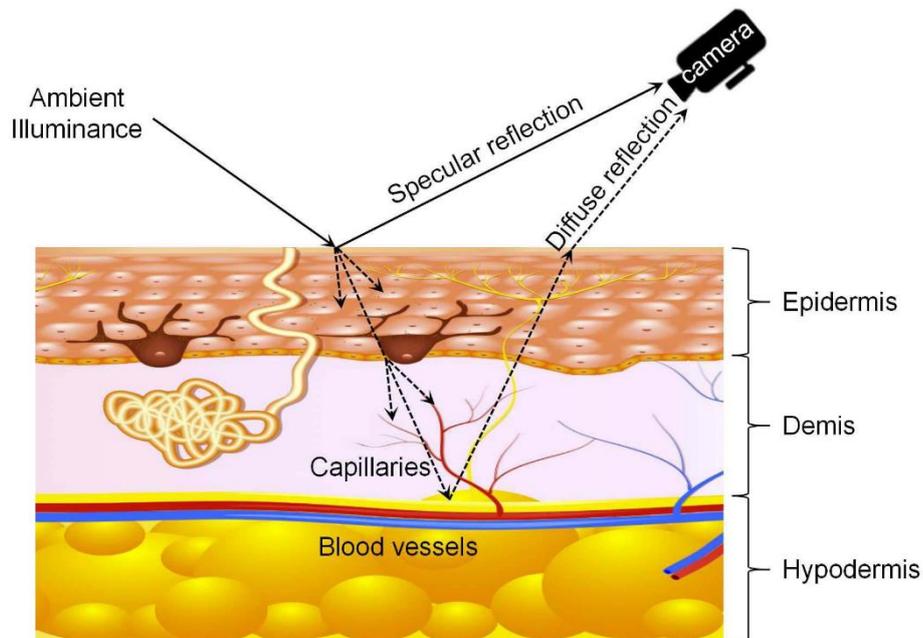


Fig. 1.8: Modelo de reflexión del rPPG.

Se observó que el ruido de cuantificación de la cámara podía reducirse mediante el promediado espacial de los valores RGB de todos los píxeles asociados a la piel de la ROI facial, e incluso llegar a ser insignificante [13].

El promediado temporal de los canales RGB de cada píxel $C(t)$ puede escribirse en una única ecuación (Ecuación 1) como combinación lineal de tres términos: el nivel de intensidad de iluminación $i(t)$, la parte variable de la reflexión especular $s(t)$ y las señales de pulso cardíaco $p(t)$. Además, estas tres señales son de media nula.

$$C(t) = I_0 \cdot \mathbf{u}_c \cdot c_0 + I_0 \cdot \mathbf{u}_c \cdot c_0 \cdot i(t) + I_0 \cdot \mathbf{u}_s \cdot s(t) + I_0 \cdot \mathbf{u}_p \cdot p(t) \quad [Ec. 2]$$

Respecto al resto de términos, I_0 se corresponde al nivel de intensidad lumínica estática (DC), $\mathbf{u}_c \cdot c_0$ es la parte variante temporal de la amplitud de intensidad, \mathbf{u}_s denota el vector unitario de color del espectro de luz y \mathbf{u}_p el vector unitario de amplitud de pulso.

Dependiendo del conocimiento previo disponible sobre las componentes, los métodos robustos al movimiento pueden agruparse principalmente en dos categorías, métodos basados en BSS o métodos basados en modelo.

Los métodos basados en BSS pueden ser ideales para desglosar $C(t)$ en fuentes para extraer el pulso sin información previa, mientras que los métodos basados en modelo pueden emplear el conocimiento sobre los vectores de color de las diferentes componentes para controlar este desglose.

Además de estas dos categorías, los métodos utilizados para determinar y trackear las ROIs pueden ser tratados como estrategias de compensación de movimiento. Estas tres categorías, al margen de otros métodos existentes, se agrupan en la fig. 1.9.

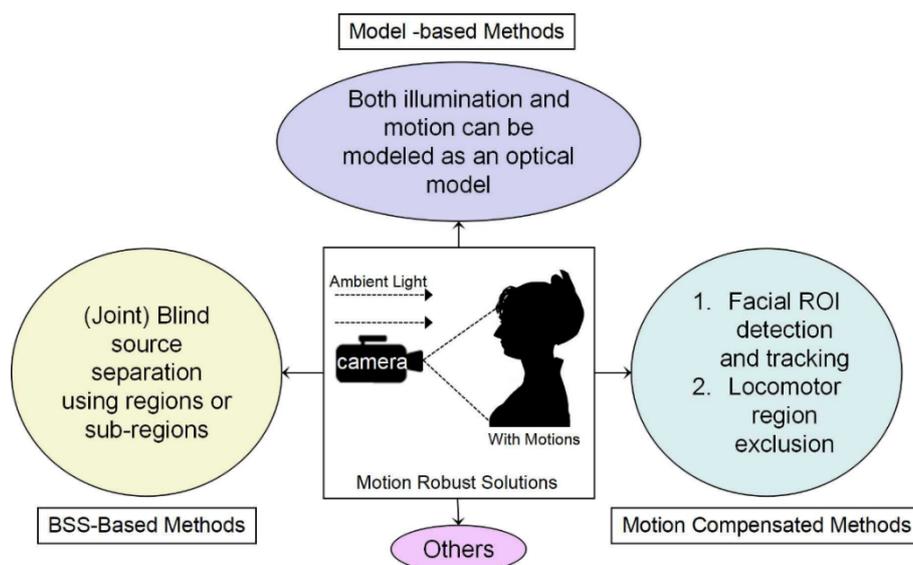


Fig. 1.9: Cuatro categorías de soluciones robustas al movimiento para técnicas rPPG.

A continuación, se describen las dos categorías fundamentales (métodos basados en BSS y en modelo) en el desarrollo de algoritmos robustos al movimiento.

Métodos basados en BSS

Dentro de esta categoría se pueden diferenciar la BSS convencional y BSS conjunta (JBSS).

➤ **BSS (Blind Signal/Source Separation) convencional**

La separación ciega de fuentes (BSS), consiste en recuperar señales o fuentes no observadas a partir de un conjunto de mezclas observadas, sin información previa respecto al proceso de mezcla.

Generalmente, estas observaciones se corresponden a salidas de los sensores siendo cada una de éstas una combinación de fuentes ^[45]. Un método común de BSS es el análisis de componentes independientes (ICA), cuya eficacia ha sido probada en muchos campos ^[46]. Basado en la suposición de que las señales de los canales RGB son ahora una combinación lineal de la señal del pulso cardíaco y otras señales, Poh *et al.* ^[47] propuso el ICA basado en el algoritmo JADE (del inglés, *Joint Approximate Diagonalization of Eigen-Matrix*).

El objetivo era suprimir las correlaciones y el alto orden de dependencia entre los canales RGB, para extraer la componente de FC mientras el sujeto está sentado y con movimiento natural. Tras esta aplicación, el error cuadrático medio (RMSE) correspondiente a las situaciones de movimiento se reducían de 19.36 a 4.63 bpm, demostrando la viabilidad del ICA para evaluar la FC.

Sun *et al.* ^[48] introdujo nuevos métodos para la reducción de artefactos que consistían en la compensación de movimiento plano y la BSS, referida principalmente al ICA monocal (SCICA). Su comportamiento en la extracción de la FC fue evaluado y probado a través de un testeó con video faciales capturados de un mismo individuo.

Monkaresi *et al.* propuso una aproximación combinando *machine learning* y este mismo ICA, para mejorar la precisión en la estimación de la FC en situaciones reales.

Wei *et al.* ^[49] propuso estimar la FC aplicando BSS de segundo orden a las señales RGB, en este caso de 6 canales, provenientes de la doble ROI facial. Los métodos BSS tenían de algún modo la capacidad de tolerar movimientos, pero aún de forma limitada, especialmente al tratar con movimientos severos ^[50].

Dado que el orden de las componentes extraídas vía BSS es aleatorio, generalmente se emplea la FFT para determinar la FC más probable. Así pues, los métodos BSS no se pueden emplear en casos en los que la frecuencia de la periodicidad de los artefactos en movimiento coincide con el rango de frecuencias del pulso cardíaco.

Recientemente, Al-Naji *et al.* [51] han propuesto la combinación de EEMD completa con ruido adaptativo (CEEMDAN) y análisis de correlación canónica (CCA) para estimar la FC a partir de secuencias de video grabadas por un dron.

La combinación de los métodos CEEMDAN y CCA logra un mejor rendimiento que los métodos ICA o PCA en presencia de ruidos inducidos por la variación de iluminación y los movimientos del sujeto y de la cámara.

➤ **JBSS (Joint Blind Signal/Source Separation)**

Las técnicas de la BSS convencional son originalmente diseñadas para tratar un solo conjunto de datos a la vez, p. ej., descomponiendo las distintas señales de los canales RGB de una única ROI facial en sus respectivas componentes independientes [21].

Recientemente, los canales RGB de múltiples subregiones faciales han sido empleadas para una mayor precisión en la medida de la FC [20] [52]. Con el incremento de la disponibilidad de múltiples conjuntos de datos (*multisets*) para el análisis, varios métodos JBSS han sido propuestos para acomodar simultáneamente varios *multisets*. La meta del JBSS es extraer las fuentes subyacentes de cada conjunto de datos y mientras tanto, mantener un orden consistente de las fuentes extraídas a través de múltiples conjuntos de datos [46].

Chen *et al.* proporcionó una completa descripción de los métodos más representativos de JBSS. Estos métodos, tras ser testeados en varias aplicaciones neurofisiológicas realistas, han destacado por su eficacia y grata perspectiva a futuro en el análisis de datos neurofisiológicos.

Guo *et al.* [53] introdujo el método JBSS en el rPPG aplicando el análisis de vectores independientes (IVA) para analizar conjuntamente las señales de color derivadas de las múltiples subregiones faciales. A priori los resultados experimentales mostraron una mayor precisión en la medida de la FC en comparación con el método ICA-BSS.

Más tarde, Qi *et al.* [54] propuso un nuevo método sin contacto para la medida de la FC mediante la exploración de correlaciones entre las subregiones faciales del set de

datos con JBSS. Las pruebas en una gran base de datos pública también demostraron que el método JBSS propuesto superó a las metodologías anteriores basadas en ICA.

Métodos basados en Modelo

Desde que la información de los vectores de color puede ser utilizada por los métodos basados en modelo para controlar la separación de componentes, estos métodos tienen en común que la dependencia del promedio temporal de las señales RGB con la reflexión media de la piel puede ser eliminada ^[13].

Los métodos basados en modelo son típicamente métodos basados en modelos de crominancia (CHROM), métodos que utilizan la firma del volumen de sangre (PBV) para distinguir la señal del pulso de las distorsiones por movimiento ^[55], y métodos basados en un plano ortogonal a la piel (POS) ^[13].

de Haan y Jeanne ^[44] desarrollaron un CHROM que consideraba las componentes de reflexión difusa y las contribuciones de reflexión especular, que en conjunto hicieron que el color observado variara dependiendo de la distancia (ángulo) desde (entre) la cámara a la piel y a las fuentes de luz. Por tanto, el impacto de tales artefactos de movimiento podría eliminarse mediante una combinación lineal de los canales R, G y B individualmente. Resultados experimentales demostraron que CHROM superó a los métodos predecesores basados en ICA y PCA en presencia de movimiento.

Basándose en el mismo método CHROM, Huang *et al.* ^[56] aplicó un filtro adaptativo (tomando la posición de la cara como referencia) seguida de la transformada discreta de Fourier (DFT) para señales rPPG. Los resultados del experimento mostraron que la robustez al movimiento del método propuesto era viable incluso cuando los sujetos realizaban ejercicios periódicos en máquinas deportivas.

Wang *et al.* ^[90], utilizando CHROM como base, propuso un marco novedoso para suprimir el impacto de los artefactos del movimiento, aprovechando la redundancia espacial de los sensores de imagen para distinguir el pulso cardíaco del ruido inducido por el movimiento.

Posteriormente, de Haan y van Leest ^[55] propusieron un método basado en PBV para mejorar la robustez al movimiento. Este método utilizaba la firma de los cambios en el volumen de sangre para distinguir cambios de color provocados por el pulso en las trazas del RGB a lo largo del video. Los resultados experimentales obtenidos de los sujetos mientras realizaban ejercicio en 5 máquinas deportivas diferentes, mostraron una mejora significativa respecto al método CHROM.

Recientemente, Wang *et al.* ^[13] ha propuesto otro algoritmo rPPG basado en modelo, referido al POS. El método POS define un tono POS en el espacio RGB temporalmente normalizado, para la extracción del pulso cardíaco. De todos estos métodos comentados, tras ser testeados en una base de datos privada el POS obtuvo el mejor rendimiento en general, principalmente debido a que el tono POS definido era fisiológicamente razonable. Esto hizo que el POS fuera especialmente ventajoso en situaciones de ejercicio, donde la máscara de la piel era ruidosa. También se demostró que el POS y CHROM funcionaban bien bajo situaciones estacionarias y de movimiento, aunque pueden tener problemas en distinguir la componente pulsátil de distorsiones cercanas en nivel de amplitud. Sin embargo, PBV fue especialmente diseñado para situaciones de movimiento.

Apartado 2

Materiales y Métodos

Esta sección recopila todas las etapas fundamentales en las cuales se han segmentado los análisis de rPPG para la estimación de la frecuencia cardíaca (FC) en los modos offline y online.

Principal e independientemente del modo en cuestión, existen tres fases en las que se asienta esta técnica: recopilación de datos, análisis y procesado de señal y, por último, cálculo de resultados.

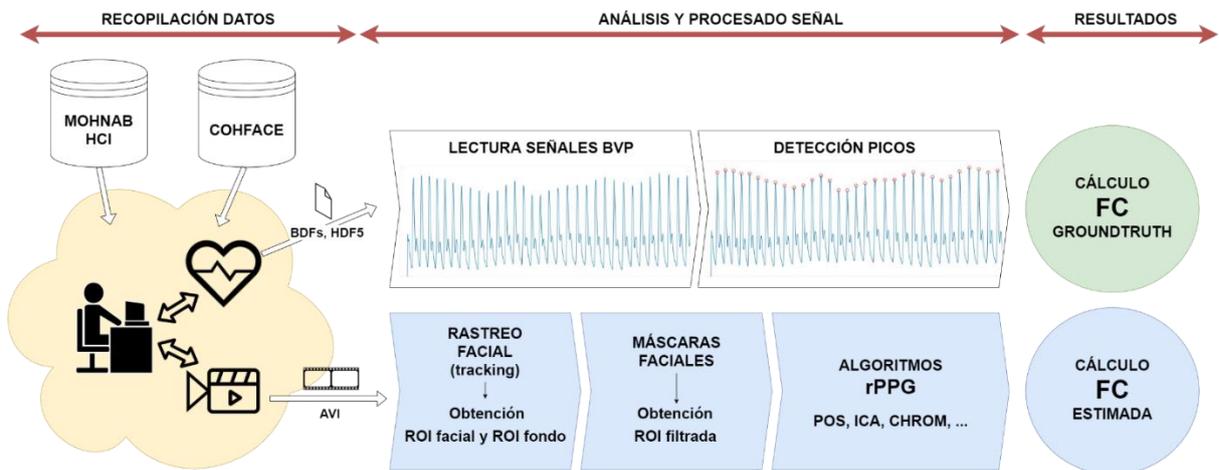


Fig. 2.1: Diagrama de fases en rPPG, modo offline.

En el modo offline, y como se puede intuir del diagrama de la fig. 2.1, los datos a procesar son los videos de los individuos y los archivos que contienen la información capturada por los sensores fisiológicos. Estos conjuntos de datos están recopilados en dos bases de datos independientes.

Por otro lado, y a diferencia de este primer modo, en el modo online la fase de recopilación de datos va asociada únicamente a la secuencia de video generada por la webcam en tiempo real. En esta vertiente, los datos adquiridos se actualizan constantemente durante el tiempo de grabación y no son almacenados ni exportados para su disposición futura.

Respecto al resto de fases, tanto el análisis y procesado de señal como el cálculo de resultados son comunes en ambos modos. Si bien, en el modo online existe una

ejecución iterativa de las tres fases, en el modo offline esta ejecución se realiza una vez por cada video.

A continuación, en la fig. 2.2, se muestra el diagrama de fases del modo online. A la vista de la fig. 2.1 y fig. 2.2, se aprecia la similitud de ambos diagramas intrínseca al método rPPG.

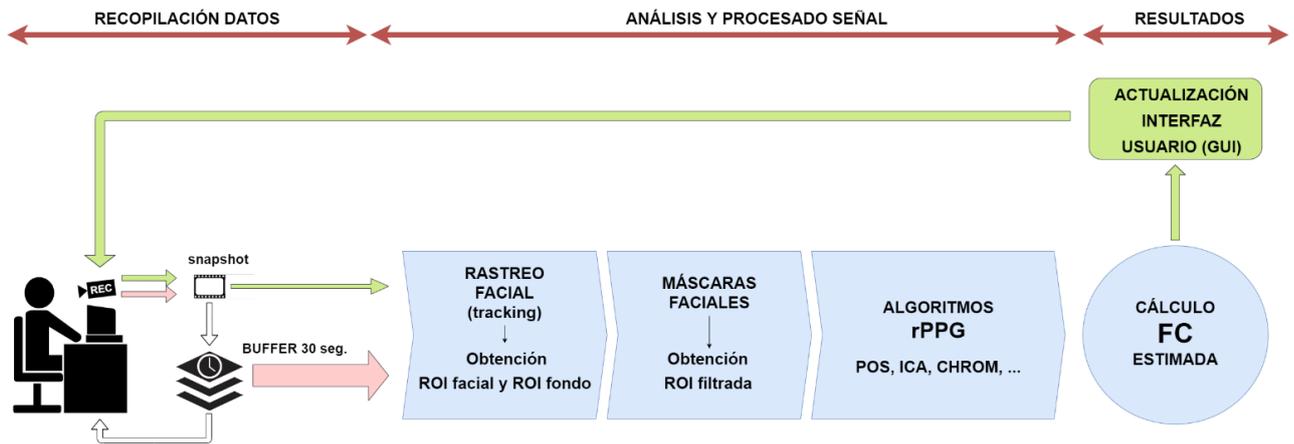


Fig. 2.2: Diagrama de fases en rPPG, modo online.

2.1 Bases de datos

A lo largo del estudio de las técnicas de rPPG existentes sobre rostros de individuos para la estimación de su frecuencia cardíaca, se deben tener en cuenta numerosos condicionantes que afectan intrínsecamente a la precisión de los resultados obtenidos.

Debido a que el análisis se fundamenta en el procesado de imágenes (fotogramas), estos condicionantes están asociados directamente a cómo se dispone la información, tanto de señal útil como espuria, en la propia imagen. En este caso, los principales factores que intervienen en la imagen son el sujeto (tono de piel, vello facial, maquillaje, etc) y la iluminación presente en el escenario en el que se encuentra (si es luz natural o artificial, tipos de fuentes, etc).

Además, para poder comparar fielmente los resultados obtenidos mediante el método rPPG y los reales es imprescindible disponer, al margen de los videos de sujetos, de la señal del pulso cardíaco de estos durante el periodo de grabación del video.

Para cubrir este aspecto y recopilar los resultados de análisis bajo distintas casuísticas, se han empleado los videos recogidos en dos bases de datos públicas: *MAHNOB HCI* ^[1] y *COHFACE* ^[2].

Ambas bases de datos contienen tanto los vídeos de los sujetos como los ficheros generados por los sensores de medición de frecuencia cardíaca (colocados a cada sujeto en los test) con la información del ECG correspondiente a cada uno de ellos.

Por otro lado, cabe destacar que actualmente la mayoría de los trabajos publicados sobre rPPG son evaluados sobre bases de datos privadas, lo que dificulta considerablemente la evaluación de los algoritmos propuestos de una manera estandarizada y basada en unos principios comunes preestablecidos.

Como consecuencia, en este trabajo se emplean dos bases de datos de carácter público y se analizan con los algoritmos rPPG más relevantes, de manera que los resultados obtenidos brinden una visión más realista sobre su comportamiento bajo distintos escenarios.

2.1.1 MAHNOB HCI

MAHNOB HCI-Tagging dataset ^[1] se trata de una base de datos etiquetada multimodal, creada para la caracterización de contenido multimedia basada en emociones humanas. Así pues, se trata de una base de datos afectiva.

Sus autores recopilaron los videos y las señales fisiológicas de 30 sujetos (13 hombres y 17 mujeres de entre 19 y 40 años, $M = 26.06$ años y $SD = 4.39$ años) de diferentes orígenes culturales, siendo expuestos a estímulos audiovisuales controlados mediante la visualización de imágenes y pequeñas secuencias de vídeo en dos experimentos.

En el primer experimento, los individuos visualizaban 20 videos emocionales anotando tanto su estado anímico como su nivel de excitación en una escala de valencia. En el segundo experimento, se les mostraban pequeñas secuencias de vídeo e imágenes sin etiquetar, y después etiquetadas correcta o incorrectamente. Y posteriormente, cada individuo evaluaba si estaba de acuerdo o en desacuerdo con las etiquetas propuestas.

Según los autores, durante la grabación se empleó equipo profesional en cuanto a iluminación y videocámaras, para maximizar la calidad de los datos de video. En la Fig. 2.1, se muestra un ejemplo de una captura de vídeo extraída por la cámara durante el test.



Fig. 2.3: Ejemplo de imagen capturada en MAHNOB.

A pesar de no ser su objetivo original, esta base de datos puede ser empleada para evaluar las técnicas rPPG, puesto que dispone de toda la información necesaria (sensores de ritmo cardíaco y los vídeos), además de información extra como el contenido multimedia visualizado por los individuos durante la prueba, los ficheros generados por el resto de los sensores biométricos, señales de audio, etc.

Si se ignora la mayor parte de su contenido escogiendo las señales fisiológicas de los sensores más relevantes y los datos de vídeo, sólo faltaría correlacionar ambas informaciones mediante software para que vayan perfectamente sincronizadas. Con esto, la obtención del pulso cardíaco para cualquier instante de vídeo es inmediata.

Captura de datos

Durante los experimentos, el comportamiento de los participantes fue grabado mediante 6 cámaras profesionales desde ángulos diferentes, 2 micrófonos, un rastreador de mirada y los sensores fisiológicos para la extracción de los ECG (32 canales), la amplitud de respiración y la temperatura de la piel. En la Fig. 2.2, se muestra un ejemplo de dicha configuración experimental.



Fig. 2.4: Ejemplo de la configuración experimental para la grabación de los vídeos en MAHNOB HCI.

El audio está grabado en dos canales. El canal 1 (o izquierdo, si se interpreta como señal estéreo) contiene la señal de audio del micrófono de sala *AKG C 1000 S MkIII*, que incluye tanto el ruido de la habitación como los sonidos de los estímulos del video. El canal 2 (derecho) contiene la señal de audio del micrófono de diadema *AKG HC 577 L* dispuesto en la cabeza del participante.

Los vídeos de cada sesión fueron tomados por 6 cámaras de video diferentes. La primera de ellas estaba situada encima de la pantalla y capturaba la señal de video a color, cubriendo íntegramente el frontal del individuo. El resto de cámaras, monocromáticas, capturaban las reacciones del individuo desde distintos ángulos, incluyendo una vista frontal, vista de perfil y vista superior del cuerpo.

Para esto, se emplearon dos tipos de cámaras: el modelo *Stingray F-046C*, para la señal a color y las otras cinco de modelo *Stingray F-046B*. Todas ellas del fabricante *Allied Vision* con una resolución de 780 x 580 píxeles y 60 frames por segundo (fps).

Antes de cada experimento se calibraba el rastreador de mirada mediante el software de estudio *Tobii* ^[58] y, todos los sensores fisiológicos incluidos los sensores de encefalograma (EEG), sensores de electrocardiograma (ECG), sensores de respuesta galvánica de la piel (GSR) de los dedos, sensor de temperatura de la piel y la cinta de respiración alrededor del pecho se colocaban adecuadamente sobre los participantes.

Las respuestas fisiológicas de los participantes se registraron utilizando el sistema *Biosemi Active II* ^[59], modelo utilizado por muchos laboratorios de investigación en todo el mundo. El sistema se conecta al resto de dispositivos (sensores) mediante fibra óptica (aislamiento galvánico) y va alimentado mediante una batería. Las señales provenientes de estos sensores están almacenadas en los archivos BDF (*Biosemi Data Format*) disponibles en la base de datos. Existe uno por cada sesión.

Formatos de archivo

Como se ha indicado previamente, el formato de salida del sistema *Biosemi Active II* es un fichero BDF único por sesión que contiene toda la información relativa a los estímulos de los participantes y las señales fisiológicas.

Los datos se distribuyen en 47 canales. Los 32 primeros se corresponden a la actividad cerebral recogida por los electrodos EEG, dispuestos en un gorro. Los canales 33, 34 y 35 contienen la información de las señales de pulso cardíaco recogidas por tres sensores ECG situados en el tórax.

Los canales 45 y 46 se corresponden a las señales recogidas por la cinta de respiración y el sensor de temperatura respectivamente. El canal 47 se asocia a una señal formada por una serie de pulsos que indican el estado del experimento, delimitando el inicio y final de los estímulos grabados dentro de la secuencia de vídeo. El objeto de este último canal es facilitar la sincronización entre las señales capturadas por los sensores y la secuencia de vídeo.

En cuanto al formato de vídeo, todas las secuencias se transmiten en un contenedor de películas AVI.

Preselección de sesiones

Para realizar los análisis rPPG en este trabajo, se ha realizado una selección previa de sesiones descartando aquellas que pudieran causar imprecisiones en la estimación de la FC por los algoritmos rPPG.

Así pues, de las 347 sesiones que contiene la base de datos *MOHFACE HCI* se han escogido 286 bajo el criterio de selección guiado por dos requisitos: que la duración del vídeo suficientemente larga (mínimo 40 segundos) para poder estimar el ritmo cardíaco y, por otro lado, que los individuos (de cualquier género y con o sin consentimiento) tengan el rostro lo más despejado posible evitando aquellos con barba, bigote o gafas.

2.1.2 COHFACE

COHFACE dataset [2] se trata de una base de datos publicada en 2017, que recoge un gran número de sujetos grabados bajo dos condiciones diferentes de iluminación. Concretamente, contiene 160 videos y señales fisiológicas extraídas de 40 individuos en buenas condiciones de salud.

Los datos adquiridos en este nuevo corpus incluyen condiciones más realistas en comparación con las grabaciones bajo condiciones de laboratorio bien controladas del set *MAHNOB HCI*.

En la fig. 2.5, se muestra un ejemplo de las señales ECG capturadas por los sensores, disponibles en la base de datos. En azul, las señales ECG de la base de datos. En rojo, el resultado de la detección QRS. En la parte superior de cada gráfico, la frecuencia cardíaca estimada en bpm.

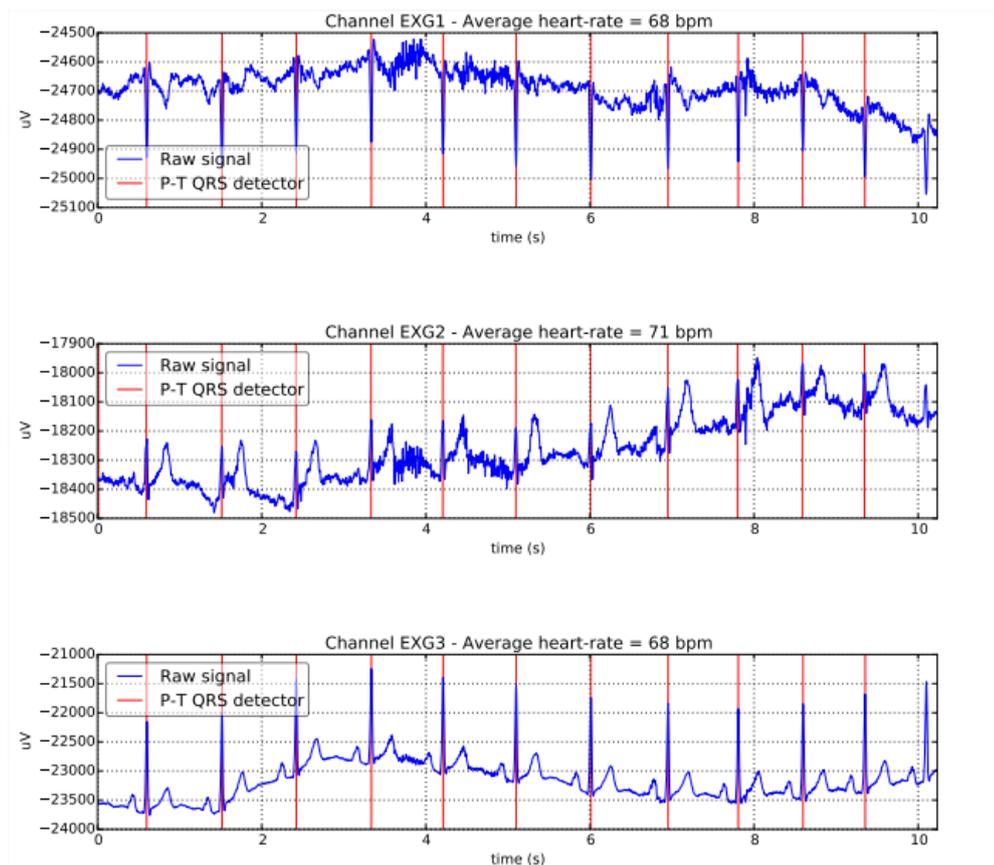


Fig. 2.5: Ejemplo de las señales ECG extraídas.

La edad promedio de los individuos es de 35.6 años, con una desviación estándar de 11.47 años. En cuanto al género, hay 12 mujeres (30%) y 28 hombres (70%). La distribución de edad de los 40 individuos se muestra en fig. 2.6.

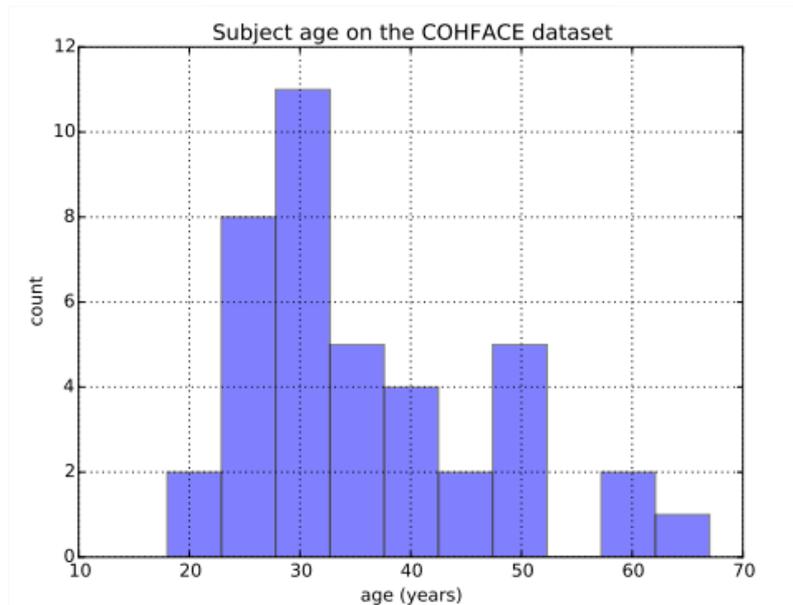


Fig. 2.6: Distribución de edad de los sujetos recopilados en COHFACE.

Captura de datos

A los individuos participantes se les grabó mediante una webcam convencional conectada a un portátil durante un periodo de aproximadamente 60 segundos. Se registraron tanto las señales de vídeo como las fisiológicas, estas últimas mediante fotopletismografía de contacto.

El equipo empleado para ello consistía en un kit de herramientas biomédicas con 3 sensores (canales de entrada) sincronizados, permitiendo medir tanto el ECG como la respiración del individuo.

Los datos fisiológicos serán utilizados como valores reales de referencia, también conocidos como valores *groundtruth* (término del inglés que significa “base verdadera o verdad fundamental”), durante las pruebas de los algoritmos rPPG.

En cada uno de los 160 videos, la cara de un sujeto es grabada por una webcam convencional (Logitech HD Webcam C525) durante un minuto completo, mientras se capturan las señales fisiológicas mediante un sensor BVP (modelo SA9308M) y una cinta de respiración (modelo SA9311M), ambos equipos del grupo Thought Technologies.

El sensor BVP mide los cambios en la reflectancia de la piel tras aplicarle luz infrarroja, causados por la variación del nivel de oxígeno en sangre debido a los latidos del corazón.

La cinta respiratoria está compuesta por un sistema de bobinado mecánico (sistema inductivo) que simplemente mide el estrechamiento torácico. Ambos sensores están

conectados (siguiendo los consejos del fabricante) a un ordenador con SO Microsoft Windows a través de su sistema de adquisición basado en USB de dos canales (ProComp2).

Los datos de los sensores fisiológicos y el vídeo se sincronizan y graban conjuntamente mediante la suite de programas Thought Technologies BioGraph Infiniti, versión 5. Con la configuración estándar, BioGraph Infiniti v5 es capaz de transmitir el video a una resolución de 640 x 480 píxeles a 20 frames por segundo. Junto a ello, y de forma paralela, realiza la medición de las señales BVP a 256 muestras por segundo y la lectura de respiración a 32 muestras por segundo.

La siguiente figura, fig. 2.7, muestra las señales adquiridas en sincronización con los datos de video, ambas en azul y correspondientes al pulso de sujeto (primera gráfica) y la lectura de respiración (segunda gráfica).

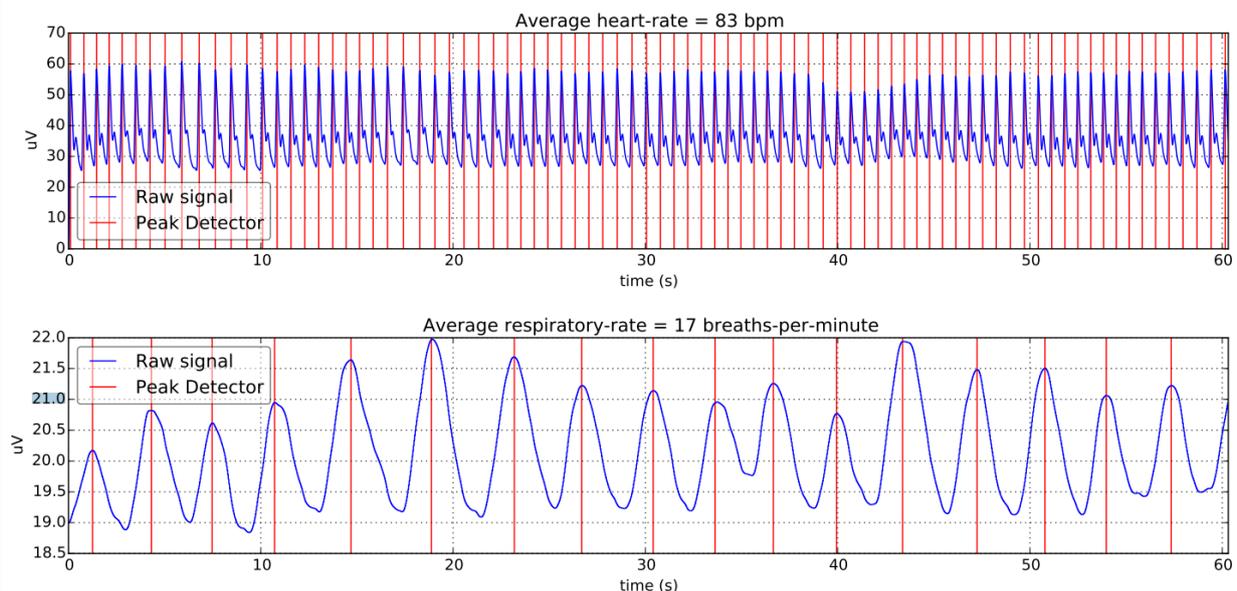


Fig. 2.7: Ejemplo de las señales adquiridas en cada sesión del set COHFACE.

Protocolo de adquisición y cambios de iluminación

Cada uno de los 40 individuos voluntarios se sentaron en frente de una webcam de manera que toda la región de la cara quedase íntegramente grabada, durante cuatro sesiones de aproximadamente un minuto de duración. A su vez, estas grabaciones se replicaron bajo dos condiciones diferentes de iluminación:

- Con luz de **estudio**: para la cual se cerraron las persianas de la sala, evitando la influencia de luz natural y aplicando una fuente de luz adicional en un punto para iluminar bien la cara del sujeto.
- Con luz **natural**: en el que todas las luces estaban apagadas y las persianas abiertas.

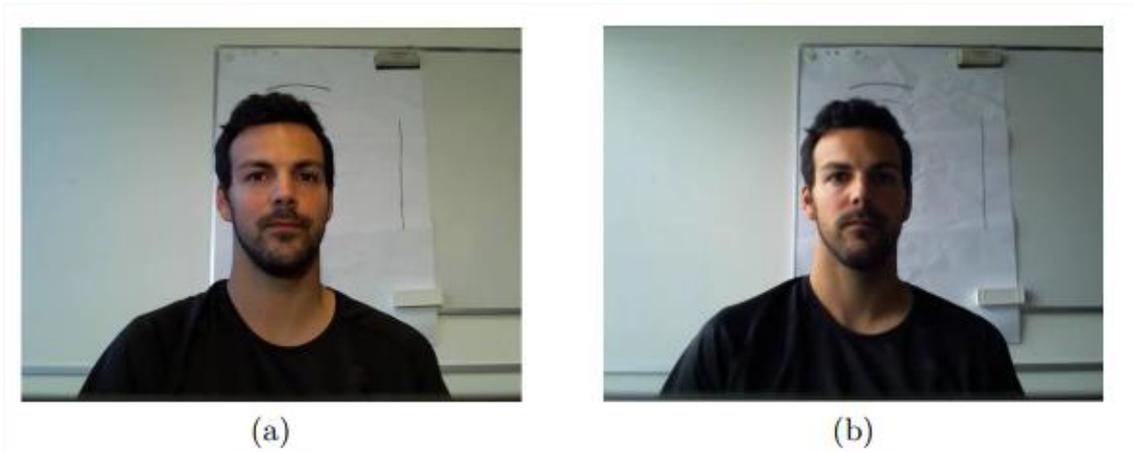


Fig. 2.8: Ejemplo de imágenes capturadas en COHFACE. (a) luz estudio, (b) luz natural.

Las cuatro secuencias de video (2 con luz de estudio y 2 con luz natural) pueden ser empleadas para evaluar el rendimiento de los algoritmos rPPG, bajo las mismas condiciones o completamente diferentes.

Formatos de archivo y metadatos

La elección de los formatos para recopilar los datos ha sido seleccionada para mejorar la legibilidad y el acceso mediante herramientas de software gratuitas. Los datos de video en cada sesión y por cada individuo, se envía en formato MP4 dentro de un contenedor de películas AVI. Los datos de los sensores fisiológicos se distribuyen utilizando contenedores HDF5 estándar, con información como la fecha de nacimiento del sujeto, su género, la condición de iluminación y por supuesto, las señales del pulso y respiración registradas.

2.2 Herramientas y algoritmos

2.2.1 Obtención de la FC *groundtruth* (modo offline)

Para poder cuantificar la precisión de los algoritmos rPPG empleados en la estimación de la FC de los vídeos en el modo offline, es necesario conocer la FC *groundtruth* obtenida a partir de los sensores fisiológicos capturados durante la prueba de cada individuo, y así comparar posteriormente el grado de correlación entre los resultados estimados y los *groundtruth*.

Los datos de los sensores fisiológicos se facilitan junto a los videos en las dos bases de datos analizadas. La lectura de estos datos se realiza a partir de los archivos BDF en el caso de la *MAHNOB HCI* y de los archivos HDF5 en el caso de la *COHFACE*.

Cálculo de la FC *groundtruth* en MAHNOB HCI

Los datos de los sensores, en el caso de la *MAHNOB HCI*, necesitan ser sincronizados previamente con los datos de video ya que en esta base de datos se incluye los treinta segundos antes y después del vídeo con el participante en reposo, sin estímulos. Para esto, se emplea el canal 47 que indica el estado del experimento con flancos de subida. Una vez delimitado el tiempo de interés (TOI), se selecciona los datos de ECG en ese intervalo.

En ese intervalo, se detectan los picos del complejo QRS con un detector de picos (algoritmo *Pan-Tompkins*) y mediante la diferencia temporal iterativa de un pico y su vecino, se calculan los índices asociados a cada pico de los latidos del ECG.

Una vez obtenidos los índices, se calcula el promedio entre ellos obteniendo el intervalo temporal en muestras, N . Con este dato, conocida la frecuencia de muestreo a la que han sido capturados los datos de los sensores, fs_ecg , y la frecuencia de muestreo de los vídeos, fs_video , el cálculo de la frecuencia cardíaca *groundtruth* es inmediato sustituyendo en la siguiente fórmula:

$$FC\ groundtruth\ (BPM) = \frac{fs_ecg \cdot fs_video}{N} \quad [Ec. 3]$$

Algoritmo Pan-Tompkins

Es un algoritmo diseñado principalmente para detectar y localizar complejos QRS en señales de electrocardiograma (ECG). La detección del QRS se basa en análisis de amplitud, pendiente y anchura de los complejos QRS.

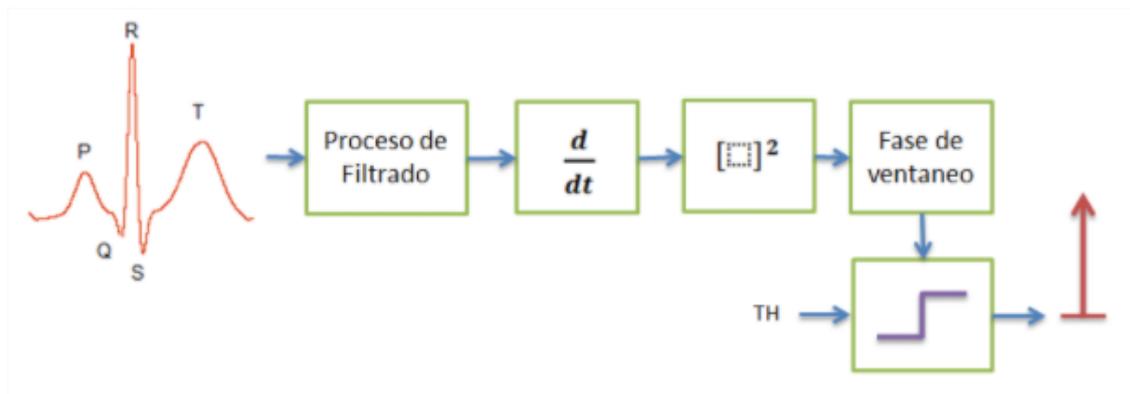


Fig. 2.9: Etapas del detector de picos Pan-Tompkins

Las etapas que describen este algoritmo son fundamentalmente cuatro: filtrado paso-banda, derivación, transformación no lineal y decisión. El filtrado paso-banda está compuesto por un filtro paso-alto y paso-bajo en cascada, y tiene como objetivo atenuar las bajas frecuencias asociadas a las ondas P y T, oscilaciones basales e interferencia de red. La etapa de derivación destaca las pendientes abruptas de la señal, que en el caso del ECG coincide con el complejo QRS. En la etapa de transformación no lineal se eleva al cuadrado la señal convirtiéndola en positiva y, se acentúan las frecuencias altas correspondientes al QRS. Y, por último, en la etapa de decisión se umbraliza la señal seleccionando únicamente los picos de interés (picos R) [60] [61].

Este algoritmo es óptimo para la selección de picos en los *bdf* facilitados en la base de datos *MAHNOB HCI*.

Cálculo de la FC *groundtruth* en COHFACE

Para la obtención de la FC *groundtruth* en la COHFACE se realizan tres medidas:

- **Análisis espectral:** se selecciona la componente principal evitando las componentes de baja frecuencia. Esta medida es única para cada vídeo. Para calcular la FC bastará con multiplicar por 60 la frecuencia (en Hz) de la componente principal.
- **Análisis en tiempo real:** se realiza detectando los picos en la señal temporal del ECG, calculando la diferencia entre vecinos y estimando a partir de ellos la frecuencia en tiempo real del sujeto.
- **Análisis temporal promedio:** se calcula a partir del análisis en tiempo real, promediando el conjunto de estimaciones y filtrando los *outliers*, por si hubiera alguna estimación errónea.

La diferencia entre la primera y tercera medida es pequeña (menos de 10 bpm para toda la base de datos como máximo, y menos de 1 bpm de media). Con esto, se escoge la tercera medida por ser la más precisa.

La detección de picos se realiza mediante una herramienta propia, ya que el algoritmo de detección de picos Pan-Tompkins empleado con la MAHNOB HCI no funciona correctamente en este caso. Esto es debido a que las lecturas del pulso cardíaco de los sensores PPG con contacto empleados en la COHFACE son más simples y de menor sensibilidad, como se puede apreciar en Fig. 2.7.

En Fig. 2.10, se muestra un ejemplo tras su aplicación en un fichero generado por los sensores. En color azul, la señal BVP leída en una sesión. En color rojo, los picos detectados del complejo QRS. El valor *groundtruth* de frecuencia cardíaca, mostrado en el título se obtiene de manera inmediata tras la detección de picos y su localización temporal.

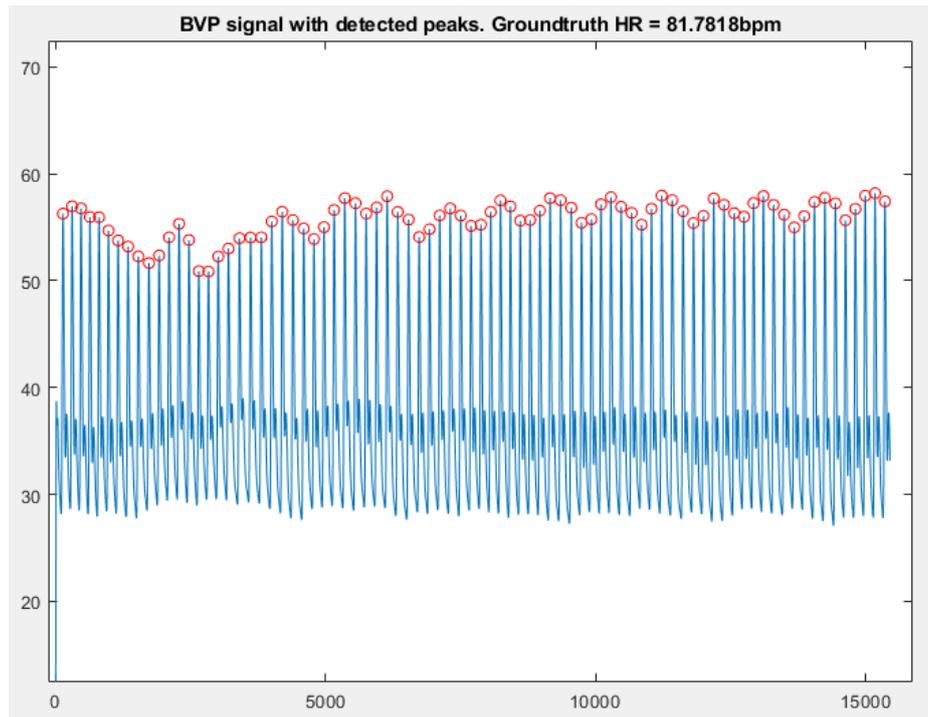


Fig. 2.10: Ejemplo detección de picos en señal BVP de COHFACE.

2.2.2 Obtención y tratamiento de la ROI facial

Algoritmo iCCR-master

El punto de partida de los algoritmos rPPG es una imagen del rostro del individuo en cuestión, por lo que una parte crítica de este trabajo es lograr una buena segmentación de la región de interés (ROI), en nuestro caso, la ROI del rostro del individuo para posteriormente ser tratado con la mayor pulcritud posible.

Por ello, se ha decidido emplear el algoritmo iCCR-master para la detección y seguimiento de rostros (*face-tracking*) ya que destaca por su precisión y rapidez en aplicaciones a tiempo real. Este algoritmo realiza un seguimiento de marcas faciales, basado en la regresión continua en cascada (CCR) con actualización incremental, lo que le permite agilizar el proceso de aprendizaje para la regresión en cascada disminuyendo el tiempo de actualización a una fracción de segundo. Esto le permite un seguimiento del rostro en tiempo real.

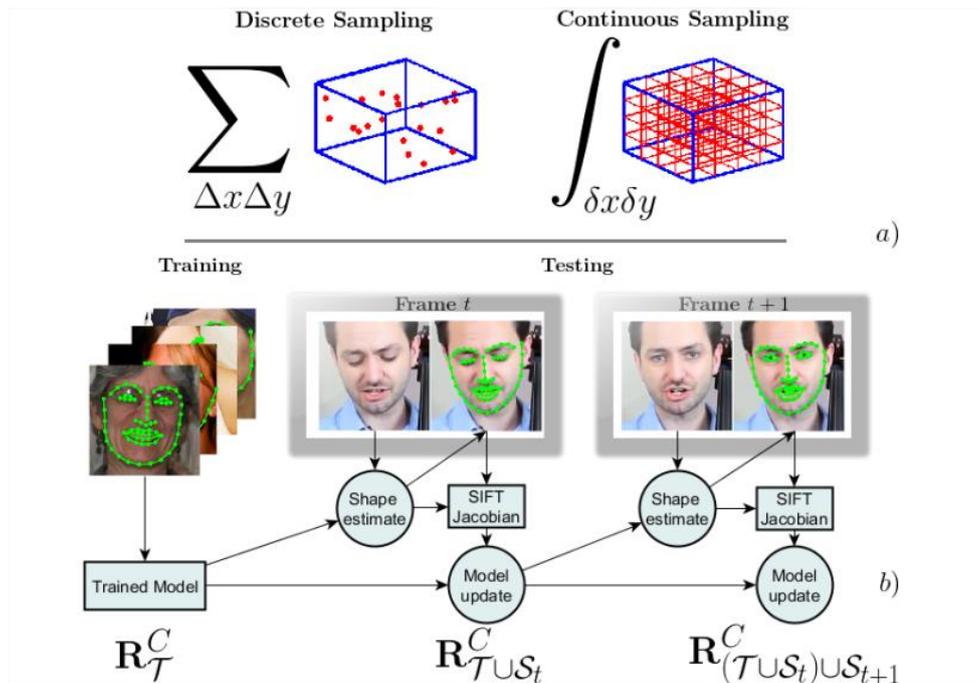


Fig. 2.11: Visión general del algoritmo iCCR.

El enfoque estándar del seguimiento facial es emplear un algoritmo de detección de marcas faciales inicializado con las marcas detectadas en el *frame* anterior. Esto se debe a que la forma de la cara varía de manera continua con una tasa de *frames* lo suficientemente alta y, por lo tanto, si las marcas iniciales son detectadas con suficiente precisión, la forma inicial del modelo será lo suficientemente cercana para que el algoritmo converja a un óptimo local aceptable también para el *frame* actual. Por lo tanto, es más probable que los algoritmos de seguimiento produzcan resultados de ajuste altamente precisos que los algoritmos de detección que se inician mediante el cuadro delimitador del detector facial.

Sin embargo, en este algoritmo, el *tracker* emplea un modelo deformable genérico de la cara, creado mediante entrenamiento previo utilizando una base de datos con imágenes faciales de otros individuos y marcadas manualmente. Por ello, la mayor limitación de este algoritmo es que el modelo genérico de referencia empleado en la detección de marcas faciales es el mismo durante todo el proceso de seguimiento, pudiendo o no adecuarse con los rasgos faciales del individuo analizado. Lo ideal, sería emplear un modelo adaptativo que de manera inteligente (con un entrenamiento previo bajo distintos modelos de apariencia) se amoldase al rostro del sujeto.

Para probar el rendimiento de este algoritmo, se puede observar la siguiente fig. 2.12 extraída del *Reto de Seguimiento de Objetos Visuales (VOT)* [63], donde se aprecia el comportamiento del método CCR frente al SDM (*Supervised Descent Model*, modelo referente actual) y el comportamiento del iCCR frente al CCR tradicional.

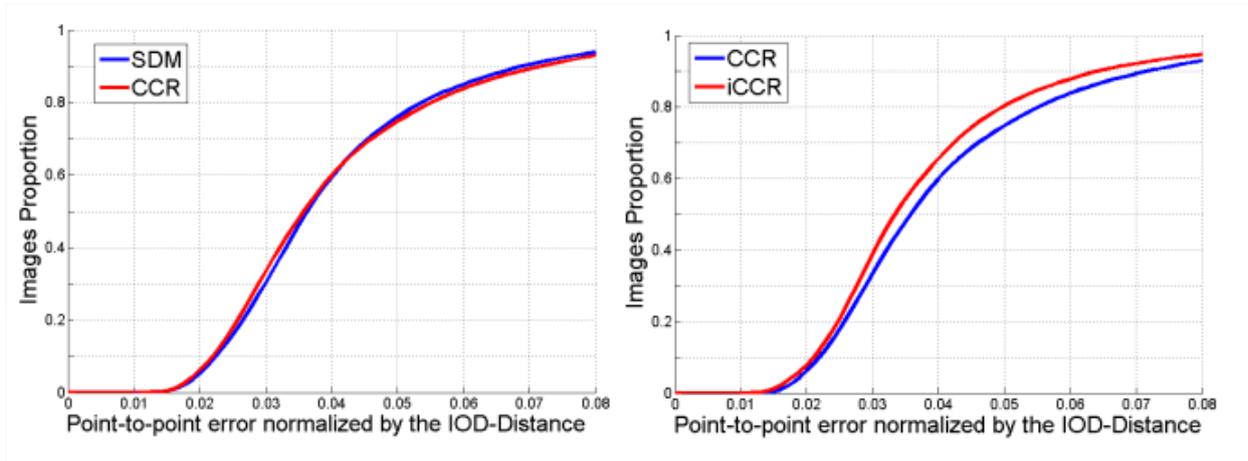


Fig. 2.12: Gráficas comparativas: SDM vs CCR (izda), CCR vs iCCR(dcha).

En fig. 2.12, la gráfica izquierda muestra que tanto en el eje de la proporción de imágenes como en el eje de error punto a punto normalizado, el SDM y CCR presentan un rendimiento similar. Y la gráfica derecha, muestra claramente un beneficio en cuanto a rendimiento empleando un aprendizaje incremental (iCCR) frente al aprendizaje tradicional con modelos genéricos [62].

Por lo tanto, en vista a los resultados anteriores, este algoritmo presenta unas cualidades idóneas para el seguimiento de rostros en aplicaciones realistas a tiempo real como es el objeto de este trabajo.

Máscaras: filtrado espacial.

Una vez rastreados y registrados los puntos que delimitan el rostro del individuo por el algoritmo iCCR-master en cada fotograma, el siguiente paso es establecer una región de interés óptima para los análisis posteriores. Por ello, el siguiente paso es desechar aquellas regiones de la cara (dominio espacial) que contengan muy poca información útil o ninguna para la aplicación de rPPG como, por ejemplo, los ojos, cejas, boca, barba, bigote o las zonas de sombra bajo las narices. Cabe recordar que la información de interés para la estimación de la FC mediante técnicas remotas está fundamentalmente en la imagen de la tez del individuo, por lo que todo lo “añadido” a ésta, se considera espurio.

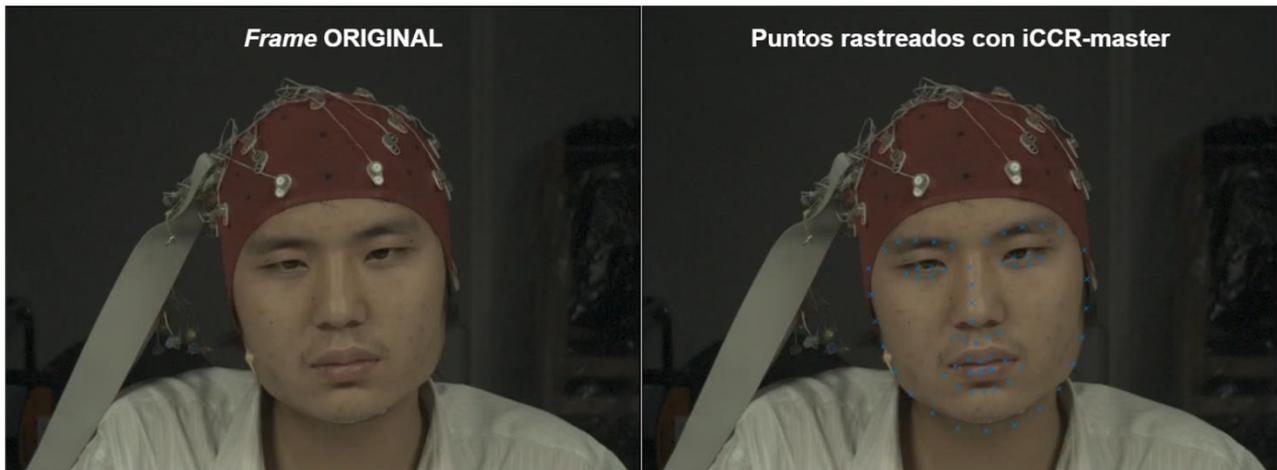


Fig. 2.13: Ejemplo de puntos rastreados con iCCR-master.

Para ello, se han diseñado tres máscaras distintas con el objeto de observar su efecto en los resultados obtenidos por los algoritmos rPPG, aprovechando los puntos faciales obtenidos en la etapa de rastreo facial o *face-tracking*.

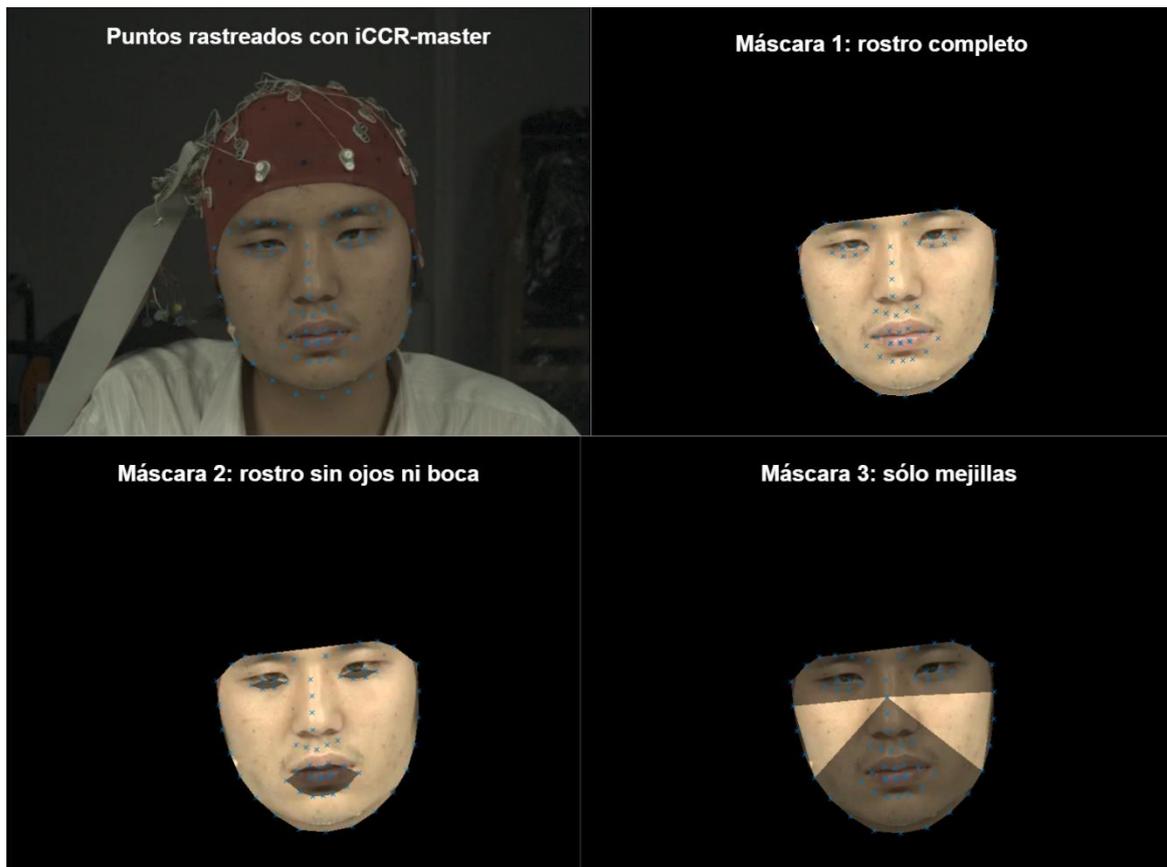


Fig. 2.14: Ejemplo de aplicación de las máscaras sobre un individuo.

Estas máscaras, se han diseñado en un orden incremental de restricción consiguiendo que cada máscara sea más selectiva y consideren un menor número de píxeles válidos que la anterior diseñada, para poder comparar el efecto del filtrado espacial en los resultados finales de rPPG. En la imagen de la fig. 2.14, se puede observar un ejemplo de aplicación de estas máscaras a un individuo. La primera máscara (Máscara 1), está diseñada para segmentar el rostro completo. La segunda (Máscara 2), es una versión más restrictiva en la que se desecha la región asociada a ojos y boca. Y en la tercera (Máscara 3), y más restrictiva, únicamente se selecciona la región asociada a las mejillas.

La obtención de la ROI del fondo necesaria para la rectificación de iluminación es trivial, ya que partiendo del *frame* original, bastará con multiplicarlo por la máscara del fondo, *bg_mask*, obtenida por:

$$bg_mask = 1 - face_mask \quad [Ec. 4]$$

Donde *face_mask* se corresponde con la *Máscara 1* de la fig. 2.14.

Rectificación de iluminación

Tras el proceso de selección de las ROIs faciales y del fondo, el siguiente paso es intentar mejorar la ROI de partida para que los algoritmos rPPG sean lo más precisos posible en la extracción de la señal BVP. Se debe tener en cuenta que, en una situación real, la señal BVP de interés se ve interferida por otras señales tanto lumínicas (provenientes de los reflejos de la pantalla empleada durante la prueba y de la iluminación empleada) como del ruido introducido principalmente por la cámara de video. Para reducir el efecto de estas interferencias, antes de aplicar los algoritmos rPPG, se rectifica la iluminación con la información de las ROIs obtenidas anteriormente.

En primer lugar, se analiza la variación de iluminación en el rostro. Concretamente, se observa el canal *G*, que contiene los niveles de verde y es la componente donde mejor se aprecian estas variaciones, ya que de las tres componentes *RGB*, el verde es el que más absorbe la hemoglobina y la oxihemoglobina de la sangre. Así pues, el nivel de verde de la cara g_{face} puede ser descrito mediante la siguiente ecuación:

$$g_{face} = p + i \quad [Ec. 5]$$

Donde el término p se corresponde con las variaciones del nivel de verde producidas por el pulso cardíaco y el término i con las variaciones del nivel de verde producidas por la iluminación ambiental.

La componente i no se puede obtener directamente, pero tomando el nivel medio de verde en el fondo de cada *frame* como referencia y haciendo uso de la función lineal h , se puede estimar la correlación entre las variaciones del nivel de verde asociadas a las variaciones de iluminación y el nivel medio de verde del fondo $g_{background}$, pudiendo describir i como:

$$i = h \cdot g_{background} \quad [Ec. 6]$$

El vector $g_{background}$ contiene los valores medios de verde de la ROI del fondo, obtenida anteriormente. Los coeficientes de esta función lineal h se pueden encontrar iterativamente por cada fotograma, aplicando un filtro adaptativo NLMS (*Normalized Least-Mean Square*). Una característica de este filtro es que asegura la convergencia de la media y varianza mediante el parámetro μ . Este filtro comienza a partir de un coeficiente h inicial y se actualiza iterativamente en cada fotograma mediante μ hasta que $h(k)$ converge al valor óptimo que minimiza $i - (h \cdot g_{background})$. Esta actualización iterativa viene dada por la siguiente ecuación:

$$h(k+1) = h(k) + \mu \cdot \frac{g_{IR}(k)}{g_{background}^H(k) \cdot g_{background}(k)} \cdot g_{background} \quad [Ec. 7]$$

El término g_{IR} denota la señal de pulso rectificadas de la iluminación, y se obtiene mediante la siguiente fórmula iterativa:

$$g_{IR}(k) = g_{face}(k) - h \cdot g_{background}(k) \quad [Ec. 8]$$

De esta manera, como se puede observar en la fig. 2.15 mediante el filtrado NLMS se reduce el ruido causado por las variaciones de iluminación.

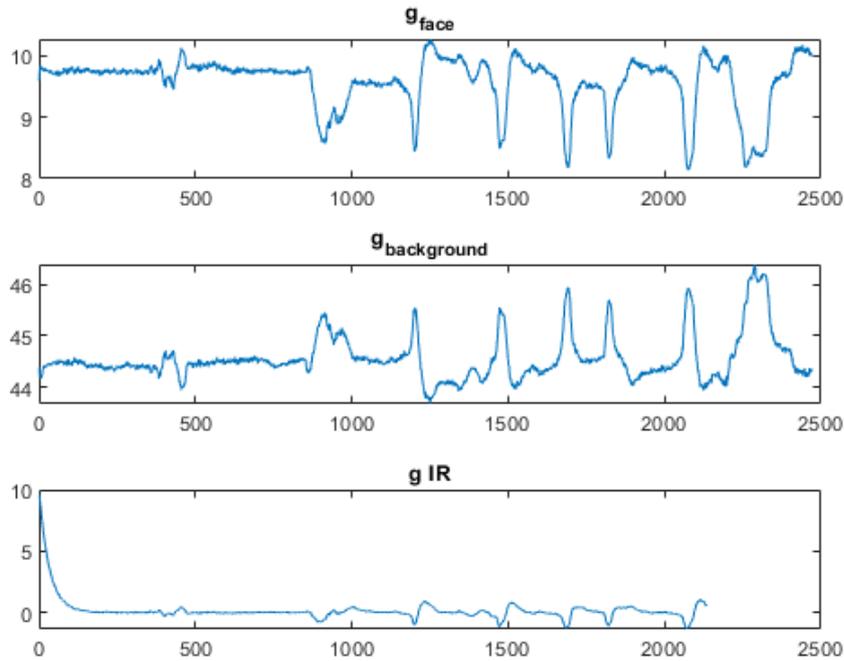


Fig. 2.15: Ejemplo variaciones de nivel de verde en cara, fondo y tras el filtrado NLMS.

Filtrado temporal

Una vez rectificada la iluminación y reducidas las interferencias generadas por las variaciones de iluminación, el siguiente paso es seleccionar únicamente aquella información considerada relevante en un rango de interés. Para esto, se intentan discriminar todas aquellas componentes frecuenciales irrelevantes mediante una serie de filtrados temporales, gracias a la relación directa de los dominios *tiempo-frecuencia*.

El primer tratamiento aplicado ^[64], elimina la línea de base de la salida obtenida del filtrado adaptativo NLMS previo, es decir, del término g_{IR} para analizar la variabilidad de la FC.

Este tratamiento aplica como un filtro paso-alto de respuesta finita variante-temporal. Se emplea un parámetro de suavizado λ que depende directamente de la frecuencia de muestreo de la señal siendo la frecuencia de corte del filtro inversamente proporcional a λ . Cuando λ aumenta, la frecuencia de corte disminuye y viceversa.

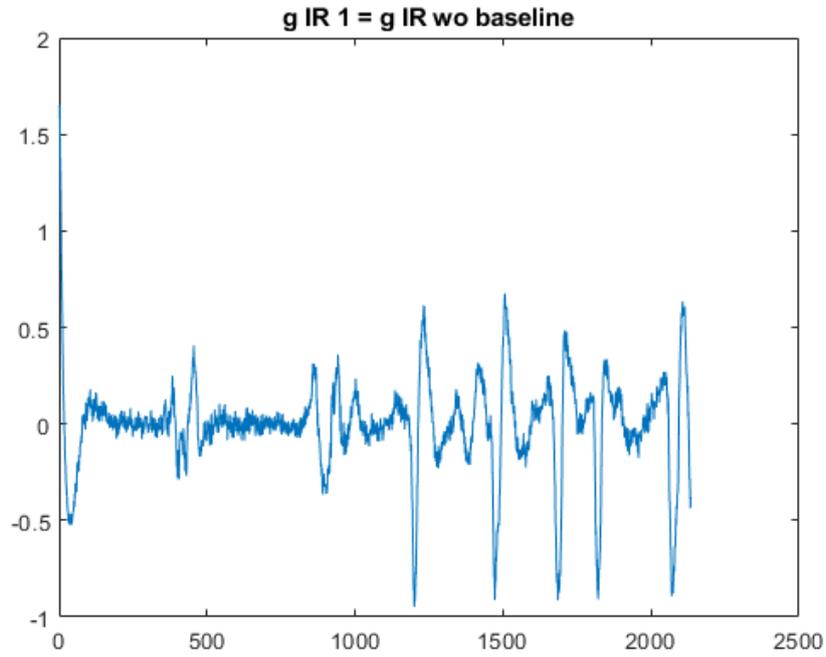


Fig. 2.16: Ejemplo salida del filtro de eliminación de la línea de base.

Para suavizar y eliminar el ruido blanco en la señal de salida del filtrado anterior g_{IR_1} se aplica un filtro de media móvil. Se trata de un filtro paso-bajo que aplica el promedio sobre un grupo de M muestras de entrada para dar una única salida ^[65]. Su efecto en la salida, g_{IR_2} , se puede apreciar con claridad en fig. 2.17.

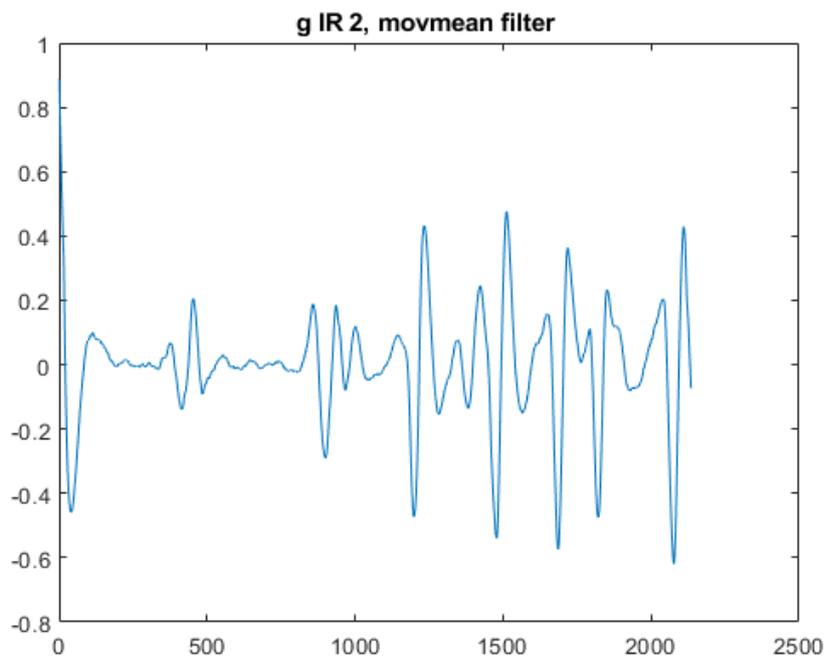


Fig. 2.17: Ejemplo salida filtro de media móvil.

Por último, se aplica un filtrado paso-banda de respuesta finita mediante un enventanado Hamming en el rango de frecuencias de interés, es decir, el rango asociado al pulso cardíaco comprendido entre 0.7 y 4 Hz (Véase pág. 7).

Los filtros de respuesta finita son siempre estables y con una respuesta en fase lineal. Sin embargo, para ciertas especificaciones requieren un orden muy superior a los filtros de respuesta infinita. Por ello, se emplea un enventanado para intentar suavizar los efectos del truncamiento abrupto.

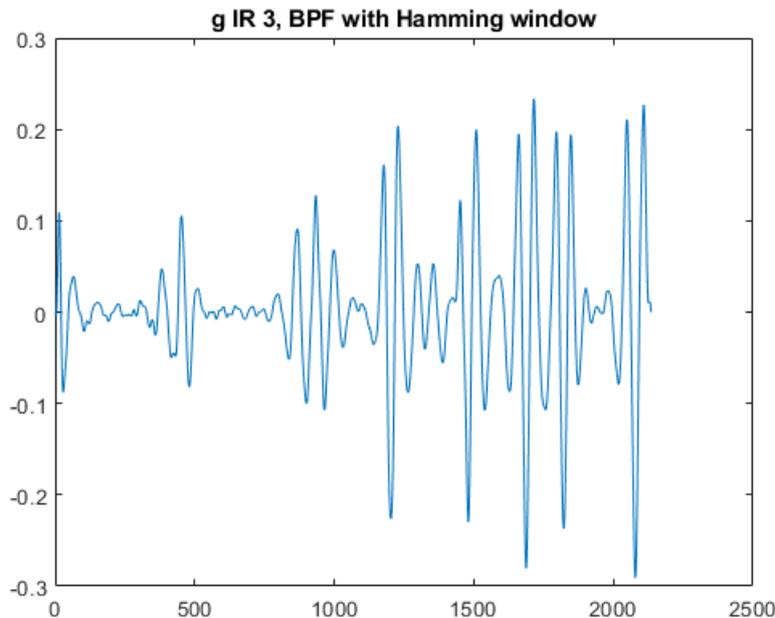


Fig. 2.18: Ejemplo salida filtrado paso-banda con enventanado Hamming.

2.2.3 Extracción de la señal BVP

Una vez preparadas y optimizadas las ROI faciales de cada individuo, el siguiente paso es extraer la señal asociada al pulso cardíaco (señal BVP) implícita en ellas y poder obtener una estimación de la frecuencia cardíaca del individuo. Para ello, se hace uso de los algoritmos de rPPG más relevantes como son el Green Verkruyse ^[66], POS ^[67], CHROM ^[68] e ICA ^[69], los cuales se describen *grosso modo* a continuación.

Green Verkruyse (GVERK)

Es un algoritmo desarrollado por Wim Verkruyse *et al.* y su aplicación es muy sencilla. En primer lugar, se extrae el canal verde (G) de la ROI de interés, que como ya se ha mencionado anteriormente, es donde las variaciones de iluminación producidas por el ritmo cardíaco son más notables. Una vez extraída la componente de interés, considerada también fuente BVP, se aplica un filtro paso-banda Butterworth de 3er orden para seleccionar las componentes dentro del rango frecuencial del pulso cardíaco (0.7-4Hz).

Tras aplicar el filtrado el espectro de la señal, se calcula el periodograma. Se considera que la componente asociada al pulso cardíaco es aquella con mayor potencia, por lo que basta con localizar la máxima componente de potencia, y multiplicar la frecuencia de ésta por 60 para obtener la FC estimada.

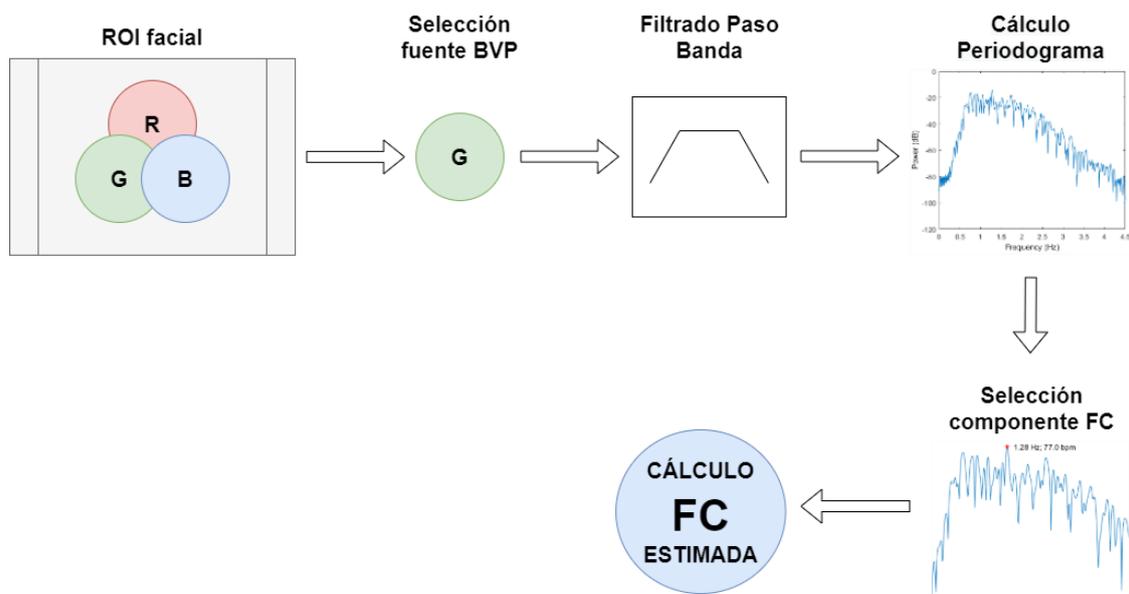


Fig. 2.19: Visión general del algoritmo Green Verkruyse.

POS

El algoritmo POS (*Plane-Orthogonal-to-Skin*) desarrollado por Wang *et al.* emplea el plano ortogonal al tono de la piel en el espacio RGB normalizado temporalmente, para realizar la estimación del pulso cardíaco. Además, este algoritmo prescinde del filtro paso-banda empleado comúnmente para la selección de frecuencias en el rango de interés.

En primer lugar y *frame a frame*, se realiza un promediado espacial de las componentes RGB. Una vez obtenido, este promediado se normaliza temporalmente estas componentes RGB. El siguiente paso es generar el espacio de variación especular mediante la proyección del RGB normalizado. En este espacio, se determina y afinan iterativamente las variaciones asociadas al pulso cardíaco. Por último, mediante suma y superposición de estas variaciones, se extrae la señal BVP. La estimación de la FC a partir de esta señal se realiza de forma análoga al algoritmo anterior, mediante periodograma.

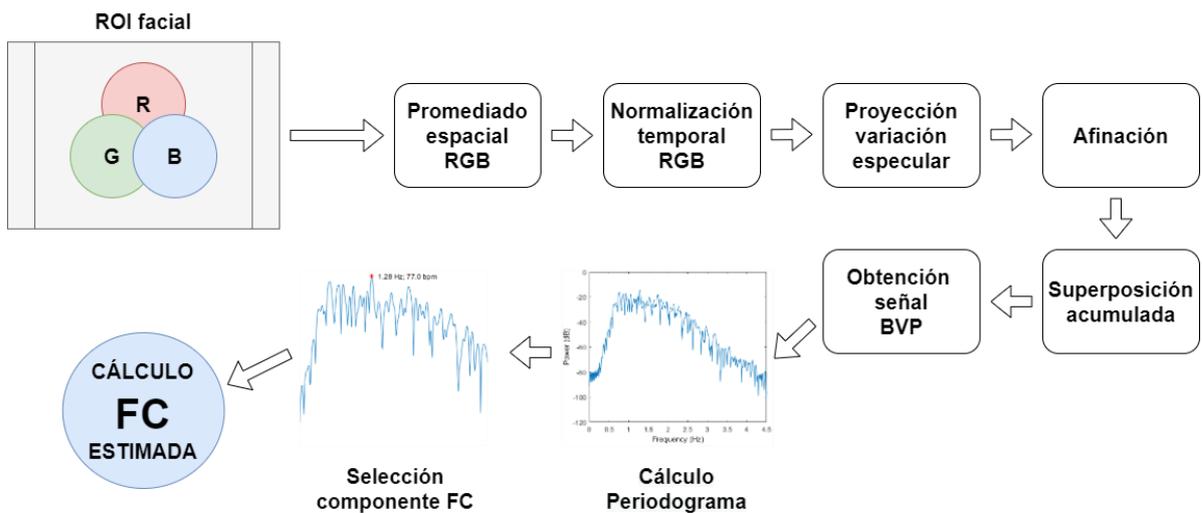


Fig. 2.20: Visión general del algoritmo POS.

CHROM

El algoritmo CHROM desarrollado por Gerard de Haan y Vicent Jeanne está basado en las crominancias de la imagen. Parte de la base que las reflexiones especulares afectan de igual manera en las tres componentes RGB ya que, asumiendo luz blanca, se suma una porción especular idéntica sus respectivas componentes de reflexión difusa.

Esto implica que se puede eliminar la componente de reflexión especular aplicando la diferencia de color, es decir, de las señales de crominancia. Para ello, con los tres canales de color (RGB) normalizados se construyen dos señales ortogonales de crominancia aplicando la diferencia de color. En estas dos señales de crominancia, las variaciones debidas a los cambios del volumen de sangre en la piel serán probablemente distintas, mientras que el movimiento afecta a ambas señales de crominancia por igual.

A estas dos señales de crominancia se les aplica un filtro paso-banda Butterworth de orden 3, cuya banda de paso coincide con el rango de frecuencias asociado al pulso cardíaco. Posteriormente, se aplica un enventanado Hanning con una superposición constante del 50% en cada iteración. Tras realizar este proceso de forma iterativa por cada *frame*, la salida del filtro es la señal BVP. Como ya se ha visto anteriormente, bastará calcular el periodograma de esta señal para estimar la FC.

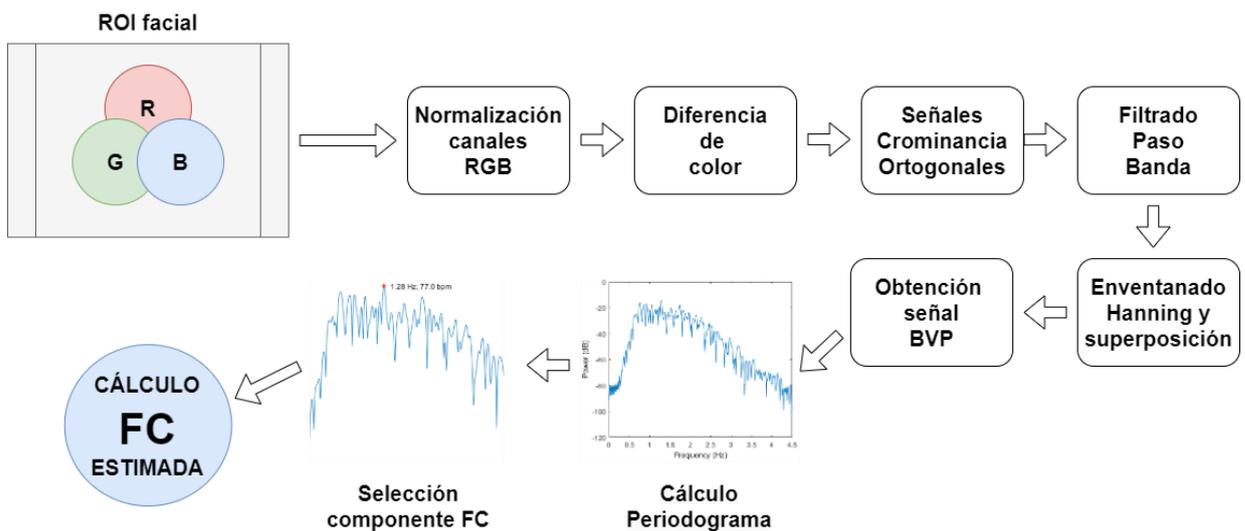


Fig. 2.21: Visión general del algoritmo CHROM.

ICA

El algoritmo ICA ha sido desarrollado por Ming-Zher Poh *et al.* y se fundamenta en el método de separación ciega de fuentes (BSS), concretamente en el análisis de componentes independientes (ICA). Recordando que las componentes RGB provenientes de la ROI facial recogen tanto la información asociada al pulso cardíaco (señal BVP) como de otras señales espurias, es de interés encontrar el modo de discernir qué componentes están asociadas a la señal BVP. Para ello, el análisis de componentes independientes tiene como objeto encontrar la matriz de separación o “desmezclado” de estas señales. Para descubrir las fuentes independientes, esta matriz debe maximizar la no-Gaussianidad de cada fuente.

Tras calcular y aplicar la matriz de separación de fuentes, se obtiene como resultado una estimación del vector que contiene las señales de las fuentes subyacentes. Estas señales, en este caso son tres componentes procedentes de las tres componentes RGB tras aplicar el ICA.

Así pues, para estimar la frecuencia cardíaca se aplica la Transformada Rápida de Fourier (FFT) a cada componente para obtener el espectro de potencia. Por último, el pico de mayor potencia de los tres espectros calculados se considera el asociado a la FC. Multiplicando su frecuencia por 60 obtenemos la FC de interés.

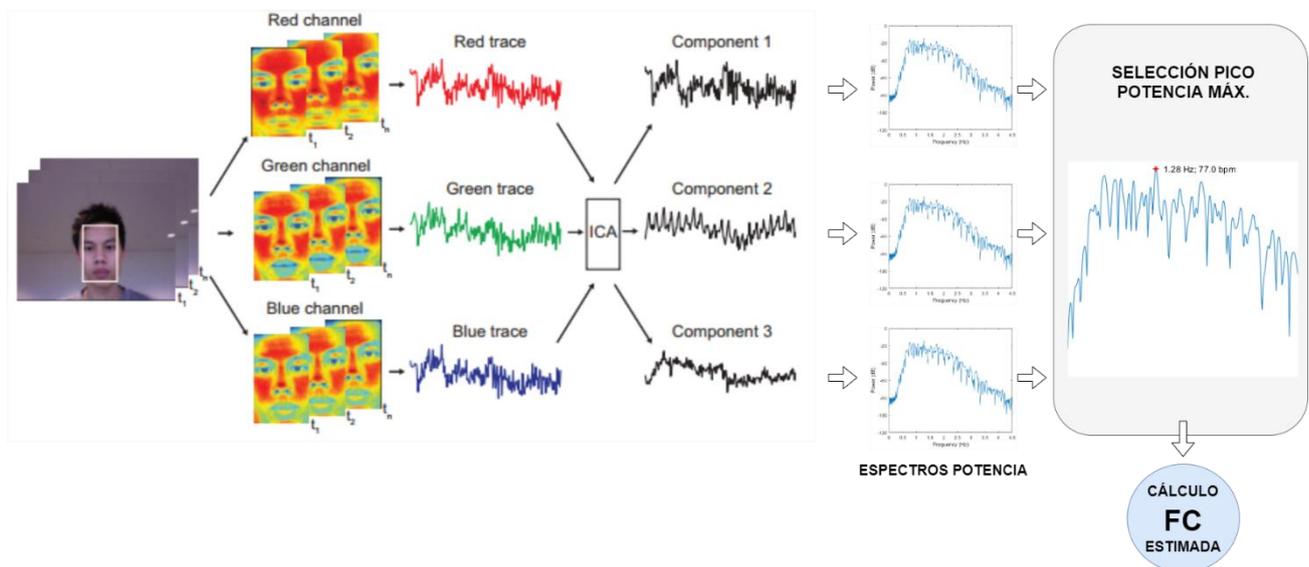


Fig. 2.22: Visión general del algoritmo ICA.

2.3 Arquitectura del modo online (GUI)

Para la implementación de un análisis rPPG en tiempo real (véase fig. 2.2), modo *online*, se ha diseñado una interfaz desarrollada en *Matlab* [70] y a partir de la cual el usuario puede grabar su cara mediante una webcam para obtener en directo una estimación de su frecuencia cardíaca.

La hoja de ruta a seguir en este análisis es similar al comentado anteriormente para el modo *offline*, a excepción de que esta versión en directo requiere una captura, *face-tracking* y procesado de señal de forma iterativa a tiempo real.

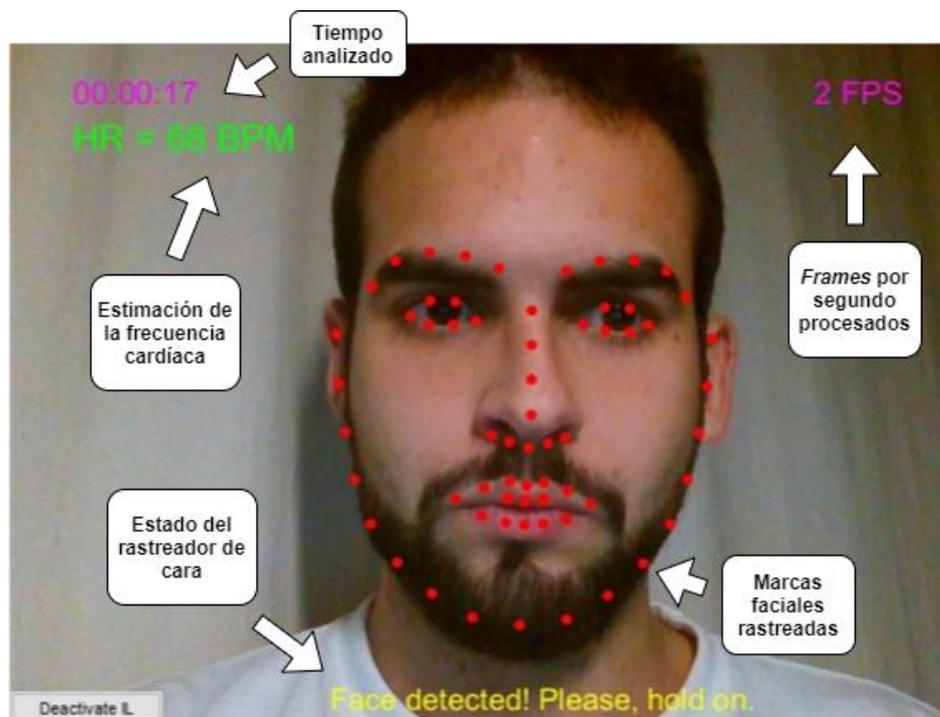


Fig. 2.23: Ejemplo de la GUI del modo online.

Para poder realizar una estimación aceptable de la FC, se requieren al menos 30 segundos de captura de video para que el promedio de las pulsaciones en este tiempo sea significativo. Por esta razón, en la fase de recopilación de datos del modo online, se aplica el rastreo facial a los *frames* comprendidos en los 30 primeros segundos de captura (mediante iCCR-master) y se almacenan las ROI faciales obtenidas en un *buffer*, aunque de momento no se muestra ninguna estimación al usuario.

Una vez completado el *buffer*, se analiza el conjunto de ROI faciales almacenados mediante el algoritmo rPPG elegido y se obtiene una primera estimación de la FC que es mostrada en la interfaz al usuario (GUI).

Tras este primer resultado, en la siguiente captura se aplica *face-tracking* para determinar su ROI facial, y de nuevo, esta ROI se almacena en el *buffer*. El tamaño del *buffer* es fijo, por lo que por cada nueva ROI que entra en él se desecha la ROI más antigua, siguiendo un sistema FIFO (*First In First Out*).

Con esta nueva ROI añadida del nuevo *frame* capturado, se vuelven a aplicar los algoritmos rPPG y se estima la FC nuevamente. El proceso, a partir de este momento, es recursivo a lo largo del tiempo de grabación.

Apartado 3

Resultados

Esta sección recopila todos los resultados extraídos en los análisis rPPG realizados en el modo *offline* para la estimación de la frecuencia cardíaca en los videos siguiendo la línea cronológica del trabajo realizado.

En primer lugar, se optó por adaptar el *script* open-source de rPPG publicado por Kooij-Naber ^[71] para realizar el análisis en cascada de la base de datos *MAHNOB HCI*.

Mas tarde, se decidió probar los métodos rPPG incluidos en *iPhys-toolbox* ^[72] publicados por Daniel McDuff. el cual hace una excelente recopilación de los algoritmos rPPG más relevantes hasta la fecha. Estos algoritmos son los descritos en este trabajo, y han sido testeados en las bases de datos *MAHNOB HCI* y *COHFACE*. En esta segunda prueba, los resultados fueron notablemente mejores (como se verá en páginas posteriores) y por ello, se tomaron como base para una posible optimización de este análisis.

Los resultados obtenidos se muestran en forma de gráficas, donde se comparan por cada video analizado, su resultado de la estimación de FC por rPPG con el resultado *groundtruth* calculado a partir de los datos ECG de los sensores. Y para añadir rigor a esta comparación, en los casos mayor relevancia se preparan unas tablas donde se incluyen los siguientes parámetros estadísticos: error medio (*err_avg*), desviación estándar del error (*err_std*), error cuadrático medio (*mse*), correlación lineal (*corr*) y coeficiente de Pearson (*p*), todo ello calculado entre los valores *groundtruth* y estimados de la FC.

3.1 Resultados Kooij-Naber (MAHNOB HCI)

En el método Kooij-Naber, se emplea el algoritmo ICA para la estimación de la FC y se proponen dos etapas extra de mejora de la señal BVP: filtrado paso-bajo (FPB) y filtrado de la señal de respiración (Resp).

Según el artículo ^[71] del autor y como se puede observar en la fig. 3.1 extraída del mismo, los mejores resultados de este método se obtienen tras aplicar las tres etapas ICA + FPB + Resp llegando incluso, en el caso de la cara, a correlaciones muy próximas a la unidad (0.97).

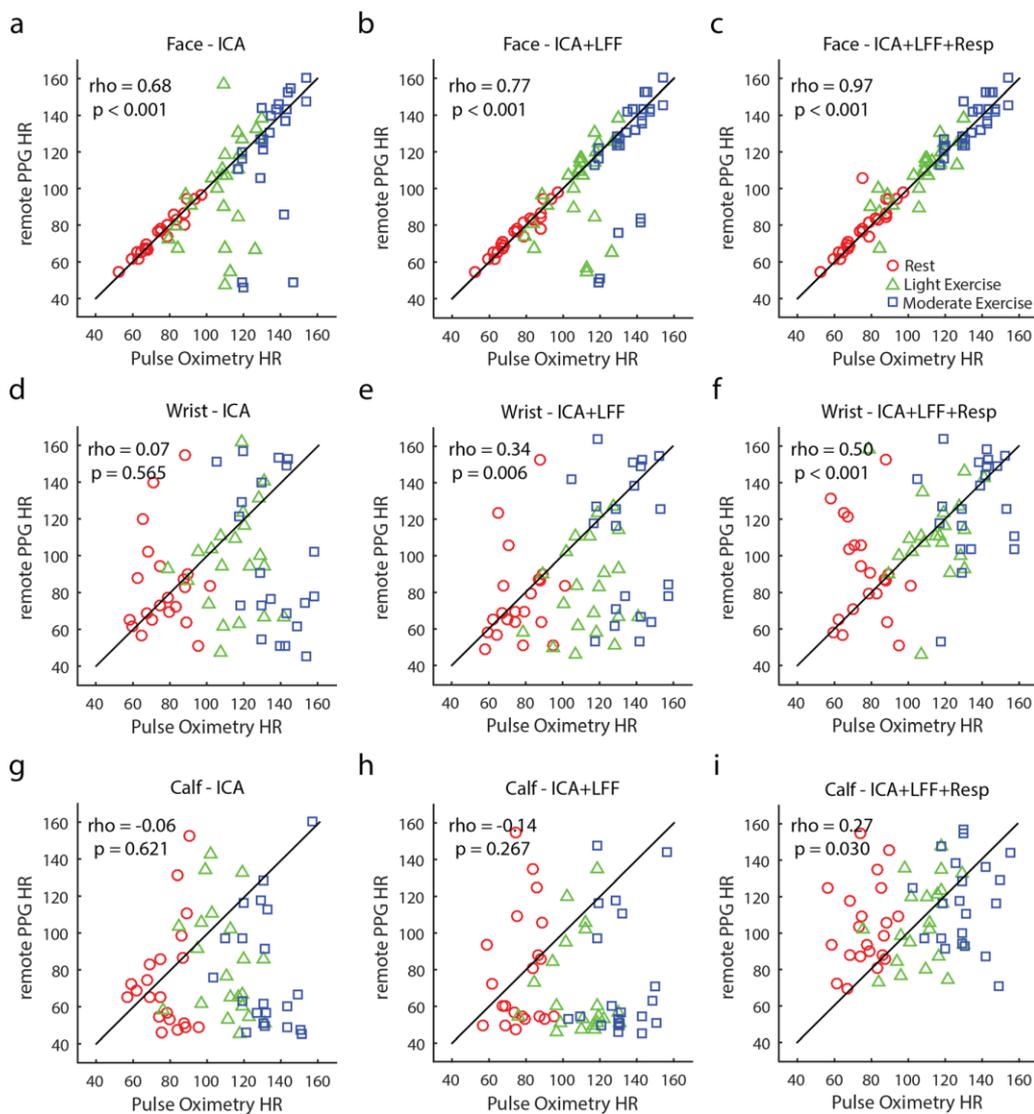


Fig. 3.1: Resultados obtenidos por Kooij-Naber.

Por este motivo, en la adaptación realizada en este trabajo para el análisis en cascada de la MAHNOB HCI, se han decidido extraer los resultados aplicando las tres etapas mencionadas.

Tras los análisis rPPG de la base de datos, los resultados obtenidos son los siguientes:

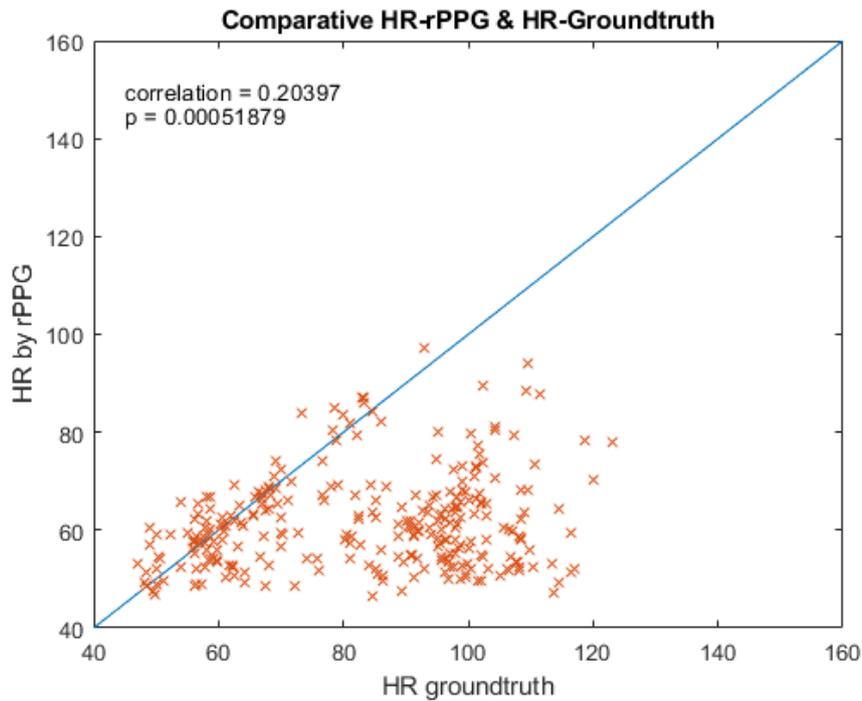


Fig. 3.2: Comparación entre FC groundtruth y estimada (coherence-based).

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 23.0672 | 18.1552 | 860.5562 | 0.2039 | < 0.001 |

Tabla 1: Valores de los parámetros estadísticos (coherence-based).

En vista de la fig. 3.2, con los resultados basados en la coherencia de la señal (coherence-based), se puede apreciar que una gran parte de las estimaciones distan mucho de sus valores *groundtruth*, teniendo en cuenta que la recta azul representa el caso ideal, donde los valores estimados y *groundtruth* coinciden. Por otra parte, se ve con claridad que el algoritmo en el rango entre 50 y 70 bpm extrae resultados más precisos y con menor error. Sin embargo, a partir de 70 bpm el algoritmo empieza a subestimar la FC. Como se refleja en la Tabla 1, esta subestimación provoca que la correlación entre las estimaciones y los *groundtruth* sea muy baja, obteniendo un error promedio superior a 20 bpm.

Por otro lado, los resultados de las estimaciones basadas en la potencia de la señal (*power-based*) no son muy alentadores, mostrando una correlación de 0.135 y un error promedio de 25 bpm.

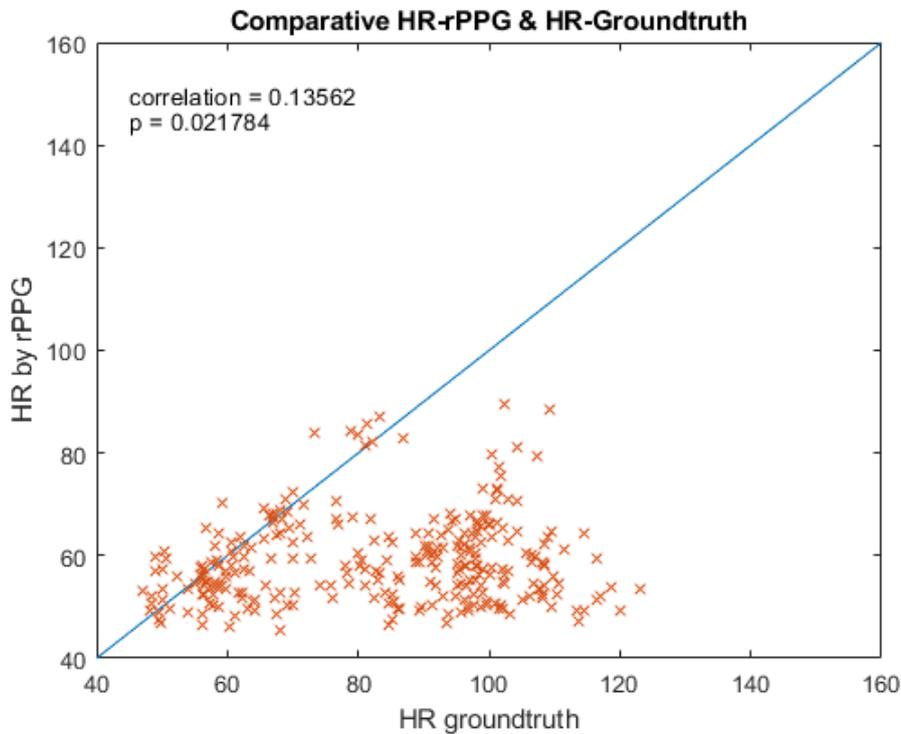


Fig. 3.3: Comparación entre FC groundtruth y estimada (*power-based*).

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 25.2316 | 18.779 | 988.04 | 0.1356 | 0.0218 |

Tabla 2: Valores de los parámetros estadísticos (*power-based*).

La gran diferencia mostrada entre los resultados de este trabajo y los publicados por Kooij-Naber empleando el mismo método rPPG, hace pensar que quizá este método sea óptimo para funcionar bajo unas condiciones de análisis muy concretas, y por las cuales se pueda llegar a unas correlaciones más altas, logrando unos mejores resultados en general.

A continuación, se muestran los resultados obtenidos por otros métodos rPPG (*iPhys-toolbox*) con la intención de ampliar la visión comparativa del comportamiento de otros métodos rPPG sobre la misma base de datos (*MOHFACE HCI*).

3.2 Resultados *iPhys-toolbox* (MAHNOB HCI)

Los resultados mostrados a continuación son los obtenidos tras analizar la base de datos MAHNOB HCI mediante el método GVERK de *iPhys-toolbox* aplicando la máscara 1 (pág. 38) para la selección de la ROI del rostro completo.

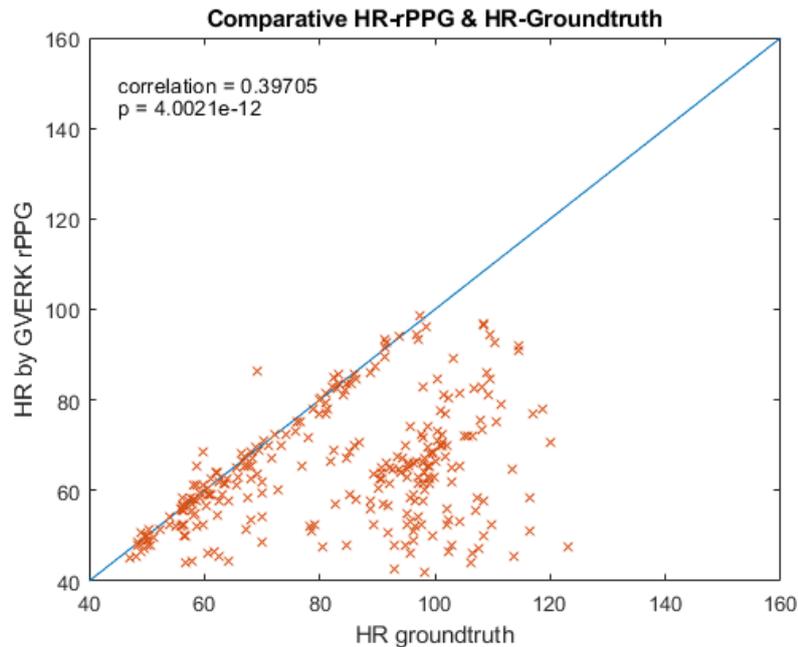


Fig. 3.4: Comparación entre FC groundtruth y estimada, GVERK.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 19.18 | 18.1047 | 697.5012 | 0.397 | < 0.001 |

Tabla 3: Valores de los parámetros estadísticos, GVERK.

En vista de la fig. 3.4 y los parámetros estadísticos de la Tabla 3, el algoritmo *Green Verkruysse* o GVERK muestra unos resultados prometedores. Como se puede observar en este caso, una buena parte de las estimaciones se sitúan en torno a la línea azul de caso ideal. Sin embargo, sigue habiendo un grupo muy amplio de medidas subestimadas a partir de los 70 bpm como ocurría en los casos previos. De hecho, si se compara la fig. 3.4 con las anteriores, se puede apreciar cierta similitud.

Sustituyendo el método GVERK por el POS, y repitiendo los análisis los resultados obtenidos son los mostrados en siguiente gráfica, fig 3.5.

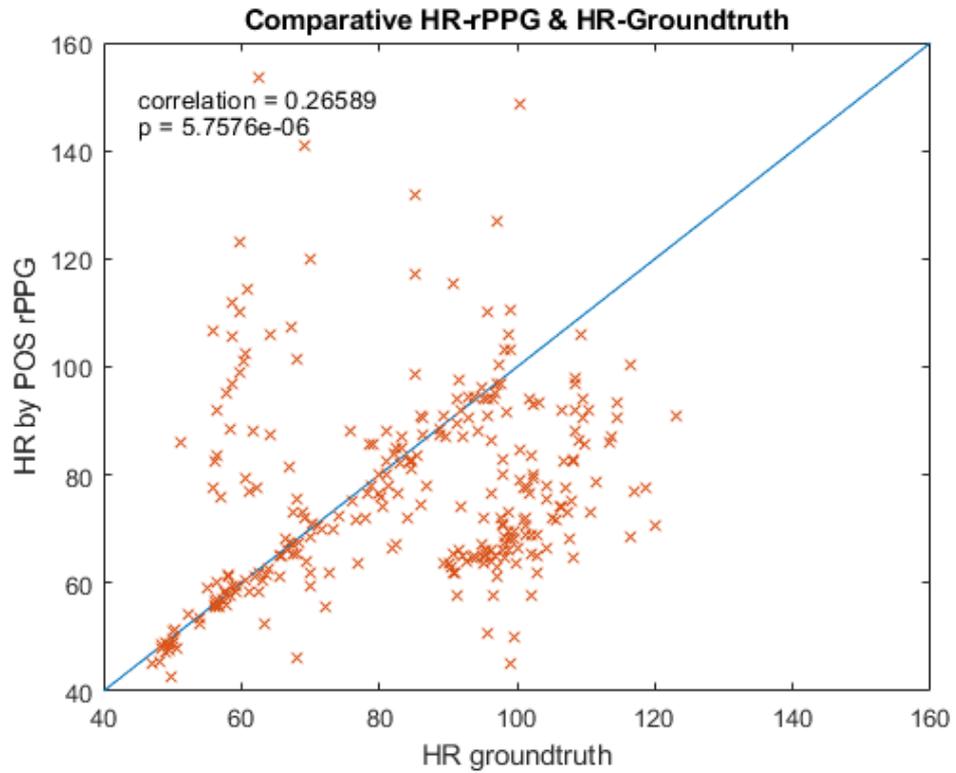


Fig. 3.5: Comparación entre FC groundtruth y estimada, POS.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 17.3214 | 16.3632 | 566.84 | 0.2658 | < 0.001 |

Tabla 4: Valores de los parámetros estadísticos, POS.

En este caso, vemos que la correlación obtenida mediante el método POS es menor a la obtenida anteriormente mediante el GVERK, aunque sigue siendo superior a la calculada mediante el método propuesto por Kooij-Naber. Sin embargo, fijándonos en los parámetros estadísticos, tanto el promedio como la desviación estándar del error es algo inferior.

Con el siguiente método del *iPhys-toolbox*, el CHROM, las estimaciones obtenidas distan bastante de sus respectivos valores *groundtruth*, por lo que la correlación es muy baja, e incluso negativa.

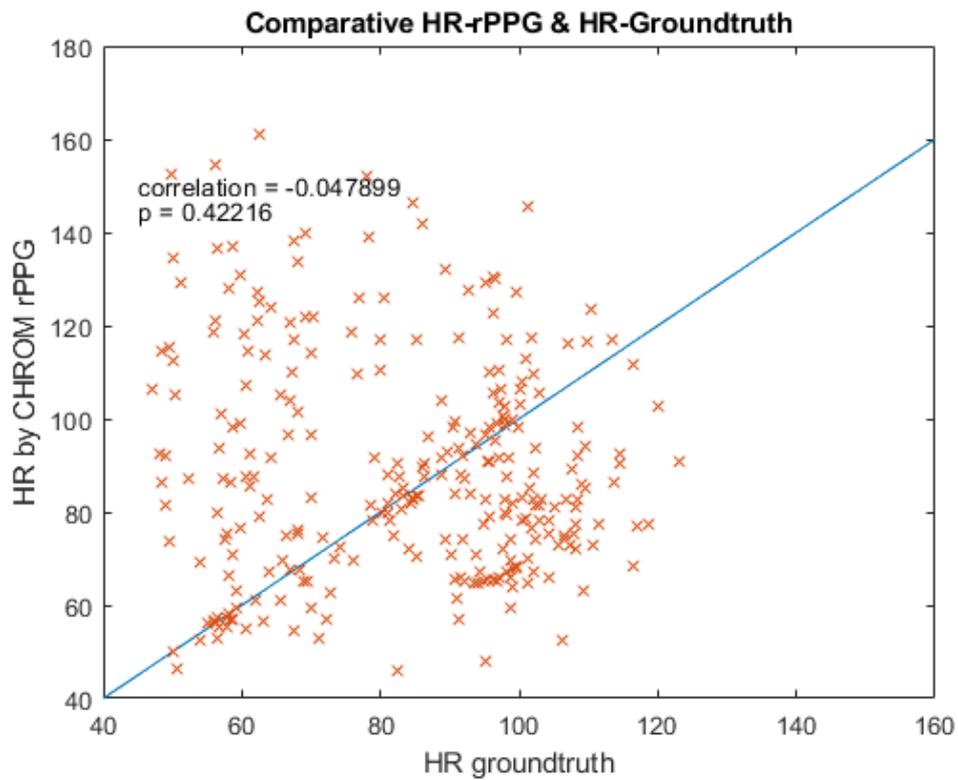


Fig. 3.6: Comparación entre FC *groundtruth* y estimada, CHROM.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 23.8 | 21.21 | 1015.2 | < 0 | < 0.4221 |

Tabla 5: Valores de los parámetros estadísticos, CHROM.

Y, por último, el cuarto método aplicado en los análisis es el ICA, y es similar al empleado por Kooij-Naber en su publicación. Tras observar estos últimos resultados en la fig 3.7, y como es de esperar, tanto el método ICA propuesto por Kooij-Naber como el incluido en *iPhys-toolbox* obtienen valores muy similares en cuanto a la correlación (0.20 y 0.23 respectivamente).

Así mismo, los algoritmos ICA y POS se comportan de manera similar en esta base de datos siendo un poco mejor el método POS, con una correlación ligeramente superior.

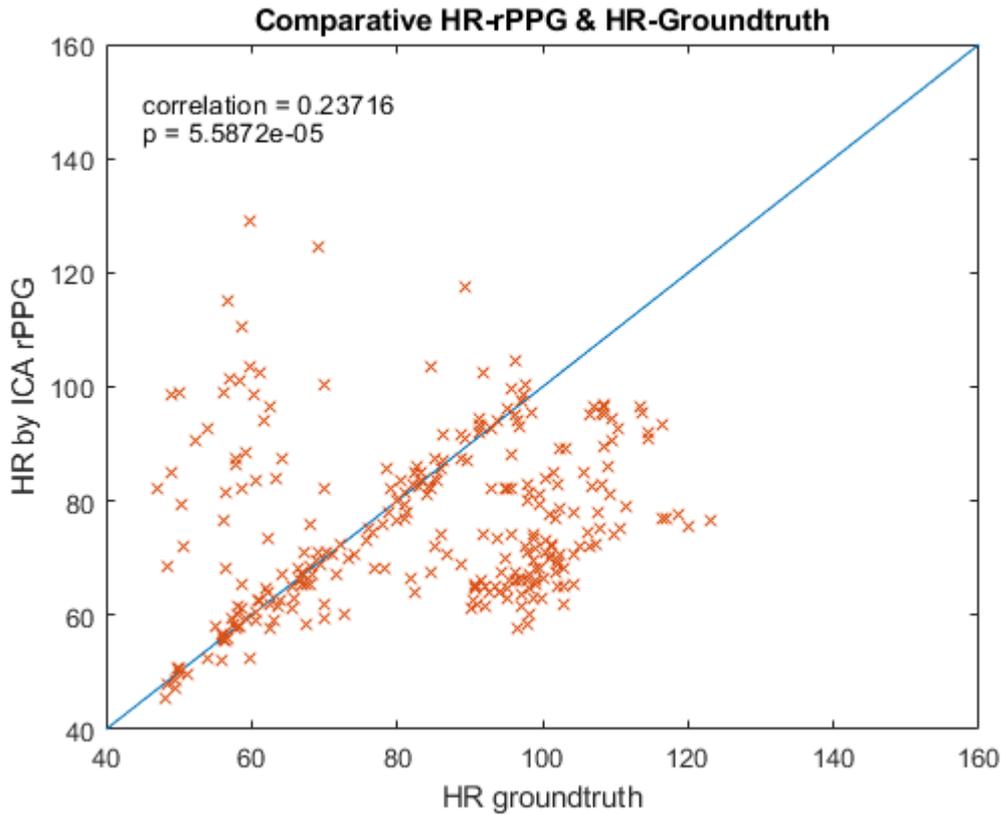


Fig. 3.7: Comparación entre FC groundtruth y estimada, ICA.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 16.93 | 14.93 | 508.7 | 0.2371 | < 0.001 |

Tabla 6: Valores de los parámetros estadísticos, ICA.

De los cuatro métodos empleados, el que mejor estimación obtiene con diferencia es el método GVERK. Este método parece muy prometedor para usarlo como base de partida a una posible optimización, debido a su sencillez de implementación y mayor precisión en los resultados.

Por concluir esta primera comparación, se presentan en orden descendente de correlación (mejor \rightarrow peor) los algoritmos empleados: **GVERK** (0.39), **POS** (0.26), **ICA** (0.23), **CHROM** (< 0).

Optimización mediante máscara más restrictiva

Para intentar mejorar los resultados anteriores, se realiza un filtrado espacial más restrictivo aplicando la *máscara 3* propuesta en el apartado anterior (pág. 38) que sustituye a la máscara de rostro completo (*máscara 1*) utilizada hasta ahora.

A priori, tiene sentido pensar que una selección más precisa, casi idónea, de ROI facial reducirá el ruido en la señal BVP causado por el procesado de las zonas sin piel (zonas espurias), aunque también se discriminarán ciertas zonas de piel con información útil. Bajo esta hipótesis, se muestran a continuación los resultados extraídos una vez aplicada la máscara que incluye únicamente las mejillas del individuo a la ROI facial (*máscara 3*).

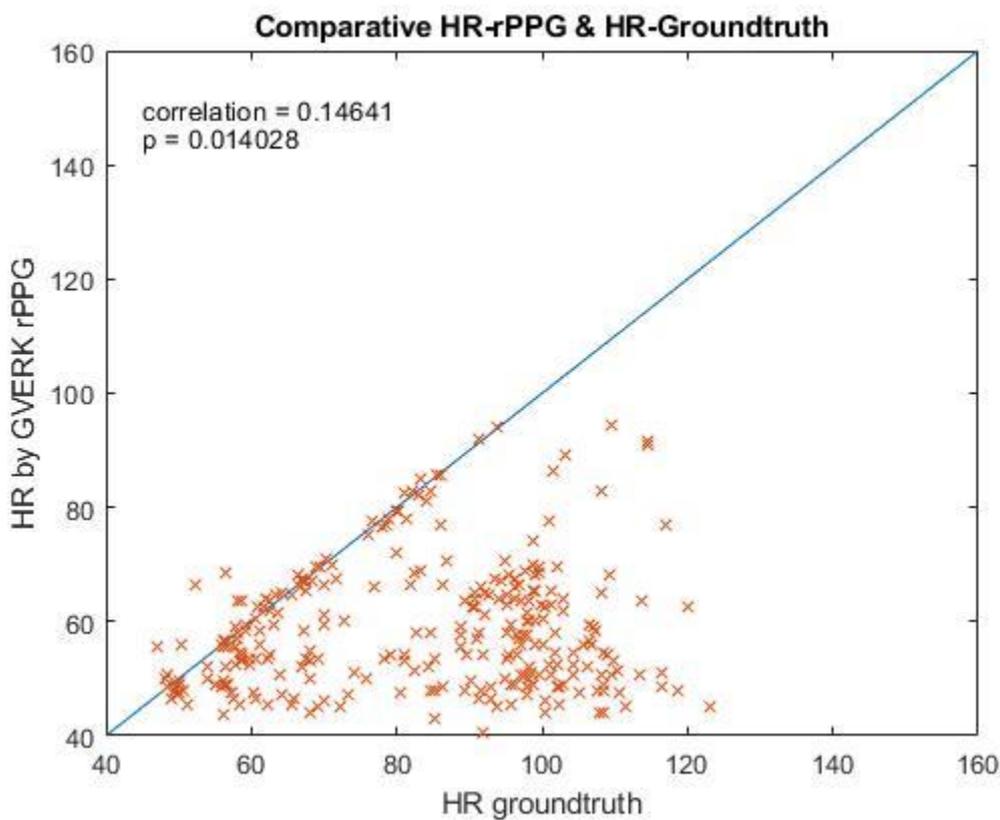


Fig. 3.8: Comparación entre FC groundtruth y estimada, GVERK + máscara 3.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 25.3079 | 20.1998 | 1047.1 | 0.1464 | 0.014 |

Tabla 7: Valores de los parámetros estadísticos, GVERK + máscara 3.

Recordando que el método GVERK era el que mejores estimaciones había realizado hasta ahora, y en vista de la fig. 3.8 y tabla 7, se puede apreciar claramente que ha disminuido drásticamente la eficacia de éste tras la aplicación de una máscara mucho más restrictiva.

Veamos los resultados obtenidos para el resto de algoritmos rPPG probados:

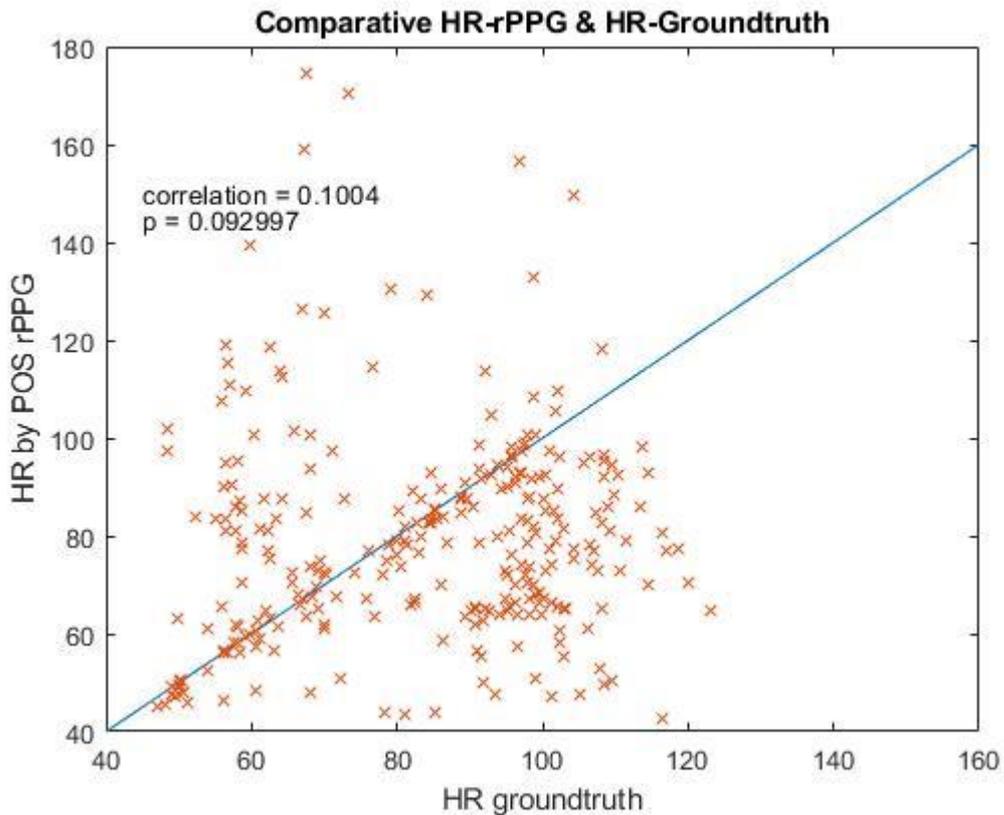


Fig. 3.9: Comparación entre FC groundtruth y estimada, POS + máscara 3.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 20.233 | 19.227 | 777.79 | 0.10 | 0.093 |

Tabla 8: Valores de los parámetros estadísticos, POS + máscara 3.

En este algoritmo (POS), se vuelve a marcar un descenso considerable en la precisión de las estimaciones obtenidas, disminuyendo la correlación de los datos de 0.26 (máscara 1) a 0.10 (máscara 3).

Veamos a continuación cómo se comporta el método CHROM (algoritmo con el peor de los resultados obtenidos con la primera máscara).

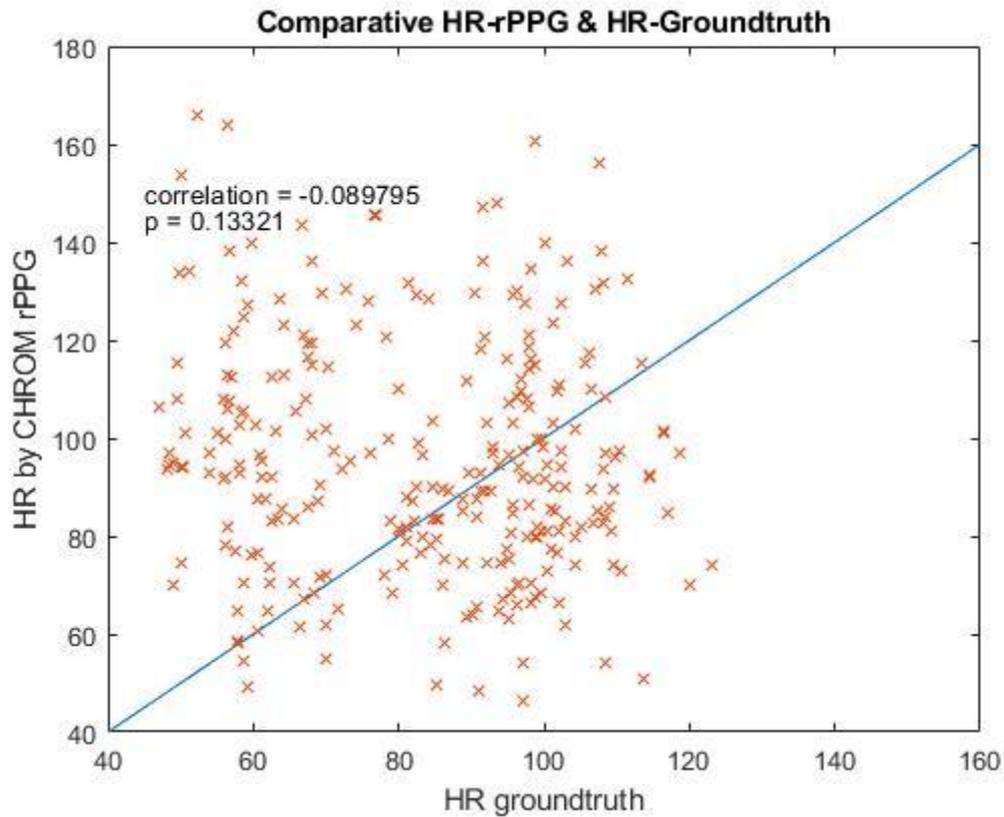


Fig. 3.10: Comparación entre FC groundtruth y estimada, CHROM + máscara 3.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 26.4849 | 21.428 | 1159 | < 0 | 0.1332 |

Tabla 9: Valores de los parámetros estadísticos, CHROM + máscara 3.

En vista de los parámetros estadísticos de la tabla 9 y la dispersión de las estimaciones mostradas en fig. 3.10, el método CHROM sigue siendo el peor método hasta ahora, en los análisis de la MAHNOB HCI.

Para acabar con los resultados de esta segunda máscara, se muestran a continuación las estimaciones por el método ICA.

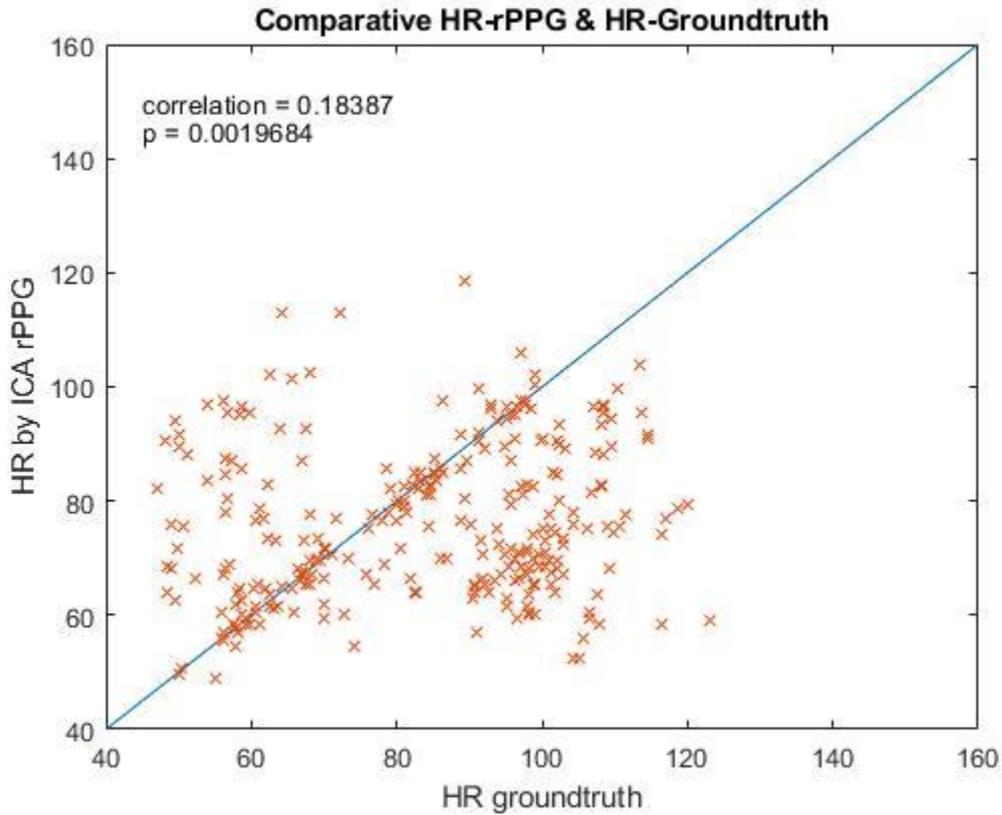


Fig. 3.11: Comparación entre FC groundtruth y estimada, ICA + máscara 3.

| <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|----------------|----------------|------------|-------------|----------|
| 17.1513 | 14.806 | 512.62 | 0.1839 | 0.002 |

Tabla 10: Valores de los parámetros estadísticos, ICA + máscara 3.

En este caso y como se puede intuir de la tendencia marcada a lo largo de este segundo análisis en cuanto a empeoramiento en las estimaciones, el método ICA para esta segunda máscara presenta también peores resultados de correlación disminuyendo de 0.23 para el caso de la *máscara 1* a 0.18.

No obstante, cabe destacar que en esta ocasión el método ICA supera a los resultados del método POS en cuanto a correlación se refiere. Así pues, el orden descendiente de correlación (mejor \rightarrow peor) de los algoritmos empleados sería: **GVERK** (0.39), **ICA** (0.18), **POS** (0.10), **CHROM** (< 0).

A modo de conclusión, tras el análisis de los resultados extraídos por los algoritmos de *iPhys-toolbox* a partir de una ROI facial obtenida con una máscara más selectiva (*máscara 3*, pág. 38), se demuestra claramente que los algoritmos utilizan información proveniente de zonas que no deberían tenerse en cuenta (zonas espurias) como las regiones de bello facial sin información aparente relacionada con la señal de pulso cardíaco.

Optimización mediante rectificación de iluminación

Otra propuesta para la optimización de los algoritmos es la rectificación de la iluminación presente en la ROI de la cara y cuyo proceder se describe en el apdo. 2, pág. 39. En este análisis se ha escogido como algoritmo rPPG el *GVERK* y la máscara de rostro completo (*máscara 1*) ya que son los que mejor resultados han mostrado en las secciones anteriores. A continuación, se muestran las gráficas con los resultados *g_IR_X* (ver págs. 39-43) obtenidos tras cada etapa de filtrado:

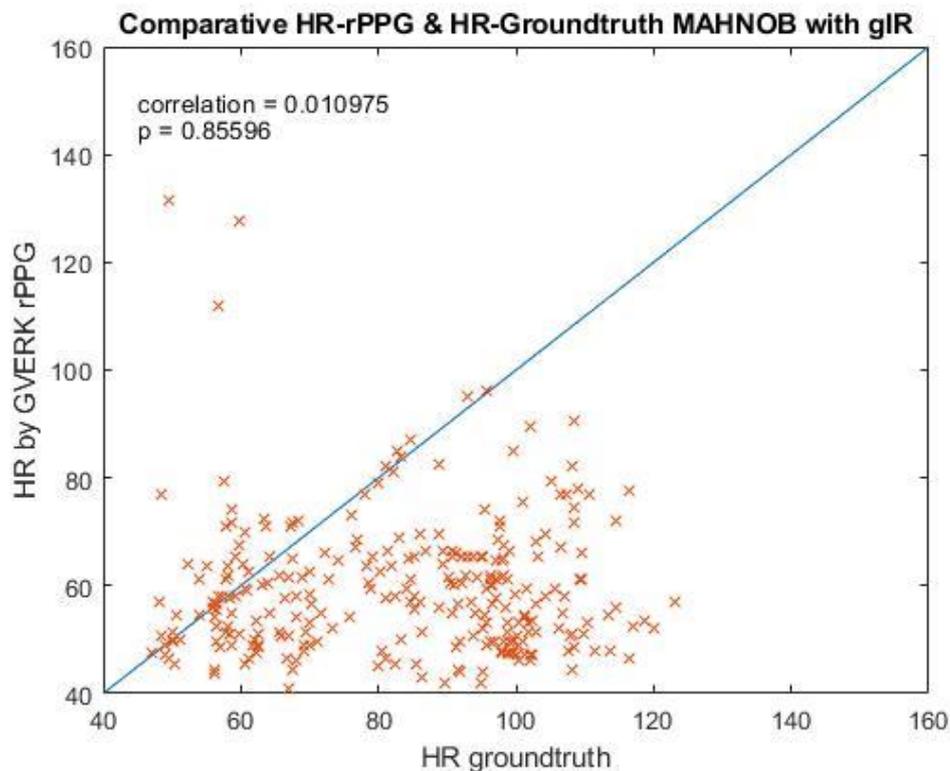


Fig. 3.12: Comparación entre FC groundtruth y estimada, *g_IR*.

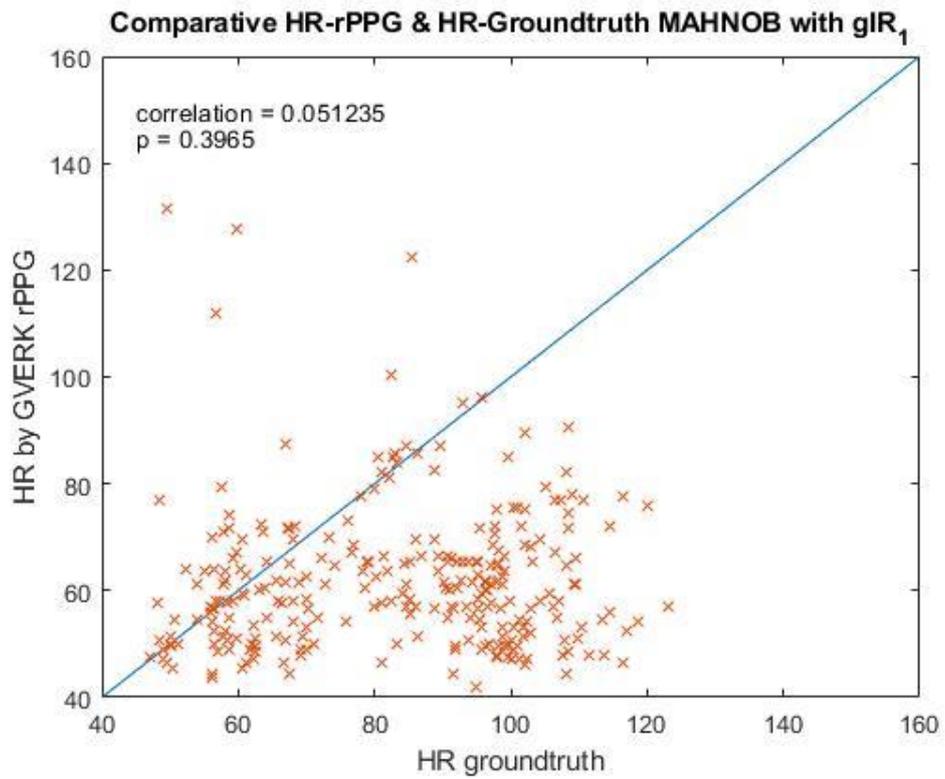


Fig. 3.13: Comparación entre FC groundtruth y estimada, g_IR_1.

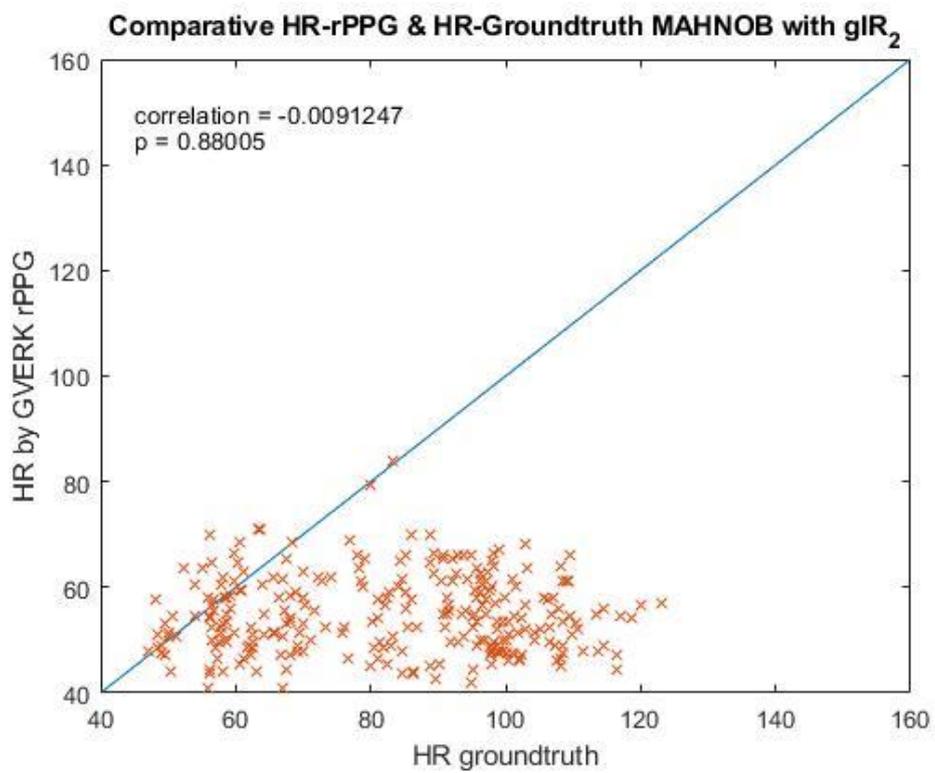


Fig. 3.14: Comparación entre FC groundtruth y estimada, g_IR_2.

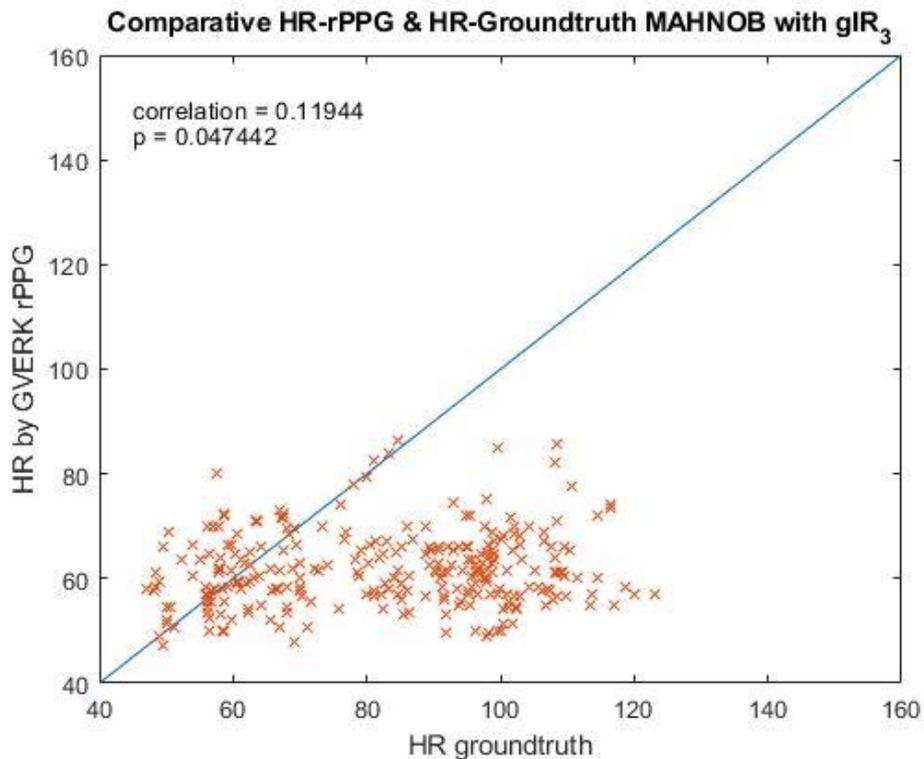


Fig. 3.15: Comparación entre FC groundtruth y estimada, g_{IR_3} .

En vista de la correlación máxima (0.12) asociada a los resultados de g_{IR_3} obtenidos tras las etapas de filtrado en la rectificación de iluminación, y comparando con el valor de partida original (0.39, con GVERK + máscara 1, pág. 54), no se logra una mejoría en los resultados.

3.3 Resultados *iPhys-toolbox* (COHFACE)

Procediendo de manera análoga a los análisis realizados en la *MAHNOB HCI*, en esta sección se presentan los resultados obtenidos con la base de datos *COHFACE*. Esta base de datos, como se ha detallado anteriormente en el apdo. 2, pág. 31, contiene videos grabados con dos tipos de iluminación: natural y de estudio (artificial) por lo que se calculan los resultados para cada tipo de iluminación, así como los globales considerando ambas fuentes.

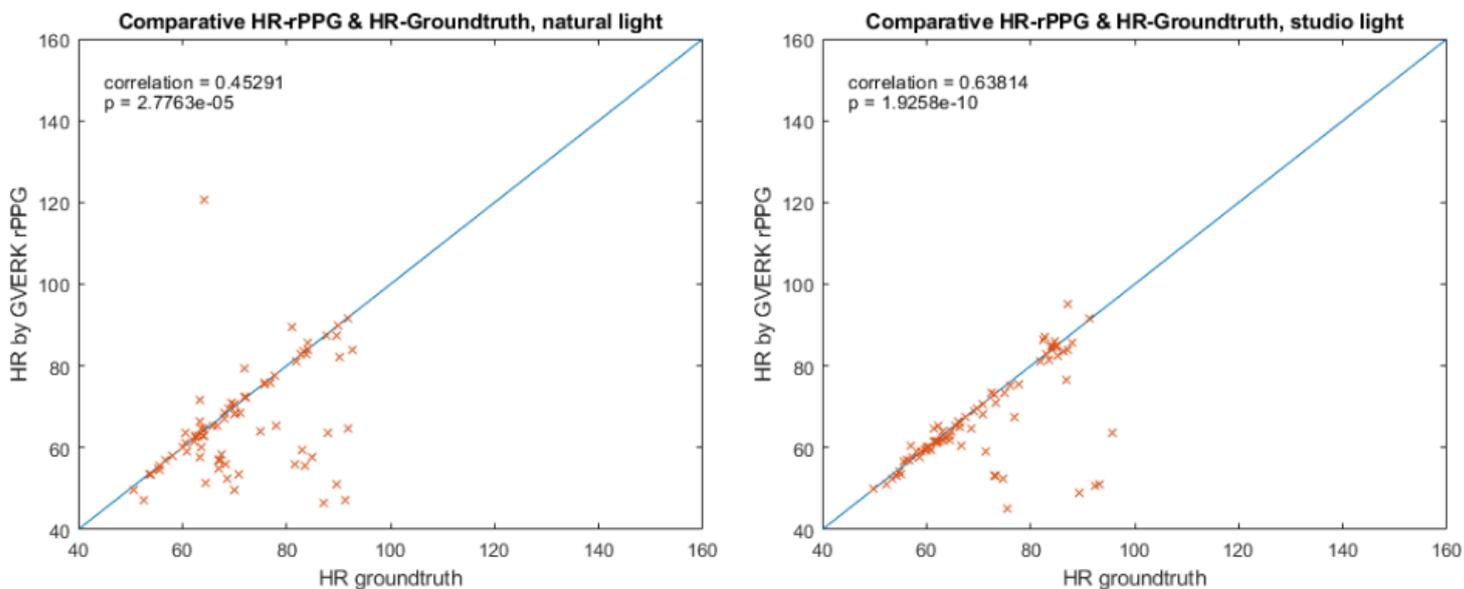


Fig. 3.16: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, GVERK.

| Iluminación | <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|-------------|----------------|----------------|------------|-------------|----------|
| Natural | 7.3055 | 11.684 | 188.169 | 0.4529 | < 0.001 |
| Estudio | 4.6842 | 9.6188 | 113.3 | 0.6381 | < 0.001 |
| GLOBAL | 5.9866 | 10.742 | 150.5 | 0.3608 | < 0.001 |

Tabla 11: Valores de los parámetros estadísticos, iluminación natural, estudio y global, GVERK.

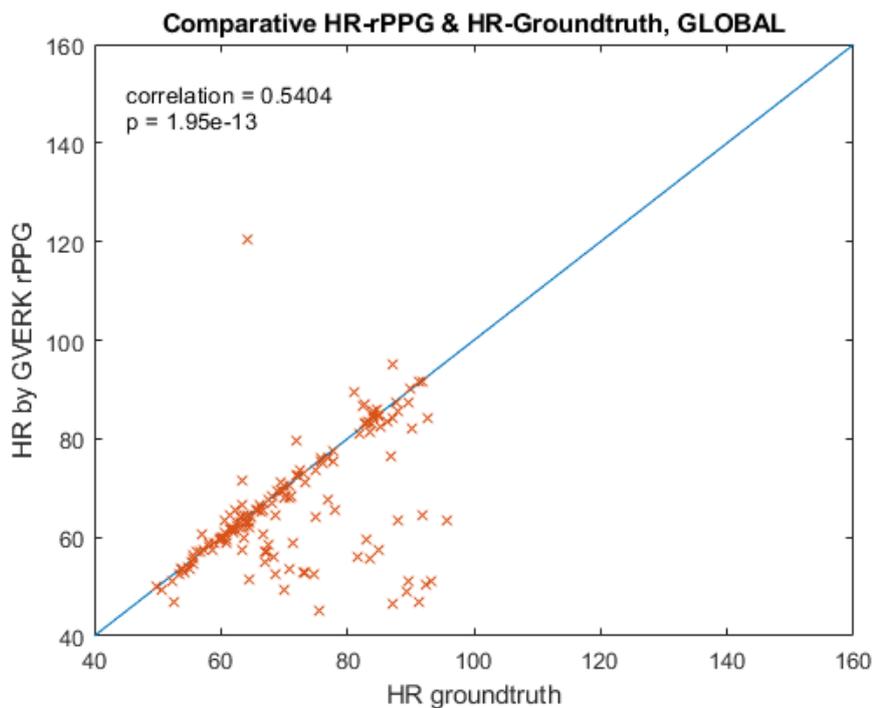


Fig. 3.17: Comparación entre FC groundtruth y estimada, conjunto global, GVERK.

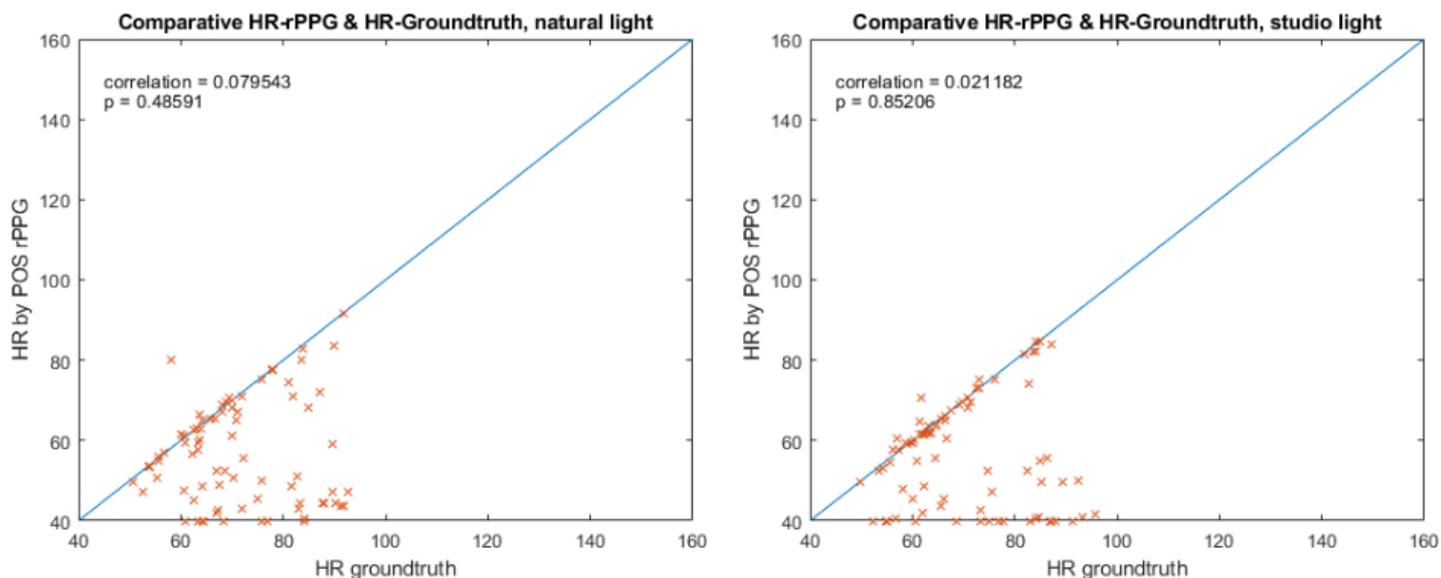


Fig. 3.18: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, POS.

| Iluminación | <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|-------------|----------------|----------------|------------|-------------|----------|
| Natural | 15.0737 | 15.7485 | 472.09 | 0.0795 | 0.0211 |
| Estudio | 14.3599 | 17.1396 | 496.3 | 0.0211 | 0.852 |
| GLOBAL | 14.7146 | 16.4149 | 484.273 | 0.0497 | 0.533 |

Tabla 12: Valores de los parámetros estadísticos, iluminación natural, estudio y global, POS.

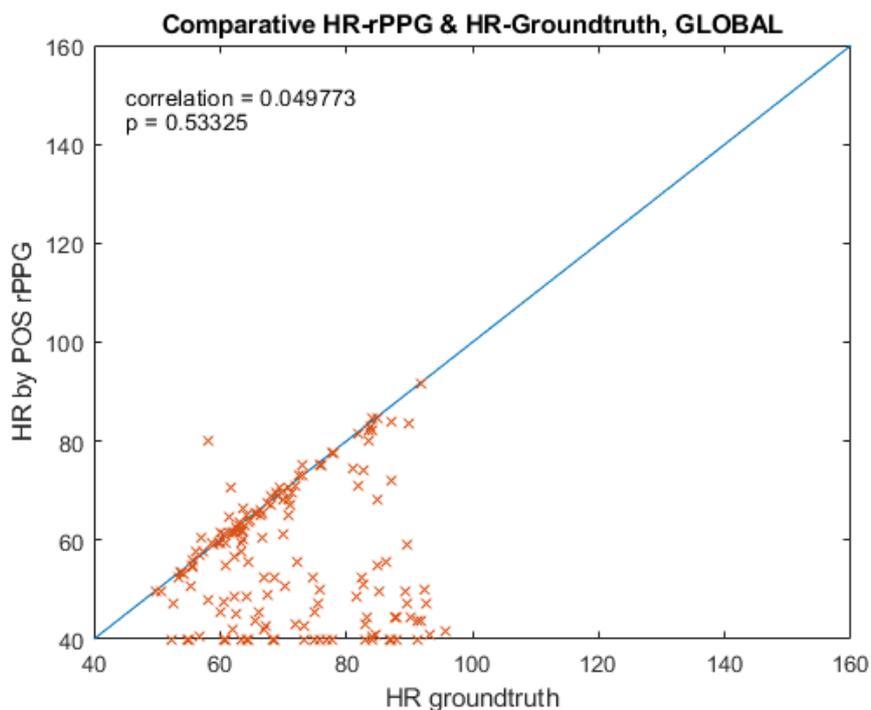


Fig. 3.19: Comparación entre FC groundtruth y estimada, conjunto global, POS.

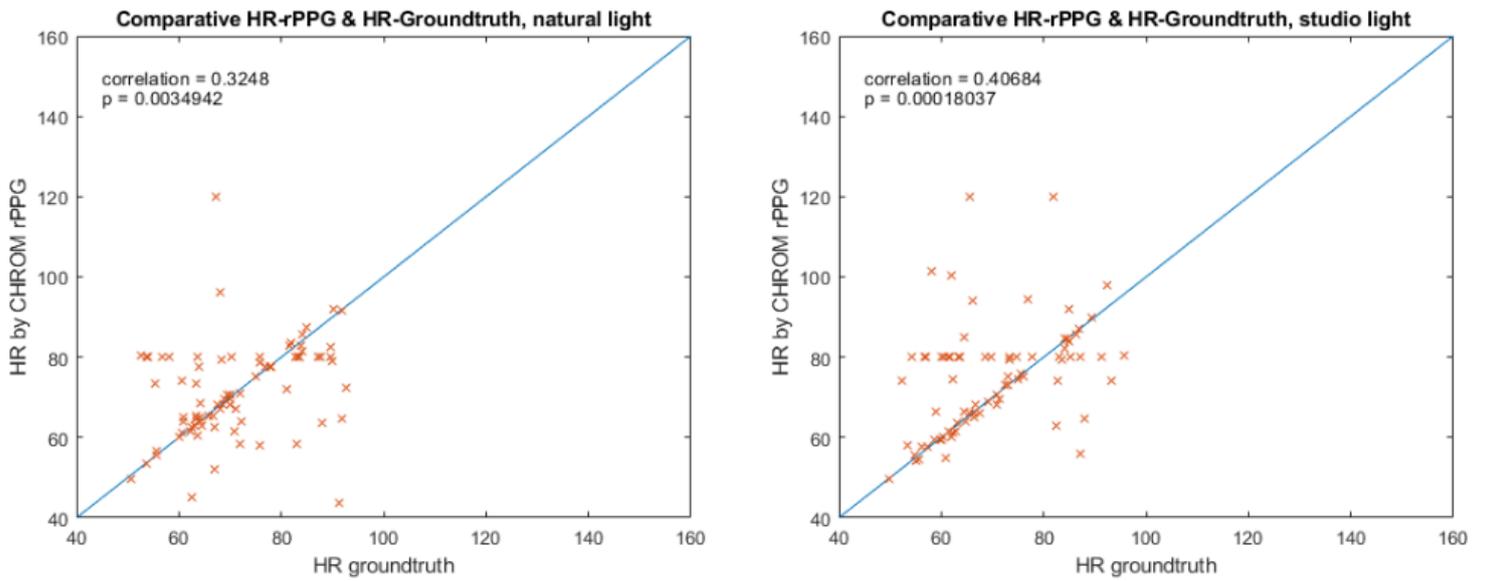


Fig. 3.20: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, CHROM.

| Iluminación | <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|-------------|----------------|----------------|------------|-------------|----------|
| Natural | 8.2114 | 10.8328 | 183.291 | 0.3248 | 0.0035 |
| Estudio | 8.8898 | 11.6519 | 213.096 | 0.406 | < 0.001 |
| GLOBAL | 8.5527 | 11.2219 | 198.288 | 0.36 | < 0.001 |

Tabla 13: Valores de los parámetros estadísticos, iluminación natural, estudio y global, CHROM.

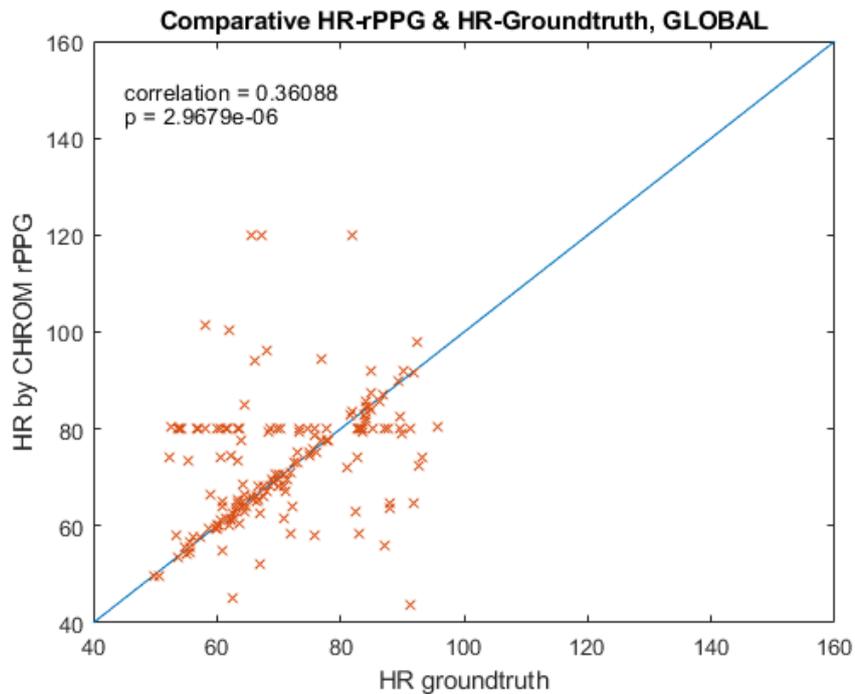


Fig. 3.21: Comparación entre FC groundtruth y estimada, conjunto global, CHROM.

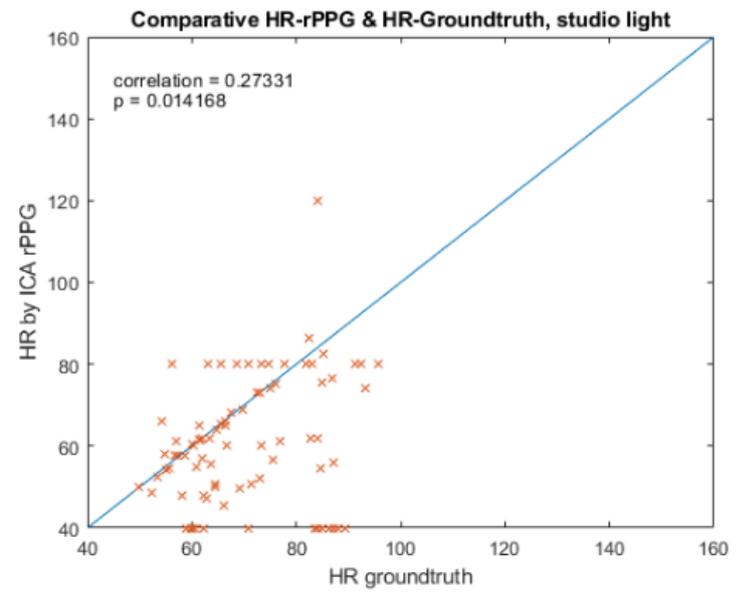
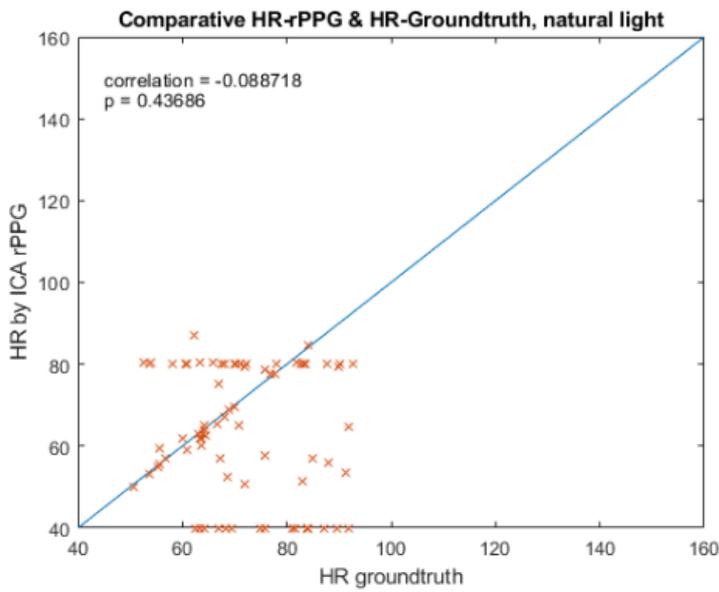


Fig. 3.22: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, ICA.

| Iluminación | <i>err_avg</i> | <i>err_std</i> | <i>mse</i> | <i>corr</i> | <i>p</i> |
|-------------|----------------|----------------|------------|-------------|----------|
| Natural | 15.5775 | 14.96 | 463.6434 | < 0 | 0.4368 |
| Estudio | 13 | 13.695 | 354.1581 | 0.2733 | 0.0141 |
| GLOBAL | 14.28 | 14.349 | 408.556 | 0.096 | 0.2268 |

Tabla 14: Valores de los parámetros estadísticos, iluminación natural, estudio y global, ICA.

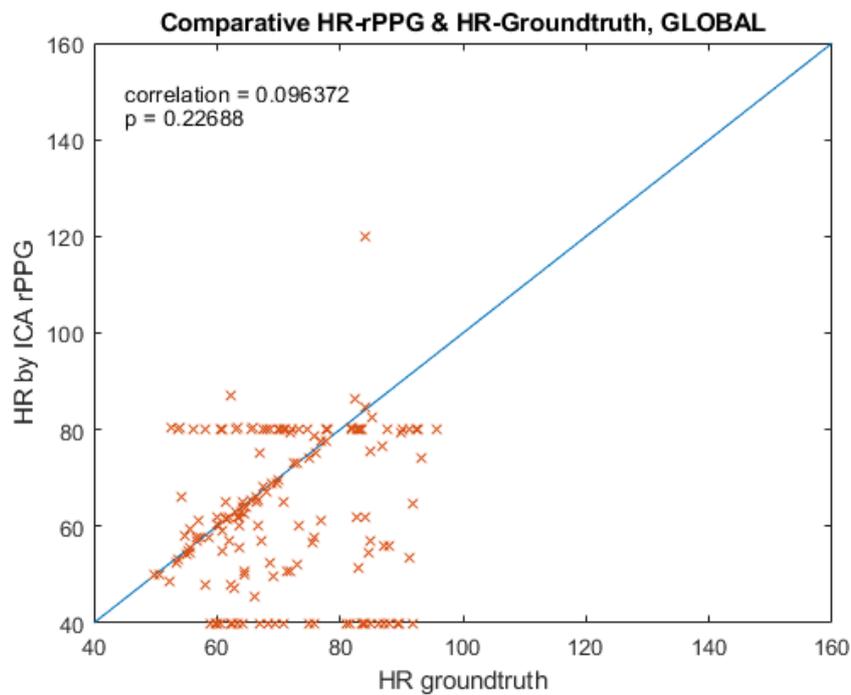


Fig. 3.23: Comparación entre FC groundtruth y estimada, conjunto global, ICA.

En vista de los resultados mostrados en las páginas anteriores se llega a la conclusión de que, en el caso de la *COHFACE*, los algoritmos del *iPhys-toolbox* empleados se comportan de manera distinta a como venían haciéndolo hasta ahora. En este caso, las mejores estimaciones se obtienen mediante los métodos GVERK y CHROM, con una correlación global de 0.36 en ambos casos mientras que las peores se corresponden a los resultados del ICA y POS, con correlaciones de 0.22 y 0.04 respectivamente.

El método CHROM presenta una mejoría muy significativa en comparación con los resultados obtenidos para la *MAHNOB HCI* (pág. 56). Por otra parte, en líneas generales los algoritmos ofrecen mejores estimaciones bajo luz artificial (de estudio) a excepción del POS que muestra resultados ligeramente mejores bajo luz natural.

Optimización mediante rectificación de iluminación

Anteriormente, se proponían unas posibles optimizaciones en los análisis de la *MAHNOB HCI* mediante la aplicación de una máscara más restrictiva y rectificación de iluminación. Tras comprobar que las estimaciones obtenidas con optimizaciones no presentaban mejoría, en esta ocasión se decide probar únicamente las tres primeras etapas de la rectificación de iluminación por las que se obtienen los resultados g_{IR} , g_{IR_1} y g_{IR_2} .

Además, se prescinden de los resultados globales ya que, para analizar los resultados en este caso tras dicha rectificación de iluminación, interesa ver el comportamiento de los algoritmos en cada tipo de luz de forma independiente.

La configuración empleada en esta prueba, es la misma que en el caso de la *MAHNOB HCI*, es decir, utilizando la máscara de rostro completo (*máscara 1*) y el algoritmo rPPG GVERK, para comparar de forma ecuánime los resultados tras esta optimización.

A continuación, se muestran los resultados obtenidos tras aplicar la rectificación de iluminación:

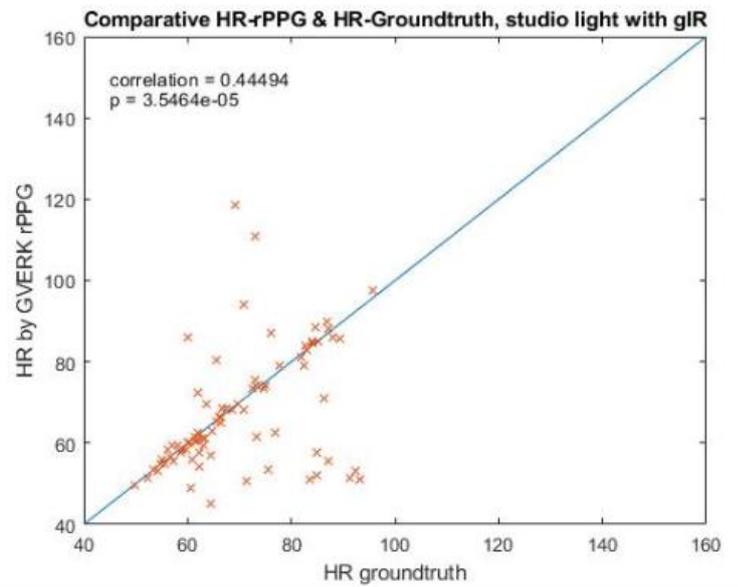
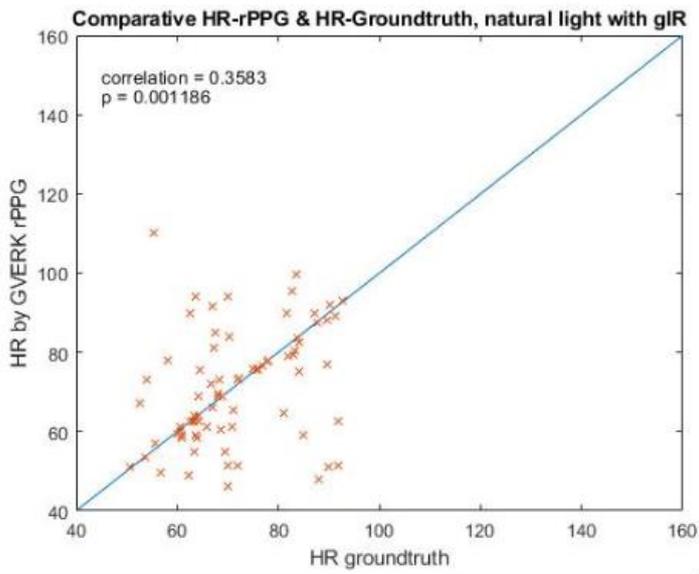


Fig. 3.24: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, g_IR.

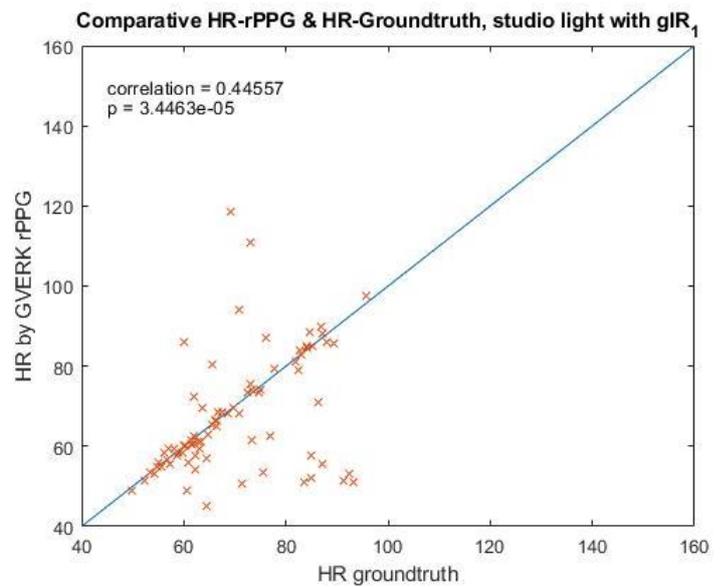
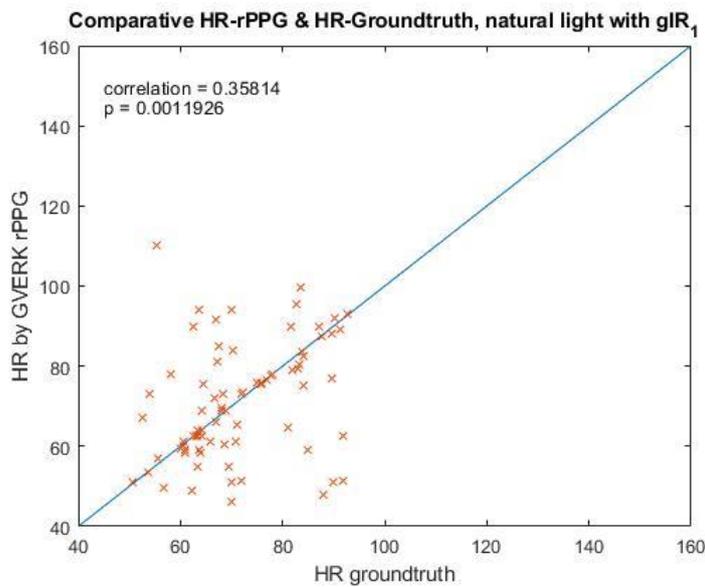


Fig. 3.25: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, g_IR_1.

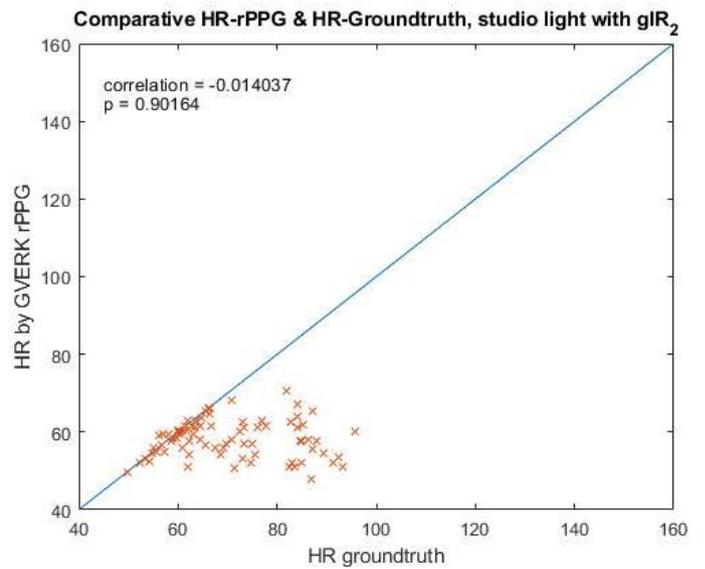
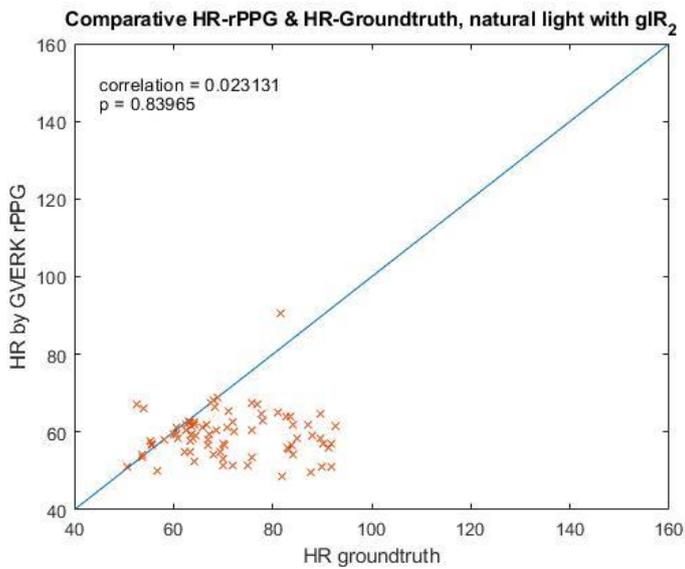


Fig. 3.26: Comparación entre FC groundtruth y estimada, iluminación natural vs estudio, g_{IR_2} .

Como se puede ver en las gráficas anteriores y se intuía de los resultados obtenidos en la *MAHNOB HCI*, las correlaciones obtenidas para g_{IR} , g_{IR_1} y g_{IR_2} son progresivamente peores, por lo que no existe una mejoría en la precisión de las estimaciones tras aplicar la rectificación de iluminación propuesta en el apdo. 2, pág. 39.

Apartado 4

Conclusiones y líneas futuras

Conclusiones

El empleo de la Fotopletismografía Remota (rPPG) como técnica principal para la estimación de la frecuencia cardíaca (FC) de un individuo a distancia, tras la captura de un vídeo de su rostro con una cámara de bajo coste es, a día de hoy, un reto a superar en los campos del procesado de imagen y señal. La infinidad de potenciales aplicaciones tanto biomédicas como sociales, despierta el interés a ser investigada. Como se ha podido ver en el repaso del estado del arte, hay numerosos estudios en este ámbito.

Para realizar un análisis de los métodos rPPG más relevantes en el modo *offline*, se emplean dos bases de datos públicas: *MAHNOB HCI* ^[1] y *COHFACE* ^[2]. Tras una selección previa de las sesiones óptimas para el análisis, se calcula la FC *groundtruth* de cada sesión a partir de los datos recogidos por los sensores fisiológicos. Por otra parte, se emplean los videos de cada sesión para estimar la FC mediante los algoritmos rPPG publicados por *Kooij-Naber* ^[71] y los incluidos en *iPhys-toolbox* ^[72]. Todo este proceso está estructurado en tres fases fundamentales: *recopilación de datos, análisis y procesado de señal y, por último, cálculo de resultados.*

Tanto en el modo *offline* como *online* se emplea el rastreador de puntos faciales iCCR-master, ya que mantiene una gran robustez y precisión incluso en aplicaciones a tiempo real. De los resultados se deduce que, cada algoritmo rPPG ofrece mejores o peores prestaciones en función de la ROI facial empleada y el tipo de iluminación presente en la escena de grabación.

Como propuesta de optimización, se emplea un filtrado espacial (máscaras) para la determinación de las ROI faciales y un filtrado para la rectificación de iluminación. A pesar de no alcanzar los resultados deseados, todo el trabajo de análisis realizado en las dos bases de datos públicas mencionadas, sirve como punto de partida para optimizaciones futuras de los métodos rPPG.

Líneas Futuras

Tras diseñar una interfaz que implementa el análisis rPPG a tiempo real (modo *online*), sería interesante realizar un estudio de cómo se comporta este análisis en función del tipo de luz empleada durante la grabación, para intentar catalogar con qué tipo de luz cada algoritmo ofrece los mejores resultados.

Bibliografía

- [1] A Multimodal Database for Affect Recognition and Implicit Tagging, <https://mahnobdb.eu/hci-tagging/>
- [2] <https://www.idiap.ch/dataset/cohface>
- [3] <https://fundaciondelcorazon.com/corazon-facil/recursos-didacticos/fichas.html?download=25:ficha-frecuencia-cardiaca>
- [4] <https://www.noldus.com/blog/what-is-rppg>
- [5] Video-Based Heart Rate Measurement. Recent Advances and Future Prospects. Xun Chen *et al.*
- [6] W. Verkruyse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Express*, vol. 16, no. 26, pp. 21434–21445, 2008
- [7] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [8] Y. Sun, V. Azorin-Peris, R. Kalawsky, S. Hu, C. Papin, and S. E. Greenwald, "Use of ambient light in remote photoplethysmographic systems: Comparison between a high-performance camera and a low-cost webcam," *J. Biomed. Opt.*, vol. 17, no. 3, p. 037005, 2012.
- [9] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Monitoring of cardiorespiratory signal: Principles of remote measurements and review of methods," *IEEE Access*, vol. 5, pp. 15776–15790, 2017.
- [10] D. J. McDuff, J. R. Estep, A. M. Piasecki, and E. B. Blackford, "A survey of remote optical photoplethysmographic imaging methods," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6398–6404.
- [11] J. Przybyło, E. Kańtoch, M. Jabłoński, and P. Augustyniak, "Distant measurement of plethysmographic signal in various lighting conditions using configurable frame-rate camera," *Metrol. Meas. Syst.*, vol. 23, no. 4, pp. 579–592, 2016.

- [12] S. A. Siddiqui, Y. Zhang, Z. Feng, and A. Kos, "A pulse rate estimation algorithm using PPG and smartphone camera," *J. Med. Syst.*, vol. 40, no. 5, p. 126, 2016.
- [13] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [14] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [15] A. Al-Naji and J. Chahl, "Simultaneous tracking of cardiorespiratory signals for multiple persons using a machine vision system with noise artifact removal," *IEEE J. Transl. Eng. Health Med.*, vol. 5, 2017, Art. no. 1900510.
- [16] A. Sikdar, S. K. Behera, and D. P. Dogra, "Computer-vision-guided human pulse rate estimation: A review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 91–105, 2016.
- [17] M. Hassan *et al.*, "Heart rate estimation using facial video: A review," *Biomed. Signal Process. Control*, vol. 38, pp. 346–360, Sep. 2017.
- [18] P. V. Rouast, M. T. P. Adam, R. Chiong, D. Cornforth, and E. Lux, "Remote heart rate measurement using low-cost RGB face video: A technical literature review," *Frontiers Comput. Sci.*, vol. 12, no. 5, pp. 858–872, 2016.
- [19] X. Li, J. Chen, G. Zhao, and M. Pietikäinen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 4264–4271.
- [20] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Express*, vol. 6, no. 5, pp. 1565–1588, 2015.
- [21] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1422–1433, Sep. 2017.
- [22] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Robust heart rate from fitness videos," *Physiol. Meas.*, vol. 38, no. 6, p. 1023, 2017.

- [23] F. Zhao, M. Li, Y. Qian, and J. Z. Tsien, "Remote measurements of heart and respiration rates for telemedicine," *PLoS ONE*, vol. 8, no. 10, p. e71384, 2013.
- [24] P. S. Addison, D. Jacquell, D. M. H. Foo, and U. R. Borg, "Video-based heart rate monitoring across a range of skin pigmentations during an acute hypoxic challenge," *J. Clin. Monitor. Comput.*, vol. 32, no. 5, pp. 871–880, 2017.
- [25] V. Jeanne, M. Asselman, B. den Brinker, and M. Bulut, "Camera-based heart rate monitoring in highly dynamic light conditions," in *Proc. Int. Conf. Connected Vehicles Expo (ICCVE)*, Dec. 2013, pp. 798–799.
- [26] S. Xu, L. Sun, and G. K. Rohde, "Robust efficient estimation of heart rate pulse from video," *Biomed. Opt. Express*, vol. 5, no. 4, pp. 1124–1135, 2014.
- [27] D.-Y. Chen *et al.*, "Image sensor-based heart rate evaluation from face reflectance using Hilbert–Huang transform," *IEEE Sensors J.*, vol. 15, no. 1, pp. 618–627, Jan. 2015.
- [28] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward, "Removing muscle artifacts from EEG data: Multichannel or single-channel techniques?" *IEEE Sensors J.*, vol. 16, no. 7, pp. 1986–1997, Apr. 2016.
- [29] K.-Y. Lin, D.-Y. Chen, and W.-J. Tsai, "Face-based heart rate signal decomposition and evaluation using multiple linear regression," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1351–1360, Mar. 2016.
- [30] A. Lam and Y. Kuno, "Robust heart rate measurement from video using select random patches," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3640–3648.
- [31] X. Li, J. Chen, G. Zhao, and M. Pietikäinen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 4264–4271.
- [32] D. Lee, J. Kim, S. Kwon, and K. Park, "Heart rate estimation from facial photoplethysmography during dynamic illuminance changes," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2758–2761.

- [33] J. Cheng, X. Chen, L. Xu, and Z. J. Wang, "Illumination variation resistant video-based heart rate measurement using joint blind source separation and ensemble empirical mode decomposition," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 5, pp. 1422–1433, Sep. 2017.
- [34] L. Xu, J. Cheng, and X. Chen, "Illumination variation interference suppression in remote PPG using PLS and MEMD," *Electron. Lett.*, vol. 53, no. 4, pp. 216–218, Feb. 2017.
- [35] X. Chen, X. Xu, A. Liu, M. J. Mckeown, and Z. J. Wang, "The use of multivariate EMD and CCA for denoising muscle artifacts from fewchannel EEG recordings," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 2, pp. 359–370, Feb. 2018.
- [36] J. Jenitta and A. Rajeswari, "Denoising of ECG signal based on improved adaptive filter with EMD and EEMD," in *Proc. IEEE Conf. Inf. Commun. Technol. (ICT)*, Apr. 2013, pp. 957–962.
- [37] X. Chen, Q. Chen, Y. Zhang, and Z. J. Wang, "A novel EEMDCCA approach to removing muscle artifacts for pervasive EEG," *IEEE Sensors J.*, to be published, doi: [10.1109/JSEN.2018.2872623](https://doi.org/10.1109/JSEN.2018.2872623).
- [38] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [39] G. Cennini, J. Arguel, K. Ak, sit, and A. van Leest, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Express*, vol. 18, no. 5, pp. 4867–4875, 2010.
- [40] F. Bousefsaf, C. Maaoui, and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 568–574, 2013.
- [41] A. V. Moço, S. Stuijk, and G. de Haan, "Motion robust PPG-imaging through color channel mapping," *Biomed. Opt. Express*, vol. 7, no. 5, pp. 1737–1754, 2016.
- [42] W. J. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Amplitudesselective filtering for remote-PPG," *Biomed. Opt. Express*, vol. 8, no. 3, pp. 1965–1980, 2017.

- [43] M. V. Gastel, S. Stuijk, and G. D. Haan, "Motion robust remote-PPG in infrared," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 5, pp. 1425–1433, May 2015.
- [44] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [45] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [46] X. Chen, Z. J. Wang, and M. McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Process. Mag.*, vol. 33, no. 3, pp. 86–107, May 2016.
- [47] [9] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," *Opt. Express*, vol. 18, no. 10, pp. 10762–10774, 2010.
- [48] S. Yu, S. Hu, V. Azorin-Peris, J. A. Chambers, Y. Zhu, and S. E. Greenwald, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, p. 077010, 2011
- [49] B. Wei, X. He, C. Zhang, and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions," *Biomed. Eng. Online*, vol. 16, no. 1, p. 17, 2017.
- [50] Y. C. Lin, N. K. Chou, G. Y. Lin, M. H. Li, and Y. H. Lin, "A real-time contactless pulse rate and motion status monitoring system based on complexion tracking," *Sensors*, vol. 17, no. 7, p. 1490, 2017.
- [51] A. Al-Naji, A. G. Perera, and J. Chahl, "Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle," *Biomed. Eng. Online*, vol. 16, Aug. 2017, Art. no. 101
- [52] B. Wei, X. He, C. Zhang, and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions," *Biomed. Eng. Online*, vol. 16, no. 1, p. 17, 2017.
- [53] Z. Guo, Z. J. Wang, and Z. Shen, "Physiological parameter monitoring of drivers based on video data and independent vector analysis,"

- in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 4374–4378.
- [54] H. Qi, Z. Guo, X. Chen, Z. Shen, and Z. J. Wang, “Video-based human heart rate measurement using joint blind source separation,” *Biomed. Signal Process. Control*, vol. 31, pp. 309–320, Jan. 2017.
- [55] G. de Haan and A. van Leest, “Improved motion robustness of remotePPG by using the blood volume pulse signature,” *Physiol. Meas.*, vol. 35, no. 9, pp. 1913–1926, 2014.
- [56] R.-Y. Huang and L.-R. Dung, “A motion-robust contactless photoplethysmography using chrominance and adaptive filtering,” in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Oct. 2015, pp. 1–4.
- [57] W. Wang, S. Stuijk, and G. de Haan, “Exploiting spatial redundancy of image sensor for motion robust rPPG,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 415–425, Feb. 2015
- [58] <https://www.tobii.com/>
- [59] <http://www.biosemi.com>
- [60] https://docs.google.com/document/preview?hgd=1&id=1Wxa7u0gtGcF_A-ayon7RUWa0ml0yvgcFW05APOE9vD8
- [61] http://ocw.uv.es/ingenieria-y-arquitectura/1-5/ib_material/IB_T6_OCW.pdf
- [62] Cascaded Continuous Regression for Real-time Incremental Face Tracking, Enrique Sánchez-Lozano, Brais Martinez, Georgios Tzimiropoulos, and Michel Valstar
- [63] <https://www.votchallenge.net/>
- [64] <https://ieeexplore.ieee.org/abstract/document/979357>
- [65] <https://medium.com/blueeast/how-to-use-moving-average-filter-to-counter-noisy-datasignal-5b530294a12e>
- [66] <https://www.osapublishing.org/oe/viewmedia.cfm?uri=oe-16-26-21434&seq=0>
- [67] https://pure.tue.nl/ws/files/31563684/TBME_00467_2016_R1_preprint.pdf
- [68] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.726.6643&rep=rep1&type=pdf>
- [69] <https://www.osapublishing.org/oe/viewmedia.cfm?uri=oe-18-10-10762&seq=0>

- [70] <https://es.mathworks.com/products/matlab.html>
- [71] van der Kooij KM, Naber M. An open-source remote heart rate imaging method with practical apparatus and algorithms. Behav Res Methods. 2019 Oct;51(5):2106-2119. doi: 10.3758/s13428-019-01256-8. PMID: 31152386; PMCID: PMC6797647.
- [72] <https://github.com/danmcduff/iphys-toolbox>
- [73] Libro de la Salud Cardiovascular del Hospital Clínico San Carlos y la Fundación BBVA.https://www.fbbva.es/microsites/salud_cardio/fbbva_libroCorazon_Cap_4.html?gOrri=6
- [74] Medida remota del ritmo cardíaco a partir de videos grabados con cámaras de bajo coste, Marta Olaizola Gómez y Rafael Cabeza Laguna.