

UNIVERSIDAD DE GUANAJUATO

---

CAMPUS IRAPUATO – SALAMANCA  
DIVISIÓN DE INGENIERÍAS

*Detección de Ritmo Cardíaco a Partir del  
Análisis de Secuencias de Imágenes*

**TESIS**

QUE PARA OBTENER EL GRADO DE:

**MAESTRA EN INGENIERÍA ELÉCTRICA**

Opción: Instrumentación y Sistemas Digitales

PRESENTA:

*Ing. Gesley Geyn Tarazona Guerrero*

DIRECTORES:

*Dr. Raúl Enrique Sánchez Yáñez*

*Dr. Jonathan Cepeda Negrete*

Universidad de Guanajuato

# Dedicatoria

*A mis padres, Teresa y Neftalí.*

*A mis sobrinos, David y Santiago.*

# Agradecimientos

Quisiera agradecer especialmente a mis padres, Teresa y Neftalí. Su apoyo se siente incluso en la distancia. A mis hermanas Patsy, Sharyn y a mi hermano Pat. Ellos son los pilares de mi vida y la razones principales de mi esfuerzo y compromiso.

A Cristhiam y su familia, en cabeza de Doña Ceci, quienes han sido parte importante en mi desarrollo personal y profesional. Gracias por su apoyo incondicional en esta aventura.

A mis asesores en este camino, el Dr. Raúl Sánchez y el Dr. Jonathan Cepeda por su guía profesional y el tiempo invertido en la realización de este trabajo. Gracias por recibirme y ayudarme en el proceso.

Al Dr. Victor Ayala, por su apoyo y el tiempo compartido durante mi estancia en el laboratorio.

Al Dr. Uriel, Dra. Rocío, Dr. Fernando y Dr. Juan Pablo, por las pequeñas pláticas y los momentos compartidos.

A mis amigos del laboratorio, José Jesús, Diana, Alberto, Leonardo, Marco, Eduardo, Abrahan, Javi, Ana Karen, Ricardo, Lubin, Andrea, Alán y Juan Pablo, por permitirme un espacio en su día a día y hacer de esta experiencia la más enriquecedora.

A mis amigos Dani, George, Jesús, Marve, Manuel, Felipe, Emmanuel y Juan Luis, a quienes conocí durante este proceso. Gracias por el tiempo que compartimos y los momentos vividos.

A todas las personas en Colombia que han estado siempre a lo largo de mi vida, gracias por su apoyo.

# Agradecimientos Institucionales

Expreso mi más sincera gratitud hacia la Universidad de Guanajuato, especialmente a la División de Ingenierías del Campus Irapuato-Salamanca por la formación y el apoyo que he recibido. A todos los profesores, mis agradecimientos por su duro trabajo durante estos dos años que me permitieron obtener el grado de Maestro en Ingeniería Eléctrica.



Este trabajo fue realizado gracias al apoyo recibido por el Consejo Nacional de Ciencia y Tecnología de México, CONACYT, bajo la beca otorgada en la convocatoria número 291025 titulada "CONVOCATORIA DE BECAS NACIONALES 2015 SEGUNDO PERIODO" con el número (CVU/Becario) 718796/592285.

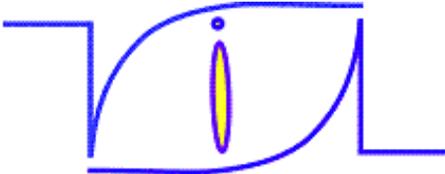


**CONACYT**

*Consejo Nacional de Ciencia y Tecnología*

---

Al Laboratorio de Visión, Robótica e Inteligencia Artificial (LaViRIA) por la formación académica recibida y el trabajo realizado durante mi estancia en este laboratorio.



Perception, cognition and  
robotics synergy

# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Ciclo y Ritmo Cardiaco . . . . .	2
1.2. Medición de la Frecuencia Cardiacas . . . . .	2
1.2.1. Medición por Electrocardiografía . . . . .	3
1.2.2. Medición por Pletismografía . . . . .	5
1.2.3. Medición por Fotopletismografía . . . . .	6
1.3. Planteamiento del Problema . . . . .	7
1.4. Estado del arte . . . . .	8
1.5. Justificación . . . . .	10
1.6. Objetivos . . . . .	11
1.6.1. Objetivo General . . . . .	11
1.6.2. Objetivos Particulares . . . . .	11
1.7. Estructura del trabajo . . . . .	11
<b>2. Estimación del Ritmo Cardiacas</b>	<b>13</b>
2.1. Descripción del Sistema Propuesto . . . . .	13

2.2. Extracción de Cuadros desde una Secuencia de Video . . . . .	15
2.3. Segmentación . . . . .	15
2.4. Detección de Piel . . . . .	16
2.4.1. Representaciones de Color . . . . .	17
2.4.2. Delimitación explícita de la Región de Piel . . . . .	26
2.5. Cálculo del Color Promedio . . . . .	32
2.6. Proceso de Filtrado y Amplificación . . . . .	35
2.6.1. Eliminación de la Tendencia no Lineal . . . . .	35
2.6.2. Filtro Pasa Banda . . . . .	37
2.6.3. Amplificación de la Señal Filtrada . . . . .	39
2.7. Puntos Máximos y Análisis Espectral FFT . . . . .	40
<b>3. Evaluación del Sistema</b>	<b>42</b>
3.1. Protocolo de Pruebas . . . . .	43
3.2. Pruebas y sus Resultados . . . . .	44
<b>Conclusiones</b>	<b>51</b>

# Índice de figuras

1.1. Señal proveniente de un electrocardiograma. . . . .	4
1.2. Forma original de una onda PPG de pulso. . . . .	6
2.1. Marco de trabajo propuesto para la Estimación de Ritmo Cardíaco sin contacto. . . . .	14
2.2. Diagrama de bloques de la Extracción de imágenes. . . . .	15
2.3. Representación general del módulo de segmentación de piel. . . . .	17
2.4. Sistema de coordenadas cartesianas tridimensional para <i>RGB</i> . . . . .	18
2.5. Imagen original y las componentes individuales del espacio de color <i>RGB</i> . . . . .	19
2.6. Representación cónica del modelo <i>HSV</i> . . . . .	20
2.7. Imagen original y cada componente del espacio <i>HLS</i> en escala de grises. . . . .	21
2.8. Representación gráfica de la componente de color <i>H</i> . . . . .	23
2.9. Representación gráfica de la componente de color del valor de luminosidad. . . . .	24
2.10. Representación gráfica de la componente de color Cromo. . . . .	25
2.11. Representación geométrica del modelo Munsell. . . . .	26

2.12. Distribución espacial de las muestras de color piel. . . . .	27
2.13. Conjunto de imágenes muestra con su salida segmentada. . . . .	30
2.14. Descripción general del módulo del color promedio. . . . .	33
2.15. Señales de pulso extraídas del valor promedio de las intensidades de los píxeles dentro de la región de interés. . . . .	34
2.16. Señal pura con tendencia no lineal. . . . .	35
2.17. Señal filtrada, sin tendencia no lineal. . . . .	37
2.18. Señal filtrada en el rango de interés. . . . .	38
2.19. Espectro de potencia de la señal PPG. . . . .	41
3.1. Resultados del video <i>face1</i> . El ritmo cardiaco de referencia es de 54 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR. . . .	45
3.2. Resultados del video <i>face2</i> . El ritmo cardiaco de referencia es de 55 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR. . . .	46
3.3. Resultados del video <i>baby</i> . El ritmo cardiaco de referencia es de 152 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR. . . .	47

- 3.4. Resultados del video *Bush*. El ritmo cardiaco de referencia es de 60 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR. . . . 48
- 3.5. Resultados del video *Gesley*. El ritmo cardiaco de referencia es de 80 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR. . . . 49

# Índice de tablas

2.1. Resultado del promedio de los valores obtenidos para el grupo muestra de 6 imágenes. . . . .	29
2.2. Resultado del promedio de los valores obtenidos para toda la base de datos. . . . .	32
3.1. Estimación de ritmo cardiaco por número de picos detectados y su porcentaje de error. . . . .	50
3.2. Estimación de ritmo cardiaco mediante la FFT y su porcentaje de error. . . . .	50

# Resumen

En el presente proyecto, se propone el desarrollo de un ambiente computacional para la detección del ritmo cardiaco, mediante el análisis de secuencias de imágenes extraídas de un video (cuadros). En este caso, se utilizarán tomas de video enfocadas directamente a la cara de un sujeto. En cada imagen capturada en un instante de tiempo, se deberá analizar la piel de la persona con el objetivo de detectar pequeñas variaciones en las características de color dentro de la secuencia de video. A partir de estas variaciones y mediante un filtrado digital, deberá deducirse la información de la señal de pulso cardiovascular del sujeto en estudio. Así, se plantea un sistema de monitoreo del ritmo cardiaco, que funciona de manera no invasiva y posiblemente sin intervención alguna por parte del sujeto de estudio, quien pudiera ser un paciente o alguien que necesite estar bajo algún tipo de observación.

# Capítulo 1

## Introducción

Como una introducción parcial a la solución de la estimación de ritmo cardiaco sin contacto (HR, por sus siglas en inglés), es indispensable plantear algunos interrogantes fundamentales que ofrezcan un panorama más intuitivo al respecto. En primera instancia se hace necesario definir ¿qué es el HR? y además comprender ¿por qué es posible medir el HR a partir del análisis de secuencias de imágenes? De manera general, es posible dar una noción parcial a estos interrogantes, entendiéndose que el ritmo cardiaco es un indicador fundamental de la salud física y emocional en los seres humanos, lo que resalta inicialmente la importancia que merece el presente estudio. Adicionalmente, ya que los latidos del corazón generan cambios en el tono de la piel, producidos por las variaciones en el volumen de los tejidos, se pueden utilizar las técnicas de procesamiento y análisis de imágenes para evidenciar dicho fenómeno. Para dar una respuesta detallada a cada duda, se describe en adelante cada módulo que compone el presente trabajo de tesis.

En el presente capítulo se exponen los conceptos básicos a explorar para la introducción al tema de estudio. Además, se realiza una revisión bibliográfica que trata de los métodos actuales de estimación del HR así como el planteamiento del problema junto con la justificación y los objetivos propuestos para su solución.

## 1.1. Ciclo y Ritmo Cardíaco

El ciclo cardíaco se conoce como el movimiento periódico de la sangre desde el corazón a la cabeza pasando por la aorta abdominal y las arterias carótidas [Balakrishnan et al., 2013]. Durante un ciclo cardíaco hay una variedad de eventos de distinta naturaleza: eléctricos, mecánicos, sonoros y de presión, directamente relacionados con el flujo de sangre a través de las cavidades cardíacas, su contracción y relajación (aurículas y ventrículos), así como el cierre y apertura de las válvulas y la producción de ruidos asociados a ellas. El valor recíproco de la duración de un ciclo es la frecuencia cardíaca, también conocida como ritmo cardíaco, la cual se suele expresar en latidos por minuto (bpm, beats per minute). Dado que un ciclo cardíaco transcurre en menos de un segundo, se multiplica por 60 para obtener el número de bpm, la unidad estándar para medir la frecuencia cardíaca.

El HR es un indicador importante de la salud física y emocional de las personas [Li et al., 2014]. Corresponde a la magnitud de medida para una de las funciones más básicas y vitales del cuerpo humano que puede entregar información muy significativa para conocer su estado actual e incluso detectar y monitorear problemas de salud. El seguimiento del HR permite conocer qué tanto está trabajando el corazón para detectar posibles anomalías o determinar rangos permisibles de actividad física que aporten el máximo rendimiento posible del cuerpo sin arriesgar la integridad física de la persona. De hecho, el análisis de señales de HR no es de uso exclusivo para el monitoreo de personas enfermas, sino también, por ejemplo, en el caso de atletas de alto rendimiento, deportistas o incluso para personas que realizan algún tipo de ejercicio.

## 1.2. Medición de la Frecuencia Cardíaca

La frecuencia cardíaca se refiere al número de veces que late un corazón humano por unidad de tiempo y está ligado directamente al bombeo de sangre a través del sistema circulatorio para distribuir el oxígeno absorbido en los procesos respiratorios y excretar dióxido de carbono residual. Normalmente, la medición de la frecuencia cardíaca, se hace a partir del pulso, el cual es un indicador tangible de las contracciones musculares del corazón (latidos).

El rango de frecuencia cardiaca a la que se presentan los latidos de un corazón normal está aproximadamente entre 50 y 100 bpm en personas adultas, aumentando hasta el doble si el paciente es un niño o un recién nacido. Sin embargo, este valor puede fluctuar dependiendo del metabolismo de la persona, su complejión, su estado físico y su edad. Adicionalmente, actividades físicas como el reposo (sueño) o el ejercicio, estados emocionales como la excitación o la angustia, e inclusive medicamentos y drogas recreacionales pueden afectar los niveles de frecuencia cardiaca.

A partir del rango definido anteriormente pueden presentarse varios tipos de condiciones anómalas que son de especial cuidado para mantener la salud de la persona bajo observación:

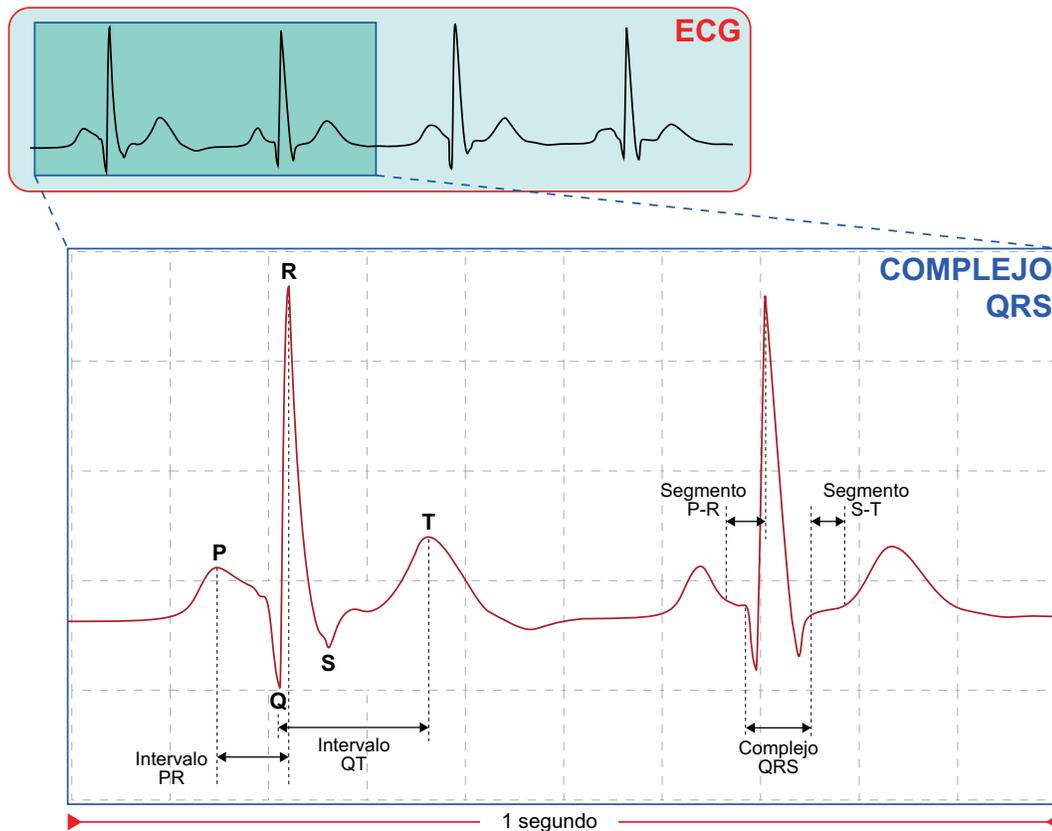
- Bradicardia: Frecuencias cardiacas inferiores a los 50 bpm en estado de reposo.
- Taquicardia: Frecuencias cardiacas superiores a los 100 bpm en estado de reposo.
- Arritmia: Falta de sincronía y periodicidad en el pulso cardiaco.

Para el trabajo desarrollado en este proyecto, es conveniente abordar ciertas alternativas previas de medición, que orientan de manera general el presente estudio.

### **1.2.1. Medición por Electrocardiofía**

La medición del HR es realizada generalmente a través de la detección de pequeños impulsos eléctricos que se crean debido a las llamadas células de marcapasos o células automáticas, localizadas en el nodo sinoauricular del corazón. Éstas producen la actividad eléctrica que se extiende al resto del músculo cardiaco a un ritmo regular, conocida como despolarización espontánea, por lo que se generan variaciones continuas a través del tiempo lo que hace posible registrar los latidos cardiacos. Este registro de la actividad eléctrica se conoce como electrocardiografía (ECG) y se puede ver representado mediante los ya conocidos electrocardiogramas.

El ECG muestra gráficamente una señal que es captada en la mayoría de los casos por electrodos de registro, puestos estratégicamente en la superficie de la piel. Los cambios eléctricos detectados son muy pequeños pero suficientes para obtener una curva del tipo QRS más conocida como complejo QRS, el cual aparece después de la onda P (la primera del ciclo cardiaco). Es llamado así debido a que lo conforman el conjunto de ondas que representan la despolarización de los ventrículos. Un ejemplo de la onda ECG es mostrado en la Figura 1.1.



**Figura 1.1:** Señal proveniente de un electrocardiograma.

La configuración de esta señal eléctrica registra una primera onda P que representa la activación auricular. Seguido de ésta, se encuentra la onda Q generalmente negativa, que indica el comienzo de la activación ventricular y a continuación la onda ascendente positiva R, terminando rápidamente en una onda descendente negativa S. El cambio brusco que es detectado en el punto más alto de la onda R se denomina deflexión intrínseca o intrinsecoide [Farreras and Rozman, 2000]. Desde el punto de vista eléctrico, existen 3 fases del ciclo cardiaco: la primera es la despolarización que tiene lugar en el miocardio, en el que se produce un cambio

brusco de potencial que da paso a la excitación de las células vecinas y de forma secuencial a todo el tejido muscular del corazón. La segunda es la repolarización o activación ventricular, observable en la región del segmento ST junto con la onda T y por último se produce el estado de reposo hasta el inicio de un nuevo ciclo. El intervalo RR inicia en la cima de una onda R y termina justo en el pico de la siguiente onda R.

### 1.2.2. Medición por Pletismografía

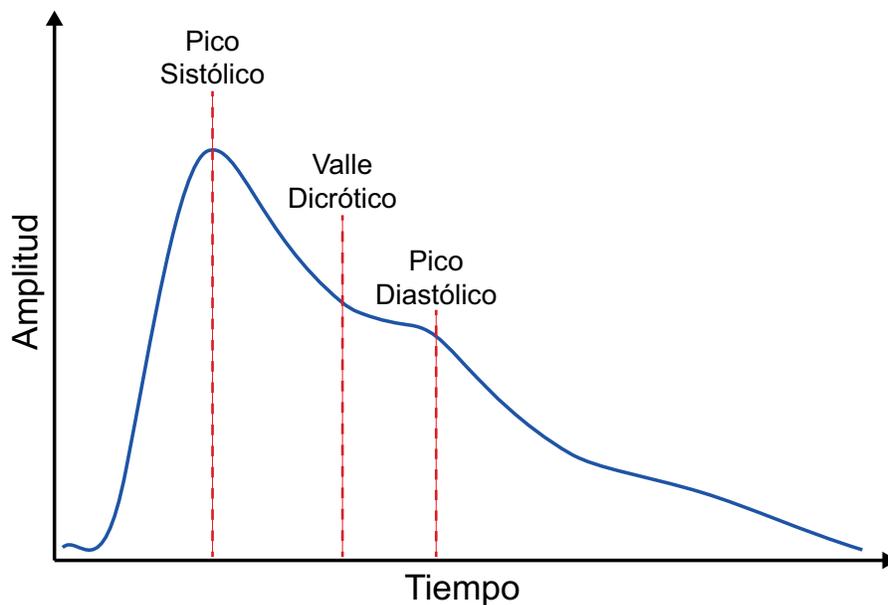
La pletismografía es un método usado como herramienta para determinar los cambios de presión y volumen causados por la sangre o el aire (en el caso de pletismografía pulmonar) que fluye por el cuerpo [Hertzman and Dillon, 1940]. Normalmente, los pletismógrafos se colocan en el perímetro alrededor de las extremidades del cuerpo, ya que presentan un flujo sanguíneo de fácil detección. Las señales que se producen con las variaciones en el volumen de un órgano o en el sistema arterial así como otros episodios naturales de expansión corporal, se relacionan directamente con la presión arterial [Weinman et al., 1977], lo que permite identificar problemas que afecten la circulación sanguínea. Encontrar y cuantificar las pequeñas diferencias en los parámetros de cambio desencadena en un gran número de aplicaciones para el estudio y diagnóstico de enfermedades principalmente cardiovasculares y pulmonares.

Para el caso de la pletismografía aplicada al flujo sanguíneo, existen dos categorías principales:

- Arterial: Puede ser utilizada para asistir en el diagnóstico de bloqueo arterial por colesterol u otros problemas de escala más pequeña en arterias menores que irrigan pies y manos. Así mismo, puede facilitar la identificación de cualquier tipo de condición que se caracterice por una constricción anormal de los vasos sanguíneos en las extremidades mayores.
- Venosa: Ayuda a identificar problemas de flujo sanguíneo a través de las venas en zonas de irrigación menor, ubicando posibles coágulos u oclusiones en el sistema circulatorio. También puede asistir en la detección de disfunciones en las válvulas del sistema venoso que pueden resultar en circulación retrograda (flujo en sentido contrario) por fallas vasculares.

### 1.2.3. Medición por Fotopleletismografía

Para contestar la pregunta ¿por qué es posible medir el HR a partir del análisis de secuencias de imágenes? debemos considerar el concepto de fotopleletismografía (PPG como abreviación de Photoplethysmography), el cual hace más clara la relación entre visión por computadora y el estudio directo de las señales que proporciona el cuerpo para su estudio y comprensión. La PPG es una técnica de bajo costo, generalmente usada para obtener señales de origen óptico, transmitidas o reflejadas en la superficie de la piel, de manera no invasiva [Allen, 2007]. Sin embargo, su uso no se limita únicamente a la superficie cutánea sino que también puede ser aplicada a órganos para evaluar su estado. Esto se debe a que dicha técnica detecta variaciones en el volumen de sangre en la cama microvascular del tejido mediante cambios en la absorción de la luz [Rhee et al., 2001, Humphreys et al., 2007]. Una onda fotopleletismográfica típica se muestra en la Figura 1.2, en donde se puede observar un pico de amplitud correspondiente a un ciclo cardiaco.



**Figura 1.2:** Forma original de una onda PPG de pulso.

A pesar de la atenuación natural en la presión del latido del corazón hasta llegar a la superficie cutánea, la piel está altamente irrigada por medio de vasos sanguíneos, por lo que resulta posible detectar el componente pulsátil del ciclo cardiaco

mediante el análisis de imágenes. En la onda fotopleletismográfica, el componente DC corresponde a la absorción neta por parte del tejido cutáneo, mientras la AC se puede atribuir a la variación en el volumen de sangre irrigando la piel, causado por la presión del pulso en el ciclo cardiaco (la diferencia entre las presiones sistólicas y diastólicas en las arterias [Kurylyak et al., 2013]). De esta manera, un análisis fotopleletismográfico no solo puede dar información directamente relacionada con la frecuencia cardiaca, sino también con la presión sanguínea.

Una vez se hace tangible la estimación del HR libre de contacto, se da lugar a las principales aplicaciones que inicialmente proponen el análisis del comportamiento humano, así como el cuidado remoto y atención primaria en entornos clínicos o de seguimiento a grupos particulares, tales como los deportistas de alto rendimiento, entre otros.

### 1.3. Planteamiento del Problema

El ritmo cardiaco es uno de los indicadores más importantes a través del cual es posible inferir el estado de salud de una persona. Estimar el ritmo cardiaco a tiempo, no solo previene de complicaciones de salud mayores, sino que también proporciona una idea general del estado físico y emocional de una persona [Zeng et al., 2015]. La mayoría de los dispositivos usados comúnmente requieren contacto con el paciente, ya sea para un electrocardiograma o por detección fotoeléctrica de la punta del dedo, entre otros, los cuales son capaces de determinar las pulsaciones por minuto (bpm) de un corazón en estudio.

Recientemente, se han desarrollado métodos para realizar mediciones eficientes sin tener contacto con el paciente, es decir, estimaciones a partir de información proveniente de video o lo que es lo mismo, secuencias de imágenes [Yu et al., 2015]. Los equipos actuales, requieren posturas especiales y contacto directo que puede incomodar e incluso causar dolor al paciente [Li et al., 2014]. Las observaciones captadas por la cámara son imperceptibles a simple vista y están directamente relacionadas con la frecuencia cardiaca [Poh et al., 2010a] [Kumar et al., 2015]. Existen cambios que van desde el movimiento periódico de la cabeza causado por los latidos del corazón [Balakrishnan et al., 2013] hasta las variaciones de color presentes en la piel del rostro [Wu et al., 2012]. Estas últimas, serán registradas y analizadas para obtener datos que permitan elaborar una estimación adecuada

del ritmo cardiaco del sujeto en estudio.

Lo que se propone, es dar solución a la necesidad de mantener contacto directo con el sujeto de estudio. Es decir, obtener de forma precisa una medida de ritmo cardiaco que no requiera contacto directo y para lo cual sea suficiente con un par de segundos de video del sujeto en observación. Así, debido a que los latidos del corazón producen variaciones en el volumen de los tejidos microvasculares generando cambios en el tono de la piel, es posible aplicar las herramientas del procesamiento de imágenes para encontrar dichas variaciones no captadas a simple vista por el ojo humano, premisa fundamental para el desarrollo del trabajo.

Los estudios en los que se correlacionan microcambios en el color de la piel con la frecuencia cardiaca [Poh et al., 2010b] presentan la idea crucial para el planteamiento de este proyecto, que comprende el análisis de cuadros de vídeo de manera individual para la extracción de características asociadas a las zonas de piel, y luego un análisis de la evolución temporal de tales características, obteniéndose así la detección del ritmo cardiaco a partir de un vídeo.

## 1.4. Estado del arte

La posibilidad de obtener el ritmo cardiaco a través del análisis de la información registrada mediante videos, extiende favorablemente los alcances de la visión por computadora. Su uso para la detección de esta característica fisiológica propone una solución a dicho problema.

En consecuencia, existen diversos trabajos en el área que han desarrollado algunas metodologías para tal fin. Dentro de los primeros, se encuentran Poh *et al.* [Poh et al., 2010a], quienes estimaron el ritmo cardiaco a partir de videos utilizando el promedio de los valores de los píxeles en el espacio de color RGB de una región de interés definida a través del algoritmo de Viola-Jones [Viola and Jones, 2004]. De este modo, las señales unidimensionales extraídas son examinadas a partir del Análisis Independiente de Componentes (ICA por sus siglas en inglés). Finalmente, utilizan la Transformada Rápida de Fourier (FFT por sus siglas en inglés) para obtener el espectro de potencia de la señal escogida y así encontrar el valor de la frecuencia cardiaca incluso de tres personas a la vez.

Por otra parte, Wu *et al.* [Wu et al., 2012] desarrollaron una técnica que aplica descomposición espacial a una entrada de video estándar, seguido de un filtrado temporal. Una vez que se tienen las bandas espaciales filtradas, éstas son amplificadas usando un factor  $\alpha$  para así continuar con un proceso de reconstrucción del video de salida. Es posible entonces, utilizar el método para amplificar las variaciones de color presentes en la piel del rostro y estimar el ritmo cardiaco de una persona. Las condiciones en las que fueron tomados los videos para tal fin, son notablemente las de un ambiente controlado.

Otro trabajo que estima ritmo cardiaco sin contacto lo proponen Balakrishnan *et al.* [Balakrishnan et al., 2013]. El sistema hace uso del fenómeno de acción y reacción que se produce con cada latido del corazón. Se analizan los picos producidos por el movimiento, casi imperceptible, de la cabeza al fluir la sangre y se aplica el Análisis Principal de Componentes (PCA por sus siglas en inglés) para distinguir, entre otras, la señal que contiene la información frecuencial necesaria para la estimación. Su desventaja más grande es requerir que la persona permanezca lo más quieta posible durante la prueba.

Dentro de las más destacadas, se encuentra el marco de trabajo presentado por Li *et al.* [Li et al., 2014]. Ellos usan un conjunto de algoritmos ya conocidos teniendo como objetivo principal la disminución de errores causados por las variaciones de la luz en el ambiente y el movimiento del sujeto en estudio. El método fue diseñado para el monitoreo continuo de al menos nueve minutos para un escenario de prueba de video juegos. El error promedio alcanzado fue de 6.87% superando a métodos como el de Poh *et al.* y Balakrishnan *et al.* con un 13.2% y 20.7% respectivamente. La desventaja principal se centra en la necesidad de realizar cortes a las tramas de video que contienen movimientos no rígidos para la eliminación del ruido y en el uso del fondo como forma de rectificar los cambios de iluminación en la piel.

Una técnica que ha sido diseñada con el propósito de hacerla robusta a los cambios de iluminación y movimiento, es la expuesta por Lam *et al.* [Lam and Kuno, 2016]. Ellos extraen pares de puntos, sobre la región de interés, que cumplen un conjunto de condiciones previamente definidas. El problema es resuelto entonces, usando la Separación Ciega de Fuentes (BSS por sus siglas en inglés) para extraer la señal, que ha sido construida a partir de estimaciones locales del ritmo cardiaco en distintas regiones. Estos resultados parciales se combinan en un histograma, generando un esquema que contiene la información de frecuencia cardiaca que

permite una mejor aproximación. La implementación actual de su algoritmo tarda alrededor de 7 minutos para procesar un video de 30 segundos. Ésta es considerada la principal desventaja de la metodología propuesta.

De igual manera, se han desarrollado métodos para la detección del ritmo cardiaco de forma remota, por ejemplo mediante el análisis del espectro infrarrojo [PHILIPS, 2011] o de imágenes térmicas [Garbey et al., 2007]. Buscándose también bajar los costos de métodos sin contacto [Bernacchia et al., 2014], se han explorado particularmente métodos basados en vídeo estándar [Poh et al., 2011, Kwon et al., 2012]. Estos métodos basados en el análisis de vídeo inclusive pueden ser implementados para aplicaciones en tiempo real [Hyun et al., 2015].

## 1.5. Justificación

El ritmo cardiaco es típicamente medido a través de señales eléctricas. En la mayoría de los dispositivos, un electrocardiograma registra una forma de onda que incluye la información presente en los latidos del corazón. Estos métodos exigen en todo caso, contacto directo y en ocasiones invasivo, lo que podría causar incomodidades en el individuo.

Recientemente, se han venido explorando alternativas de mediciones libres de contacto, abriendo extensas posibilidades de aplicabilidad no solo en áreas relacionadas con la medicina, sino también en lo referente al análisis del comportamiento humano [Li et al., 2014], así como en el mejoramiento progresivo de la calidad de vida en las personas y en general, un estudio continuado de la actividad cardiaca permitiría coleccionar a largo plazo, los datos suficientes para prevenir y/o predecir afecciones importantes en el ser humano. Aunque, la detección del ritmo cardiaco es una tarea imprescindible en medicina para la atención de los pacientes, otro aspecto a tener en cuenta es que el ritmo cardiaco puede también ser signo de nerviosismo, o de sentir algún otro tipo de emoción, por lo que se amplían las posibilidades de aplicación para un detector de variación del ritmo cardiaco.

En este proyecto, se presenta una metodología sencilla y de bajo costo para medir ritmo cardiaco. El enfoque fundamental se da a la aplicación de modelos de color en los que se busca detectar pequeños cambios presentes en el tono de la piel y

así obtener una señal que contenga la información de frecuencia cardiaca a la que se producen los latidos del corazón.

## 1.6. Objetivos

### 1.6.1. Objetivo General

Desarrollar un ambiente computacional para la detección de pequeños cambios en la tonalidad de la piel y posteriormente analizar las variaciones temporales de estos cambios, infiriendo así el ritmo cardiaco de un sujeto bajo observación.

APLICABILIDAD DE LOS RESULTADOS. Los resultados de la investigación pudieran ser aplicados en distintas áreas, principalmente para el denominado análisis de información biomédica. El monitoreo no invasivo y, en ocasiones inadvertido, abre posibilidades de aplicación en sistemas diversos, por ejemplo, en aquellos de video vigilancia.

### 1.6.2. Objetivos Particulares

- La aplicación del análisis de imágenes en una tarea de relevante importancia en la medicina.
- Aplicación de modelos de color para la detección de minúsculas variaciones en el tono de piel.
- Integración de un entorno de desarrollo para el análisis de cuadros de vídeo individuales, y posteriormente el análisis de sus cambios en un eje temporal, hasta la determinación del ritmo cardiaco de un sujeto.

## 1.7. Estructura del trabajo

En este capítulo se dio brevemente una introducción al tópico de estudio. Los siguientes capítulos del trabajo de tesis se están organizados como sigue:

Capítulo 2. En este capítulo se presenta la metodología propuesta dentro de un marco de trabajo aplicado para el alcance de los objetivos. Se describen a detalle cada uno de los subsistemas que conforman el trabajo y se da el sustento teórico necesario para el análisis de los resultados parciales para cada módulo.

Capítulo 3. Aquí se describe el protocolo de pruebas diseñado para verificar y entender lo que se desea obtener. Se muestran los resultados alcanzados y su interpretación a partir de los experimentos realizados y la comparación de éstos con trabajos relacionados.

Por último, se exponen las conclusiones derivadas del trabajo, así como el trabajos futuros que puedan realizarse a partir de los resultados obtenidos y el estudio del presente proyecto.

# Capítulo 2

## Estimación del Ritmo Cardíaco

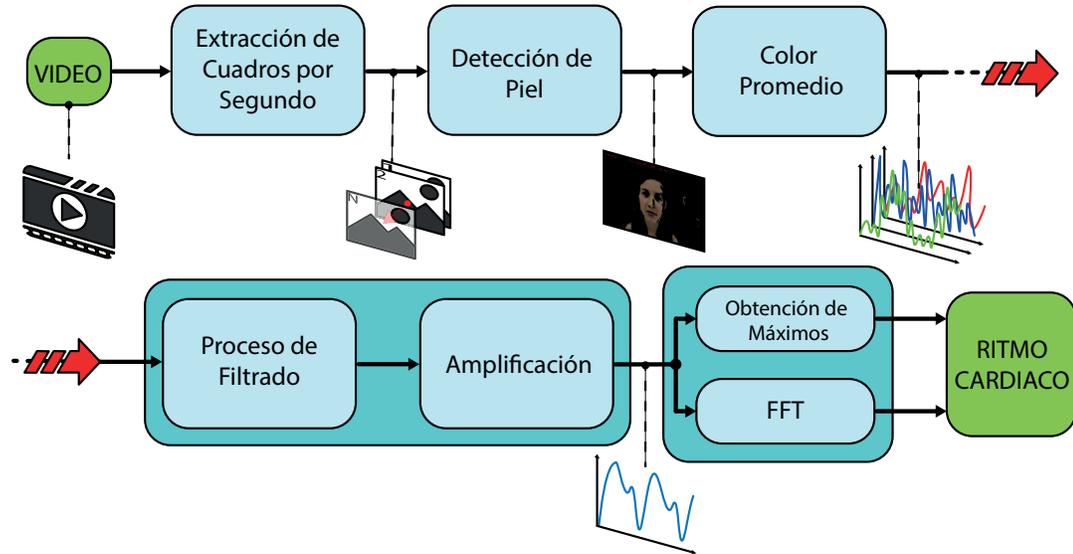
Uno de los retos más grandes que enfrenta la estimación de ritmo cardíaco sin contacto, es el diseño de una metodología robusta que permita extraer y analizar apropiadamente las características de interés en la imagen. En los sistemas que involucran la estimación libre de contacto, el objetivo primordial es encontrar el rango de valores en los que se encuentra la señal Fotopleletismográfica. Debido a que esta búsqueda se hace sobre imágenes estándar, capturadas por una cámara ordinaria y no sobre dispositivos con características particulares [Zeng et al., 2015], la detección de variaciones sutiles se hace imprescindible para su determinación. Es por eso que este trabajo se centra especialmente en los procesos de detección de piel y promedio de color, desarrollando la aplicación en tres diferentes espacios de color. De esta manera, la entrada está dada por un video de la cara de un sujeto, tomado con una cámara de bajo costo, del que se espera obtener un valor aproximado de ritmo cardíaco. En este capítulo se describe a detalle la metodología del sistema propuesto, a partir de los subsistemas o módulos que lo componen.

### 2.1. Descripción del Sistema Propuesto

Nuestro marco de trabajo está conformado por 5 módulos, como se observa en la Figura 2.1. El primer paso, es realizar la extracción de la secuencia de imágenes que componen un video de prueba. Una vez que se tiene el conjunto de cuadros

que se desean analizar, el propósito del segundo paso es obtener una región de interés. Para ello, se realiza una segmentación por color, en la que se estudian distintos modelos que permitan obtener las pequeñas variaciones en la tonalidad de la piel causados por el flujo sanguíneo, específicamente en la región de la cara.

Dadas las imágenes segmentadas, se extrae el promedio de color para cada componente en las imágenes, las cuales han sido previamente mapeadas al espacio de color *RGB*. Esto permite extraer de cada coordenada una señal unidimensional que contiene en su interior la información principal para estimar el ritmo cardiaco. Debido a la naturaleza ruidosa de las ondas, se hace necesario un cuarto paso de filtrado y amplificación. Este proceso permite descartar datos no coherentes y de nula importancia para el objetivo del sistema, producidos por el ruido de fondo. Aportando mayor visibilidad de los picos y fijando el análisis en los puntos esenciales de la señal. Por último, se obtiene el promedio de los máximos locales que corresponden a los picos sistólicos de la señal PPG estimando así un valor de ritmo cardiaco. Además, se extrae la Transformada Rápida de Fourier de esta señal y se obtiene otro resultado para la estimación de un valor de ritmo cardiaco.



**Figura 2.1:** Marco de trabajo propuesto para la Estimación de Ritmo Cardiaco sin contacto.

## 2.2. Extracción de Cuadros desde una Secuencia de Video

La extracción de una secuencia de imágenes de un video de prueba, se hace teniendo en cuenta la cantidad de Frames Por Segundo (FPS) a la que se reproduce dicho video. Conociendo la duración del video y su velocidad, es posible determinar la máxima cantidad de imágenes que se pueden extraer por cada segundo transcurrido. Por ejemplo, para un video que reproduce a 30 fps con una duración de 10 segundos, se extraerían unas 300 imágenes en total. Lo que se desea al extraer dichas imágenes, es tener una secuencia temporal con sus posibles cambios de color para analizar detalladamente cada una de ellas. Así, se tendrá un conjunto de elementos o cuadros de video creando una función  $F(x, y, t)$ , donde  $x$  e  $y$  corresponden al eje horizontal y vertical, respectivamente, y  $t$  hace referencia al tiempo transcurrido entre un cuadro y otro.

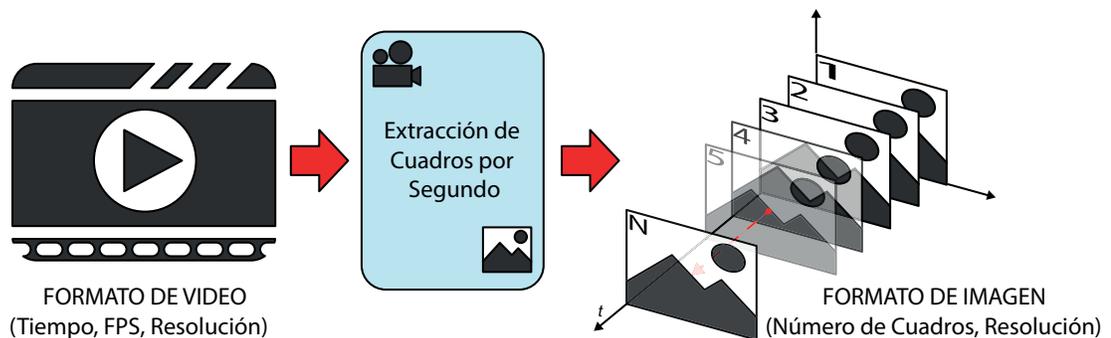


Figura 2.2: Diagrama de bloques de la Extracción de imágenes.

## 2.3. Segmentación

La segmentación de imágenes es un componente esencial en la visión por computadora. Su tarea en el análisis de imágenes es crítica y en la mayoría de los casos define el éxito del resultado final. De manera simple, la segmentación es un proceso consistente en i) separar los elementos dentro de la imagen de acuerdo a regiones con ciertas características visuales similares, y ii) asignar una etiqueta a cada píxel de modo tal que píxeles con la misma etiqueta compartan

ciertas características visuales. Para algunos autores, cumplir con la separación es suficiente para considerar hecha una segmentación, mientras que para otros, el proceso de etiquetado es también parte del problema.

El resultado de la segmentación de imágenes es un conjunto de segmentos (agrupaciones de píxeles) que de manera colectiva cubren la imagen en su totalidad. Por una parte, la segmentación debe hacerse de modo que píxeles en la misma región sean semejantes con respecto a ciertas características o propiedades calculadas, como pueden ser la intensidad, el color o la textura. Algunas de las aplicaciones más frecuentes se encuentran en la detección de rostro, estudio de las emociones a partir de gestos faciales, análisis del comportamiento humano, entre otras [Li et al., 2014].

Para este trabajo en particular, se deben detectar primero las zonas de piel en cada cuadro de la escena. Una vez que la imagen (escena) ha sido correctamente segmentada, se puede proceder directamente al procesamiento de los segmentos.

## 2.4. Detección de Piel

Como ya se mencionó antes, se requiere segmentar imágenes definiendo las zonas que corresponden a piel en ellas, específicamente piel de la cara de un sujeto. Particularmente, se sigue una segmentación orientada a la detección de grupos de píxeles considerados como piel, aislando todo lo demás en el campo visual. Esta segmentación es importante hacerla con la exactitud necesaria para evitar la inclusión de píxeles de manera errónea en la zona de interés por analizar. Es por ello que se propone la aplicación de tres espacios de color para el proceso de la detección de piel: HLS, HSV y Munsell.

Para tener una idea general de este subsistema, en la Figura 2.3 se muestra una descripción gráfica del proceso de segmentación llevado a cabo. El primer paso consiste en realizar una transformación de cada imagen, originalmente en *RGB*, a un espacio de color determinado. Una vez hecha la conversión, se fijan los límites para cada coordenada a partir del análisis de muestras de piel visualizadas en un mapa de distribución. La ubicación de los píxeles que corresponden a piel proporcionan rangos aproximados para su detección y mediante un experimento evaluado inicialmente en 6 imágenes, se les aplican los rangos que se desean probar y a través de las cuales se establecen los límites finales que serán usados en cada

espacio de color para la detección de piel. Así, se crea una máscara que contendrá la información de píxeles denominados piel, convirtiéndose ésta en la región de interés para cada imagen.

Durante el proceso de segmentación se utilizó la base de datos SFA [Casati et al., 2013]. SFA contiene 1118 imágenes de prueba, cada una de ellas con su verdad de referencia. Además incluye 3354 muestras de piel en tamaños de píxeles desde  $1 \times 1$  hasta  $35 \times 35$ . Para la construcción de los mapas de distribución, se utilizaron las muestras del tamaño de  $3 \times 3$  píxeles creando un grupo de 30168 puntos, posicionados en el espacio 3D de cada representación de color explorada.

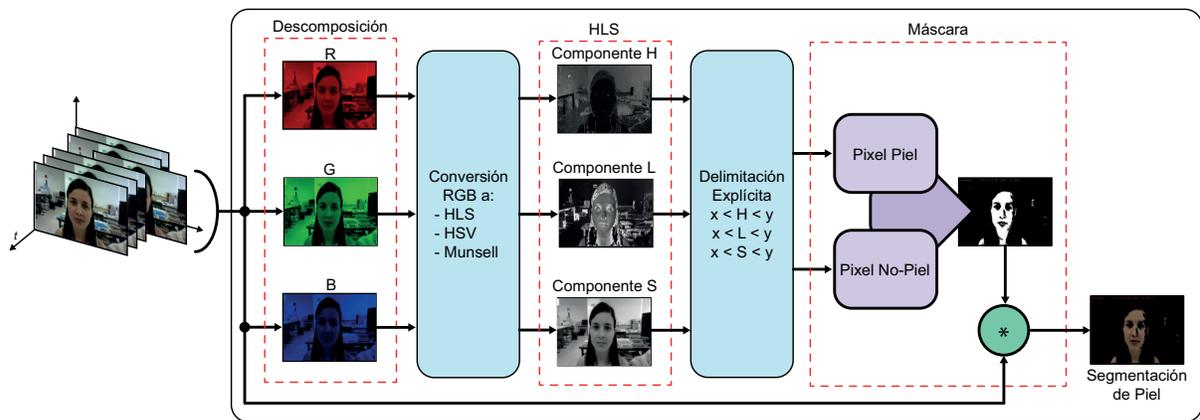


Figura 2.3: Representación general del módulo de segmentación de piel.

El color es una característica muy importante en el diseño de sistemas para segmentación de piel. Para el sistema propuesto, la elección de un espacio de color apropiado juega un papel importante en la detección de piel y el análisis de las variaciones de color. Es por eso que a continuación se amplían algunos conceptos teóricos estudiados durante el desarrollo del trabajo.

### 2.4.1. Representaciones de Color

El color es definido por Fairchild como un atributo de la percepción visual, que es formada por la combinación de contenido cromático y acromático [Fairchild, 2013]. Corresponde a una apreciación que depende puramente del ojo de quien observa, por lo tanto su interpretación puede ser subjetiva. Los espacios de color son representaciones geométricas de un conjunto de colores, que pueden expresar el

color en términos numéricos y objetivos.

Dado que la mayoría de los dispositivos de captura de video actuales utilizan de forma estándar el formato *RGB*, generalmente son usadas las conversiones entre espacios de color. En el desarrollo de la metodología se aplican de forma directa, para la detección de piel, los espacios de color *HLS*, *HSV* y Munsell. Las ecuaciones que involucran la conversión a este último, implican el uso de una transformación previa a *CIE Lab*. En consecuencia, en esta sección se presenta una breve descripción de dichos modelos que son parte del sistema general para la estimación del ritmo cardiaco.

### Espacio de color *RGB*

Se basa en la teoría tricromática y es representado por un sistema de coordenadas cartesianas tridimensional como el que se muestra en la Figura 2.4. Este espacio contiene tres colores primarios, cada uno con valores fijos: rojo (255, 0, 0), verde (0, 255, 0) y azul (0, 0, 255), que pueden ser mezclados apropiadamente para dar lugar a un nuevo color. Por ejemplo, rojo y verde combinados dan origen al color amarillo (255, 255, 0).

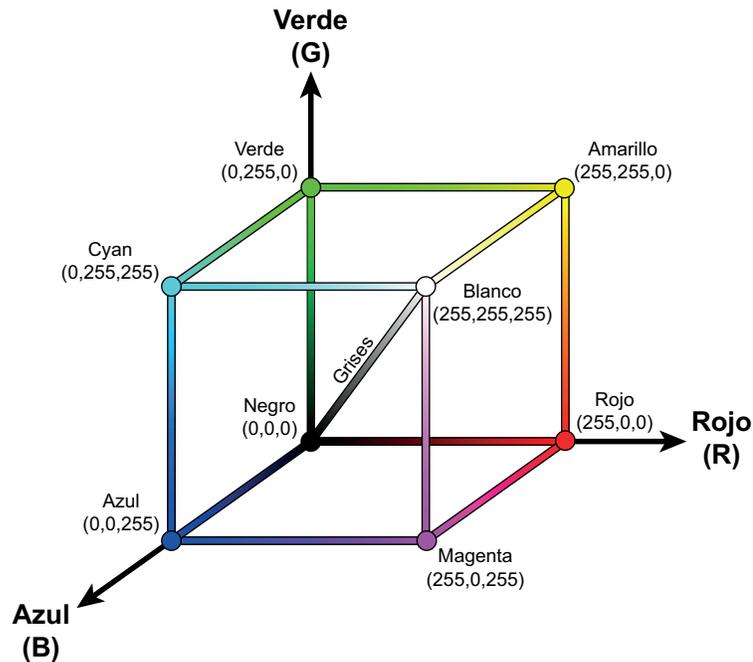
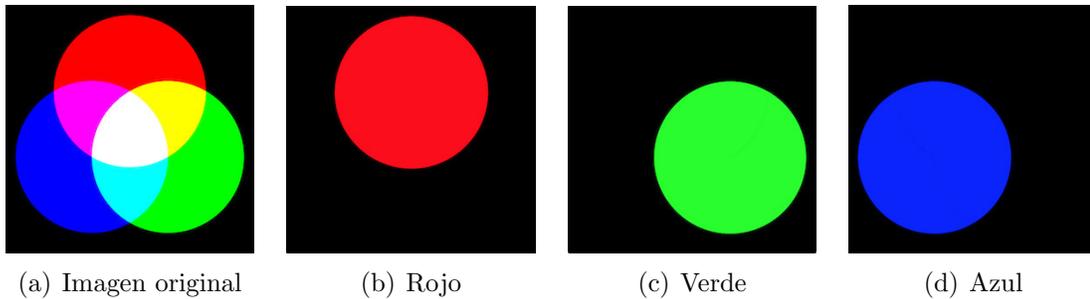


Figura 2.4: Sistema de coordenadas cartesianas tridimensional para *RGB*.

Una de las aplicaciones más comunes de este modelo se encuentra en las pantallas CRT (Cathode-Ray Tube) debido a la gran variedad de colores que pueden ser creados a partir de la mezcla de sus colores primarios. Es una de las representaciones de color más básicas y aunque es el formato de color estándar, no se prefiere usar en la detección de piel ya que no presenta suficiente robustez frente a los cambios bruscos de iluminación.

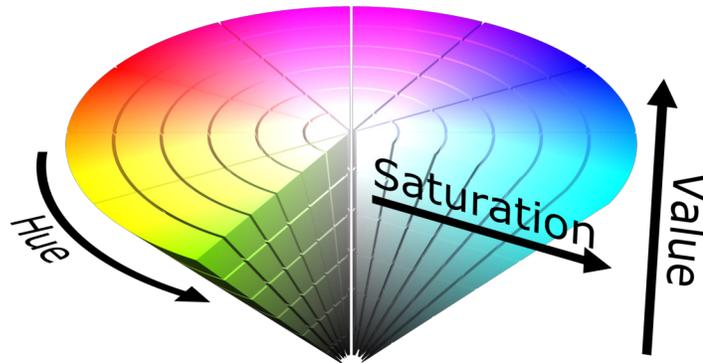


**Figura 2.5:** *Imagen original y las componentes individuales del espacio de color RGB*

En Figura 2.5 se observa una imagen descriptiva de las componente de color de manera individual.

### Espacio de color *HSV*

El modelo de color *HSV* (Por sus siglas en inglés Hue, Saturation y Value) es conocido también como *HSB* (Por sus siglas en inglés Hue, Saturation y Brightness). El tono de color es representado a través de la componente *H* y como se puede ver en la Figura 2.6 su rango está entre  $0^\circ$  y  $360^\circ$ . La pureza del color se define mediante la saturación, con una distancia que va desde 0 hasta 1, por mencionar sus equivalentes normalizados. Los colores que poseen escasa saturación tienden a ser más grises y opacos, lo que permite identificar visualmente esta característica.



**Figura 2.6:** Representación cónica del modelo HSV.

La última componente se refiere a la luminosidad o el brillo presentes en el color y su rango varía verticalmente desde 0 hasta 1, en el cono tridimensional que mapea dicho espacio. La obtención de las componentes  $HSV$ , se realizan mediante transformación no lineal, una vez se tienen las coordenadas  $RGB$ , como sigue:

$$X = -1/\sqrt{6}R - 1/\sqrt{6}G + 2/\sqrt{6}B \quad (2.1)$$

$$Y = 1/\sqrt{6}R - 1/\sqrt{6}G \quad (2.2)$$

$$H = \arctan(Y/X) \quad (2.3)$$

$$S = (X^2 + Y^2)^{1/2} \quad (2.4)$$

$$V = 1/3R + 1/3G + 1/3B \quad (2.5)$$

### Espacio de color $HLS$

Fue inventado por Georges Valensi en 1938 con el objetivo de presentar características perceptuales en la forma como se describe cada componente. En  $HLS$ , el color es definido como una coordenada en un plano circular. Existen otros espacios de color que pueden ser claramente muy parecidos en sus componentes, el mismo  $HSV$  descrito anteriormente, entre otros. Esta variedad parece surgir

de la necesidad de expresar el color lo mejor posible con respecto a la percepción humana.

La componente  $H$  que indica el tono del color, varía a través del eje circular del cilindro que representa geoméricamente este modelo de color. La saturación  $S$  cambia a lo largo del radio del cilindro. Así, el color será más intenso en la medida en que el valor de saturación esté más alejado del centro. Por su parte, la componente  $L$  se mueve a lo largo del eje vertical del cilindro e indica la luminosidad o brillantez del color. El valor de la saturación y la luminancia están en el rango de 0 a 255, mientras que el valor de  $H$  se define en un rango de  $0^\circ$  a  $360^\circ$ .



(a) Imagen Original    (b) Componente  $H$     (c) Componente  $L$     (d) Componente  $S$

**Figura 2.7:** Imagen original y cada componente del espacio  $HLS$  en escala de grises.

La Figura 2.7 muestra la forma como se perciben individualmente las componentes de este espacio de color, en escala de grises.

### Espacio de color *CIELab*

Este espacio de color es considerado dentro del grupo de los perceptuales, debido a que su principal objetivo es representar de manera más precisa la experiencia visual de los humanos. El modelo fue creado por la Comisión Internacional de Iluminación (*CIE* por sus siglas en inglés) en 1976, y se originó a partir de un modelo conocido como *CIE-XYZ* que fue diseñado en 1931, en el cual los colores son descritos en términos de las coordenadas  $XYZ$ .

Para calcular las coordenadas de *CIELab*, es preciso encontrar antes un conjunto de coordenadas *XYZ* de *CIE-XYZ* y tener los valores de otro conjunto de referencia  $X_n Y_n Z_n$  basados en el blanco de referencia. El primer conjunto *XYZ* se define linealmente a través del siguiente grupo de ecuaciones:

$$X = 0.490186R + 0.309879G + 0.199934B \quad (2.6)$$

$$Y = 0.177015R + 0.812324G + 0.010660B \quad (2.7)$$

$$Z = 0.010077G + 0.989922B \quad (2.8)$$

Una vez que se han resuelto las ecuaciones anteriores, se calculan las componentes  $L^*$ ,  $a^*$  y  $b^*$ .

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{1/3} & \text{si } \frac{Y}{Y_n} > 0.008856 \\ 903.3\left(\frac{Y}{Y_n}\right)^{1/3} & \text{si } \frac{Y}{Y_n} \leq 0.008856 \end{cases} \quad (2.9)$$

$$a^* = 500 \times (f(X/X_n) - f(Y/Y_n)) \quad (2.10)$$

$$b^* = 200 \times (f(Y/Y_n) - f(Z/Z_n)) \quad (2.11)$$

$$f(t) = \begin{cases} t^{1/3} & \text{si } t > 0.008856 \\ 7.787 \times t + \frac{16}{116} & \text{si } t \leq 0.008856 \end{cases}$$

### Espacio de color Munsell

Este sistema se forma con base en tres dimensiones perceptuales de color: Matiz ( $H$ ), Valor o Luminosidad ( $V$ ) y la Croma ( $C$ ). Fue creado en los primeros años del siglo XX por Albert Henry Munsell. Los componentes de este modelo, están

todos enfocados en los atributos de apariencia visual en los sistemas actuales. Su distribución geométrica permite definir colores con incrementos visuales iguales a través de cada una de las coordenadas.

El Matiz ( $H$ ), es representado por una escala circular que contiene 10 tonos principales. Cada uno de ellos ha sido representado de manera estándar con una notación de siglas en inglés, como sigue: Rojo (R), Amarillo-Rojo (YR), Amarillo (Y), Verde-Amarillo (GY), Verde (G), Azul-Verde (BG), Azul (B), Púrpura-Azul (PB), Púrpura (P) y Rojo-Púrpura (RP). Estos matices se subdividen en un rango de 1 a 10 formando una nueva escala de 100 puntos en esta coordenada. Un ejemplo de esta notación podría ser 5B, lo que significaría el paso de 5 en la categoría del color azul. La gráfica de la Figura 2.8 muestra una representación circular de esta componente.

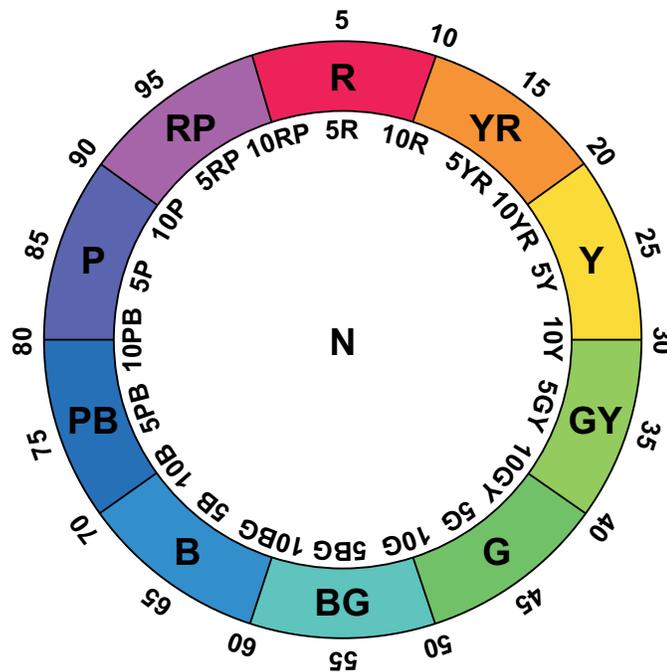


Figura 2.8: Representación gráfica de la componente de color  $H$ .

Utilizando las ecuaciones 2.6 a la 2.8, del espacio de color  $CIE_{Lab}$  y siguiendo el enfoque propuesto por Zhang *et al.* [Zhang et al., 1998] se calculan el conjunto de variables  $xyz$  correspondiente:

$$x = \frac{X}{X + Y + Z} \quad (2.12)$$

$$y = \frac{Y}{X + Y + Z} \quad (2.13)$$

$$z = \frac{Z}{X + Y + Z} \quad (2.14)$$

Dichos valores serán usados para hallar un nuevo conjunto de coordenadas  $a^*$  y  $b^*$ , las cuales permitirán calcular valor de  $H$  como sigue:

$$H = \arctan(b^*/a^*) \quad (2.15)$$

La siguiente componente del sistema Munsell se conoce como *Value* o Luminosidad. Es definida en un rango desde 0 hasta 10, donde 0 corresponde al color negro y 10 al color blanco, con tonos grises en el intermedio, como se muestra en la Figura 2.9. La precisión de la escala es tal, que la diferencia percibida de la luminosidad entre dos muestras vecinas es igual en cualquier otro par de muestras que varíen en un solo paso.



**Figura 2.9:** Representación gráfica de la componente de color del valor de luminosidad.

La ecuación para el cálculo de esta componente, se ha estudiado por McCamy [McCamy, 1992], demostrando un error menor que el generado con las aproximaciones anteriores. Entonces,  $V$  es hallado a partir del factor de luminancia  $Y$  del espacio de color *CIELab*, así:

Para  $Y \leq 0.9$

$$V = UY^W \quad (2.16)$$

Para  $Y > 0.9$

$$V = AY^{1/3} - B - C/[(DY - E)^2 + F] + G/Y^H + J \sin(KY^{1/3} + 1) + (M/Y) \sin[N(Y - 2)] - (P/QY) \sin[S(Y - T)] \quad (2.17)$$

El valor en 0.9 representa el mínimo valor de la cantidad de luz reflejada por una superficie blanca ideal. La componente de luminosidad  $V$  mantiene una relación de tipo univariante con el factor de luminancia  $Y$ , lo que explica la solución de la ecuación con referencia a un valor frontera de  $Y$ , pero no aumenta de forma lineal con respecto a la luminancia.

La tercera componente del sistema Munsell, es la Croma. La escala diseñada posee incrementos visuales iguales desde un valor mínimo de 0 hasta valores máximos no establecidos. Un cero puede indicar un color negro, gris o blanco, es decir, muestras neutras y los valores máximos indican percepciones de color gradualmente mas puras.



**Figura 2.10:** Representación gráfica de la componente de color Croma.

Para obtener una aproximación de dicha componente, se utilizan el conjunto de coordenadas  $a^*$  y  $b^*$  calculadas para hallar  $H$ , así:

$$C = \sqrt{(a^*)^2 + (b^*)^2} \quad (2.18)$$

Las representaciones lineales de las componentes de luminosidad y cromas, deben ser acopladas junto con las del matiz para conformar una geometría completa del sistema, resultando en la distribución que se muestra en la Figura 2.11

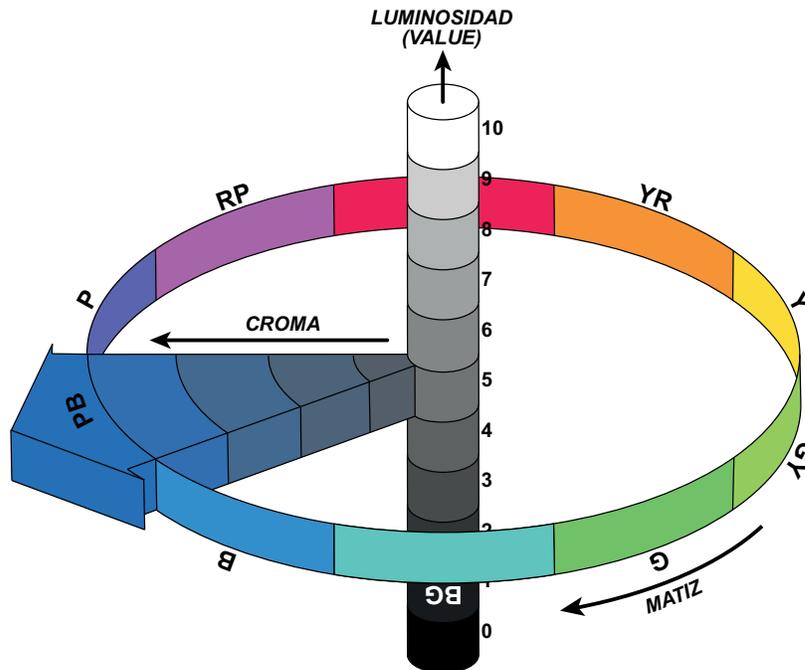


Figura 2.11: Representación geométrica del modelo Munsell.

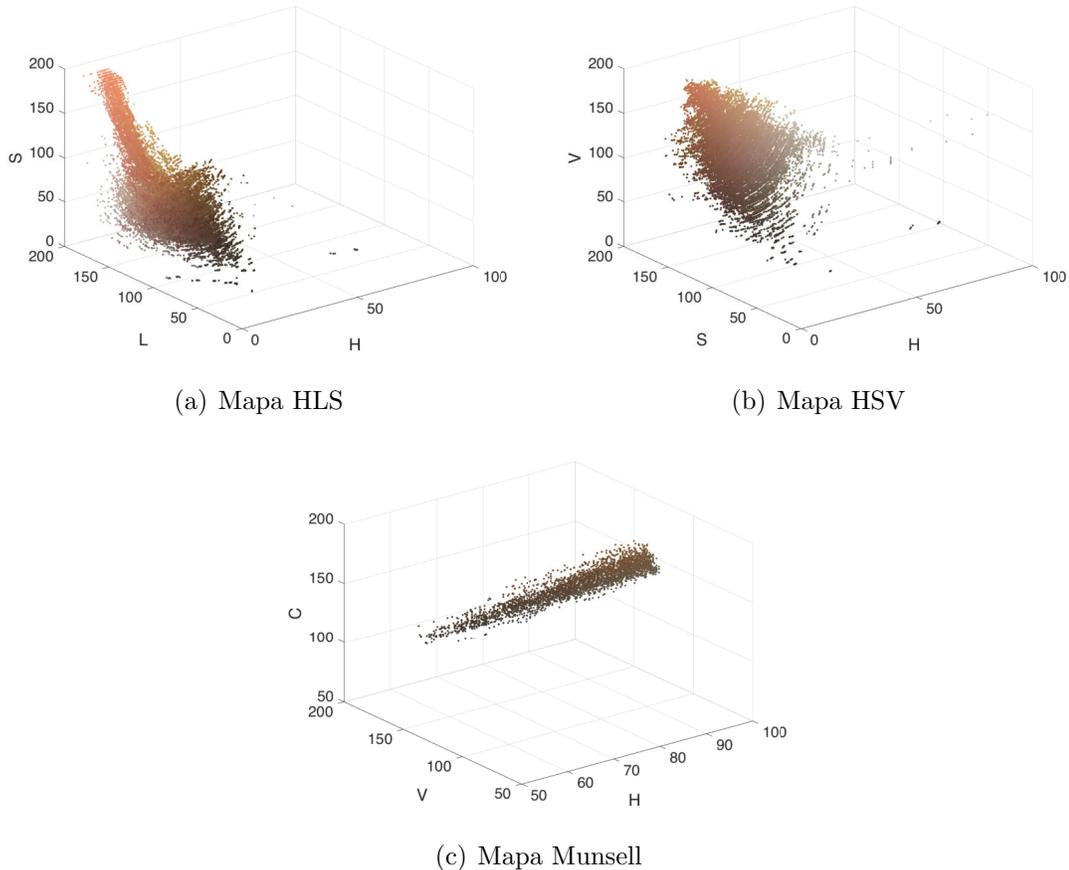
### 2.4.2. Delimitación explícita de la Región de Piel

El objetivo de la detección de piel, se sigue principalmente para limitar la región a ser analizada. Dado que el estudio se centra en los cambios de color presentes en la piel, se definen solo dos categorías en una imagen, los píxeles que son piel y los que no lo son. Es por ello, que la región de interés queda definida como el conjunto de todos los píxeles segmentados como piel en la imagen. Una vez detectada, se extraen y analizan las características de color presentes en la piel.

Para seleccionar los píxeles de color piel se usó el método de delimitación explícita [Vezhnevets et al., 2003, Kakumanu et al., 2007]. Es uno de los métodos más prácticos y utilizados, en el que se definen intervalos sobre las componentes principales de un determinado espacio de color. De esta manera, se formulan una serie de reglas condicionales a partir de los rangos establecidos para acotar regiones de interés que contengan la mayor cantidad de píxeles con intensidades del color deseado. Un ejemplo del tipo de reglas utilizadas en este método son las propuestas por [Osman et al., 2012], así:

$$RGB \in \begin{cases} Piel & \text{si } (96 < R < 255) \cap (41 < G < 239) \cap (21 < B < 254) \\ No\ piel & \text{de otra manera} \end{cases} \quad (2.19)$$

Durante el proceso de segmentación se siguen varias fases, las cuales se aplican de manera general a los espacios de color HLS, HSV y Munsell. La condición de mayor importancia es que sean consideradas las características de color que permiten encontrar pequeños cambios en el color de la piel, los cuales serán analizados en una base de tiempo.



**Figura 2.12:** Distribución espacial de las muestras de color piel.

Dados los mapas de distribución de la Figura 2.12, elaborados a partir de las muestras de piel en la base de datos SFA, se generan combinaciones de

rangos diferentes para cada espacio de color. Cada límite arroja un conjunto de resultados que ha sido evaluado mediante un procedimiento que analiza cada imagen segmentada a fin de obtener una medida que refleje la calidad lograda. Para ello, se realiza una comparación entre los píxeles piel de la segmentación evaluada y aquellos definidos como tal por la referencia, a fin de encontrar la correspondencia entre ambos. Se requiere entonces, definir tres categorías a saber: verdaderos positivos ( $TP$ ), falsos positivos ( $FP$ ) y falsos negativos ( $FN$ ). Al determinar el número total de píxeles que pertenecen a cada categoría es posible calcular los valores de Precisión ( $P$ ) y Exhaustividad<sup>1</sup> ( $R$ ) con el objetivo de encontrar el valor para la media armónica o medida-F.

La Precisión y la Exhaustividad son frecuentemente utilizadas en la evaluación de metodologías de segmentación. Lo que se desea es medir la capacidad de un sistema frente a la recuperación de información relevante. Es así como un valor bajo de Precisión refleja que el número de falsos positivos en la imagen evaluada es alto comparado con el número de verdaderos positivos y un valor bajo de Exhaustividad indica que la proporción de píxeles piel hallados en la imagen evaluada es considerablemente menor que el número de píxeles piel esperados en comparación con la imagen de referencia. Dichas medidas están definidas en función del número total de verdaderos positivos, falsos positivos y falsos negativos, como se muestra en las Ecuaciones 2.20 y 2.21.

$$P = \frac{TP}{TP + FP} \quad (2.20)$$

$$R = \frac{TP}{TP + FN} \quad (2.21)$$

Los valores obtenidos para  $P$  y  $R$  se obtienen en un espacio de valores mínimo de 0 y un máximo de 1. Los datos obtenidos a partir de las ecuaciones anteriores, sirven como base para determinar una medida única de evaluación de los límites apropiados para la detección de piel en cada modelo de color. Dicho esto, la Ecuación 2.22 define la medida-F como una función de  $P$  y  $R$  de la siguiente manera:

---

<sup>1</sup>Las siglas usadas para nombrar las variables TP, FP, FN, P y R son tomadas del inglés *True Positive*, *False Positive*, *False Negative*, *Precision* y *Recall*, respectivamente.

$$F = \frac{P \times R}{\alpha P + (1 - \alpha)R} \quad (2.22)$$

La constante  $\alpha$  en la Ecuación anterior, es un parámetro adicional que permite establecer la relevancia de las medidas  $P$  y  $R$ . De acuerdo con ello, se ajusta un valor de  $\alpha = 0.5$  para dar igual relevancia a ambas medidas.

Con el fin de determinar el conjunto de valores finales, para cada modelo de color, que serán aplicados durante la segmentación, se ha diseñado el siguiente experimento:

1. Se eligen 6 imágenes de la base de datos que contienen un tono de piel distinto entre ellas, como una muestra para definir un rango adecuado.
2. Se ajusta un límite para cada componente del espacio de color puesto a prueba.
3. Se les aplica el proceso de segmentación obteniendo su equivalente monocromático y se compara con la verdad de referencia de la base de datos.
4. Se calculan los promedios de la medida-F para un valor de  $\alpha = 0.5$ , dando el mismo peso a la Precisión y la Exhaustividad.
5. El conjunto de valores que aportan el valor de la medida-F más cercano a 1, es elegido para usarse en toda la base de datos.

Estas pruebas se realizaron para 5 combinaciones de rangos diferentes, siempre en el mismo grupo de imágenes, en cada espacio de color. Los mejores resultados promedio obtenidos de la medida-F, así como de la Precisión y la Exhaustividad fueron organizados en la Tabla 2.1. Estos valores se calcularon usando 6 cifras decimales durante el experimento.

<b>Espacio de color</b>	<b>Precisión</b>	<b>Exhaustividad</b>	<b>Medida-F</b>
HLS	0.92	0.78	0.89
HSV	0.92	0.80	0.89
Munsell	0.93	0.89	0.88

**Tabla 2.1:** Resultado del promedio de los valores obtenidos para el grupo muestra de 6 imágenes.

El conjunto de imágenes utilizadas como base en el proceso de segmentación, se puede ver en la primera columna de la Figura 2.13. La siguiente columna contiene las imágenes de referencia proporcionadas por la base de datos y los siguientes grupos de imágenes son los resultados de la segmentación para cada espacio de color.

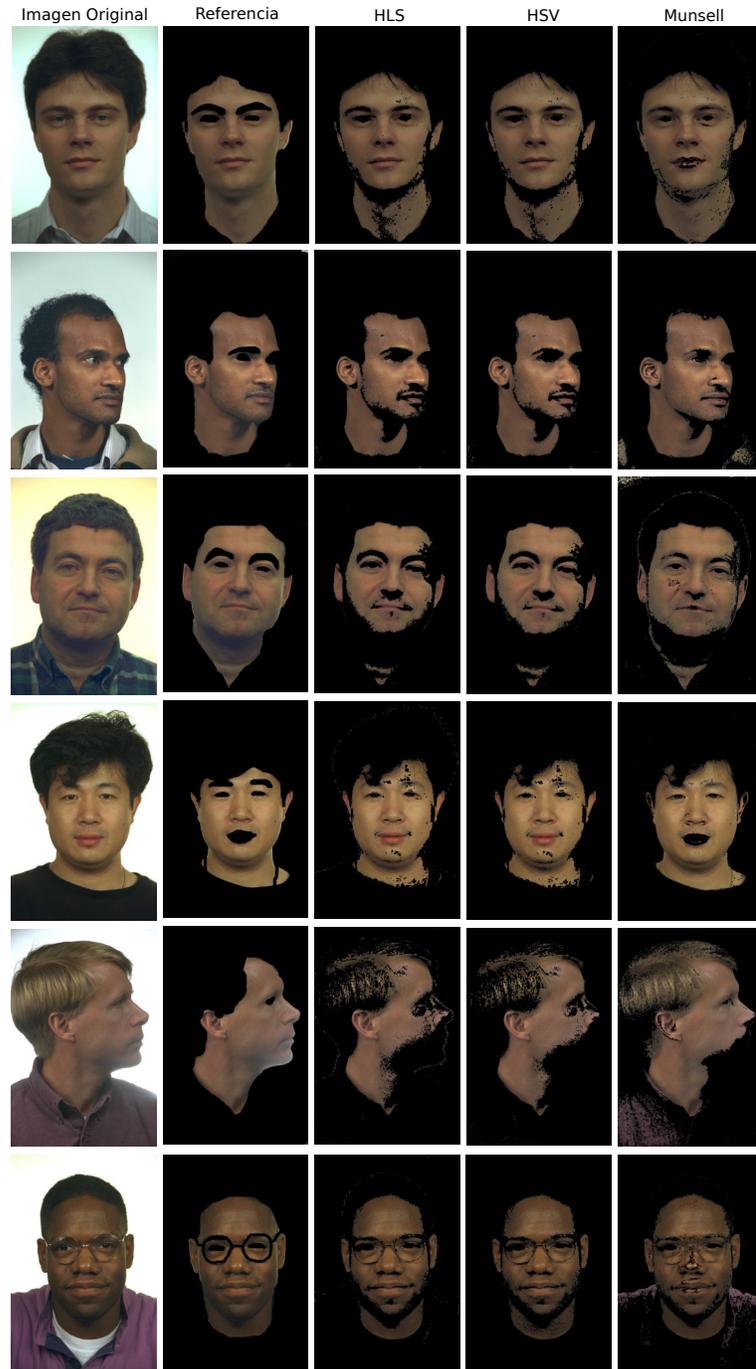


Figura 2.13: Conjunto de imágenes muestra con su salida segmentada.

El primer espacio de color usado en la detección de piel corresponde al HLS, brevemente descrito en la sección 2.4.1. Los límites establecidos que arrojan los mejores resultados según el experimento descrito, organizados en la Tabla anterior, son los siguientes:

$$HLS \in \begin{cases} Piel & \text{si } [(H \leq 18) \cup (H \geq 165)] \cap (25 \leq L \leq 150) \cap (S \geq 50) \\ No Piel & \text{de otra manera} \end{cases} \quad (2.23)$$

Para el caso del espacio de color HSV, el segundo puesto a prueba, los límites establecidos que mostraron un mejor desempeño en la detección de piel, son:

$$HSV \in \begin{cases} Piel & \text{si } [(H \leq 18) \cup (H \geq 170)] \cap (S \geq 80) \cap (40 \leq V \leq 240) \\ No Piel & \text{de otra manera} \end{cases} \quad (2.24)$$

Por último, se propone el siguiente conjunto de límites para el espacio de color Munsell.

$$Munsell \in \begin{cases} Piel & \text{si } [(Hm \leq 50) \cup (Hm \geq 100)] \cap (Cm \geq 180) \cap (0.6 \leq Vm \leq 0.736) \\ No Piel & \text{de otra manera} \end{cases} \quad (2.25)$$

El objetivo de establecer límites específicos, es probar cual proporciona mejores resultados a la hora de estudiar los cambios presentes en la señal PPG que permitirá estimar un valor de HR lo más cercano al esperado.

Una vez se tienen los límites que prueban el mejor desempeño comparado con la verdad de referencia, estas restricciones son aplicadas al conjunto total de las imágenes en la base de datos. Los resultados arrojados del promedio total de las métricas se muestran en la Tabla 2.2.

<b>Espacio de color</b>	<b>Precisión</b>	<b>Exhaustividad</b>	<b>Medida-F</b>
HLS	0.85	0.71	0.82
HSV	0.92	0.73	0.83
Munsell	0.80	0.86	0.81

**Tabla 2.2:** Resultado del promedio de los valores obtenidos para toda la base de datos.

Finalmente, la imagen segmentada será mapeada al espacio de color *RGB* de donde se extraen los promedios de color para cada componente. Estos vectores proporcionan información valiosa que será analizada para evaluar su contenido frecuencial y determinar a partir de ello la señal Fotopleletismográfica. La siguiente sección explica en detalle cómo se realiza este proceso que conforma la fase número tres del sistema.

## 2.5. Cálculo del Color Promedio

La estimación del ritmo cardiaco a través de cámaras a color con baja resolución, fue mostrado por Verkruysse *et al.* hace casi 10 años [Verkruysse et al., 2008]. Ellos encontraron que la señal PPG puede ser detectada en las componentes de color roja, verde y azul presentes en la piel. Sin embargo, diferentes estudios han demostrado que la componente verde, posee una más fuerte definición de la señal [Poh et al., 2010a, Li et al., 2014, Lam and Kuno, 2016].

La fotopleletismografía mide la intensidad de la luz reflejada por la superficie de la piel para determinar el volumen de la sangre, tal como funciona para los oxímetros de pulso utilizados a menudo para el monitoreo de la salud en los niños y en adultos mayores. Sin embargo, es indispensable el contacto físico directo con el paciente. Para el análisis que nos ocupa, se utilizó este principio y se trabajó en busca de la señal PPG a partir de videos cortos de la piel facial, esto debido a la practicidad que representa el hecho de que en su mayoría, todos tenemos siempre la cara descubierta.

Se considera como característica principal la medida individual de los cambios de color, representados en valores numéricos, hallados a partir del promedio de color para cada componente en el espacio de color *RGB*. Aun cuando la adquisición de la información representa un reto en sí mismo, un logro presentado en este

trabajo es el uso de la cámara de una computadora de escritorio convencional, en un ambiente de oficina no controlado, con la iluminación típica y sin modificación del segundo plano en la escena. Un video de tan solo 10 segundos es prácticamente suficiente para captar los cambios pulsátiles en las arteriolas de la cara, causados por el pulso cardiovascular. Es decir, estas variaciones producto de los latidos del corazón, traen consigo la información de la frecuencia cardiaca, necesaria para la estimación del ritmo cardiaco. Así, el objetivo de esta fase es registrar los sutiles cambios de color presentes en la piel, particularmente del rostro de una persona.

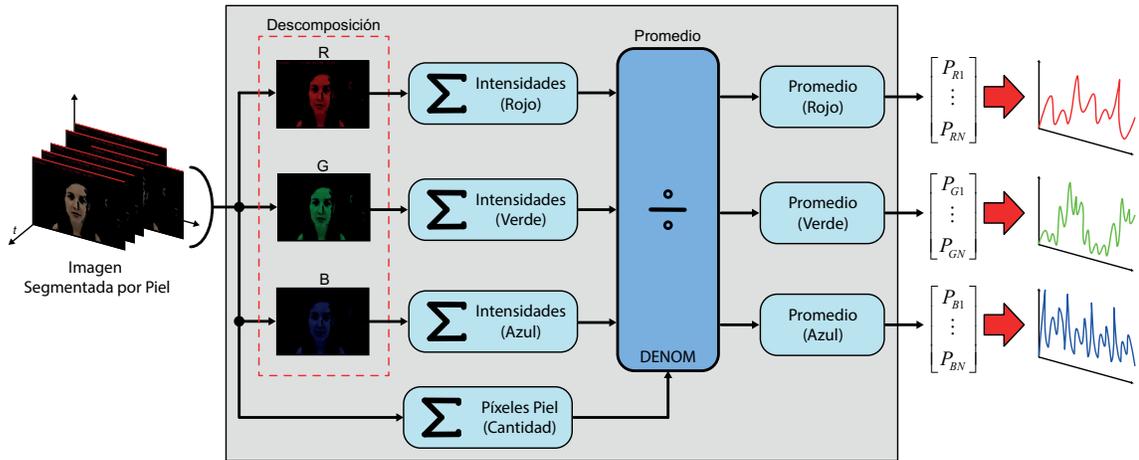


Figura 2.14: Descripción general del módulo del color promedio.

Con el objetivo de analizar la señal que guarda la información de la frecuencia cardiaca, se hace necesario encontrar las pequeñas variaciones presentes en el color. Para ello, se determina el promedio de color en cada imagen segmentada, mediante el siguiente conjunto de ecuaciones:

$$P_R = \frac{\sum R_i}{I_p} \quad (2.26)$$

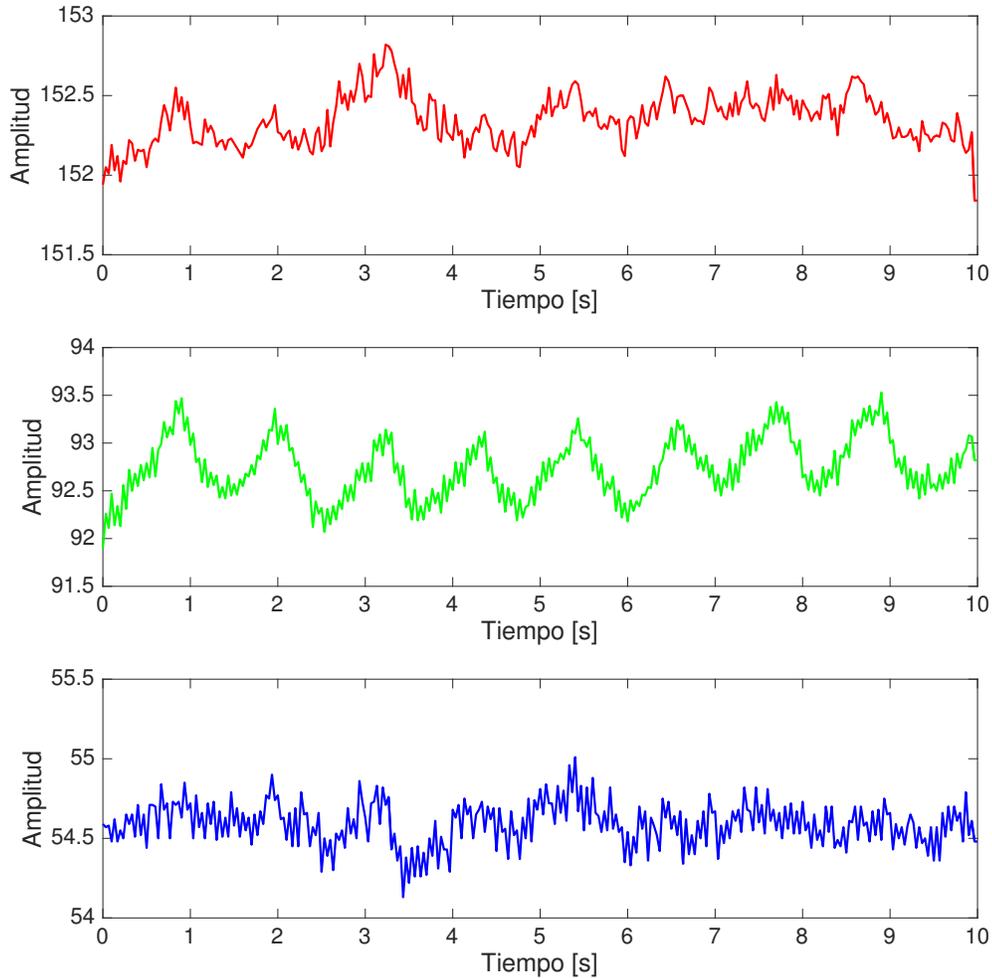
$$P_G = \frac{\sum G_i}{I_p} \quad (2.27)$$

$$P_B = \frac{\sum B_i}{I_p} \quad (2.28)$$

Donde  $R_i$ ,  $G_i$  y  $B_i$  son las intensidades de los píxeles para cada una de las componentes  $RGB$  de la región de interés e  $I_p$  son todos los píxeles piel que

componen la imagen.

Los promedios de color calculados para cada imagen en una secuencia de video, forman un un vector de elementos por cada componente, lo que genera una señal unidimensional para  $y_R$ ,  $y_G$  y  $y_B$  como se muestra en la Figura 2.15.



**Figura 2.15:** Señales de pulso extraídas del valor promedio de las intensidades de los píxeles dentro de la región de interés.

Diferentes trabajos presentados aseguran que los píxeles provenientes de la componente de color verde contiene una señal PPG más fuerte, sin embargo, en todas las demás coordenadas de color existe dicha señal aunque en un nivel más bajo. Esta observación es válida debido a que la luz verde es mejor absorbida por la

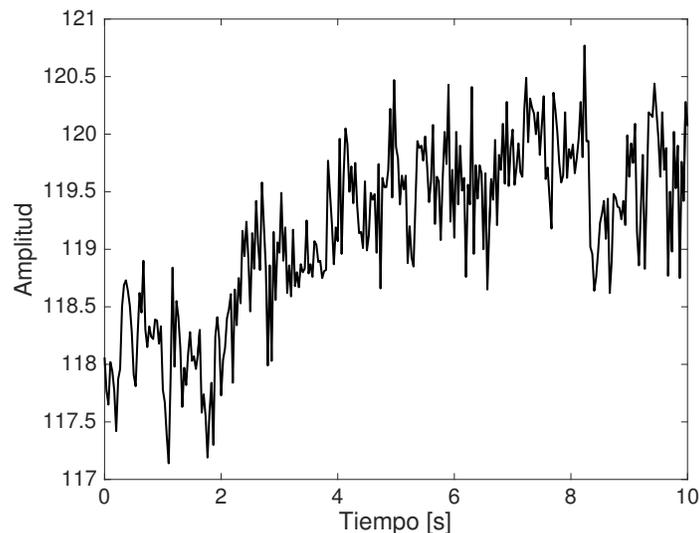
oxihemoglobina y llega más profundo en la piel. Además, durante el experimento se realizaron pruebas que verificaron finalmente el contenido de frecuencia cardiaca presente en la señal  $y_G$  y se comprobó que es mayor y se encuentra mejor definida comparada con las demás. Es por esto, que para todas las siguientes fases del sistema se utiliza únicamente el vector de datos que contiene la información de la componente de color verde.

## 2.6. Proceso de Filtrado y Amplificación

En esta sección se describe la metodología aplicada al filtrado de la señal y a la amplificación de la misma con el propósito de obtener un mejor desempeño.

### 2.6.1. Eliminación de la Tendencia no Lineal

En el proceso de filtrado se evalúa la naturaleza de la señal para definir el tipo de filtro a implementar. Particularmente, una señal PPG es de origen no estacionaria, lo que indica variaciones en el contenido de frecuencia a través del tiempo. Además, como se observa en la Figura 2.16, las formas de onda poseen una tendencia de tipo no lineal, lo que justifica el uso de un filtro que facilite la eliminación de patrones generales que no son propios de los datos en una señal PPG.



**Figura 2.16:** Señal pura con tendencia no lineal.

En el área de procesamiento de señales, los procesos de *detrending* o de eliminación de tendencias se utilizan para sustraer comportamientos indeseados o poco relevantes que se encuentran presentes en una secuencia de datos. Dichas tendencias pueden oscurecer, camuflar o simplemente dificultar la identificación y caracterización de comportamientos subyacentes que sí son de interés dentro de la señal estudiada. Al eliminar estas tendencias, se facilita el análisis de las fluctuaciones en la señal, haciendo más evidentes los comportamientos que se relacionan de manera directa con el fenómeno o parámetro físico de interés. Adicionalmente, la ausencia de comportamientos ajenos a la intención del análisis en la secuencia de datos permite el uso de herramientas más sencillas y de menor costo computacional para tratar la señal estudiada, evitando etapas más complejas de filtrado o acondicionamiento.

En este trabajo, el proceso de *detrending* para eliminar comportamientos no-lineales se realiza en dos pasos: el ajuste polinomial a la curva  $y_G$ , y su evaluación para posterior substracción de la señal  $y_G$ . En primera instancia, se utiliza la herramienta de software Matlab<sup>®</sup> y la función *polyfit* para retornar los coeficientes de un polinomio  $p(x)$  que corresponde al mejor ajuste en términos de mínimos cuadrados a una señal de entrada  $y$ . La cantidad de los coeficientes de  $p$  son  $n + 1$  (donde  $n$  corresponde al orden del polinomio) y se organizan en potencias decrecientes hasta un valor constante. La expresión que denota la forma del polinomio se muestra en la Ecuación 2.29.

$$p(x) = p_1x^n + p_2x^{n-1} + \dots + p_nx + p_{n+1} \quad (2.29)$$

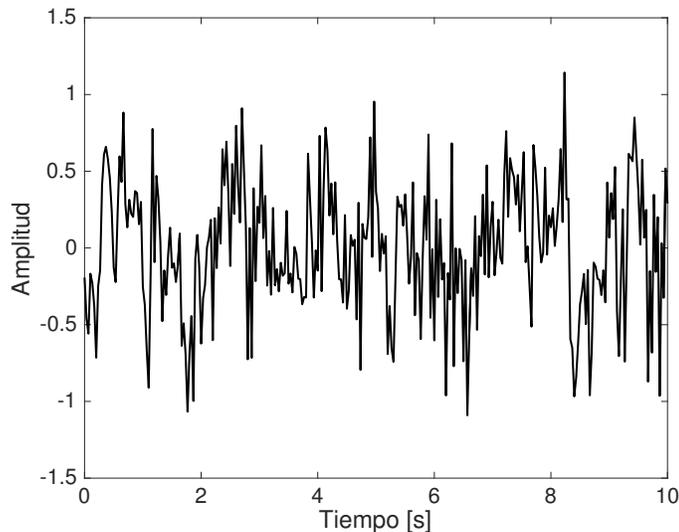
Adicionalmente la función *polyfit* retorna los valores de media  $\mu$  y varianza  $\sigma^2$  como muestra la Ecuación 2.30, información útil para centrar y escalar  $x$  de tal modo que tenga una desviación estándar unitaria, lo que mejora las propiedades numéricas para el segundo paso: el algoritmo de ajuste para evaluación polinomial.

$$\hat{x} = \frac{x - \bar{x}}{\sigma_x} \quad (2.30)$$

La segunda parte corresponde a la evaluación del polinomio y es llevada a cabo por medio de la función *polyval* a partir de la información proporcionada por la etapa anterior. Una vez se tiene la aproximación polinomial a la tendencia que

se desea eliminar, se subtrae de la señal  $y_G$  para descartar (en la medida de lo posible) el comportamiento no-lineal que dificulta el análisis de la señal de interés.

El resultado de aplicar este proceso de filtrado es una señal normalizada como la que se observa en la Figura 2.17 cuya forma se extiende a lo largo del tiempo y en la que se ha suprimido su componente DC.



**Figura 2.17:** Señal filtrada, sin tendencia no lineal.

### 2.6.2. Filtro Pasa Banda

El uso de filtros ayuda a aislar características de especial interés dentro de una determinada señal de estudio, excluyendo elementos indeseados. Dependiendo del tipo de aplicación, se pueden utilizar distintos tipos de filtros de acuerdo con el acondicionamiento que se le quiera dar a la señal para facilitar su análisis. En el área de tratamiento de imágenes suelen utilizarse filtros básicos de tipo pasa-bajas (LBP), pasa-altas (HBP), o pasa-bandas (BPF), entre otros.

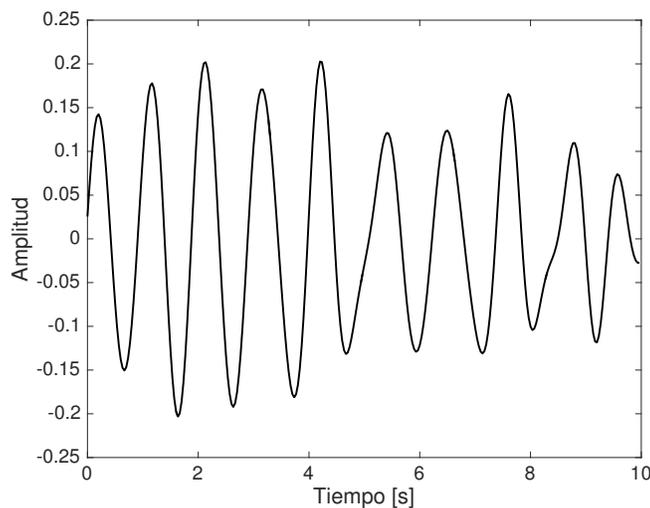
Para el trabajo realizado en este proyecto, se utiliza un BPF con el objetivo de hacer más sobresalientes los rangos en los que se encuentran las frecuencias cardiacas (parte baja del espectro) y eliminar fuentes de ruido presentes debido al fondo y la calidad de las cámaras utilizadas para capturar las secuencias de video (parte alta del espectro). Estas frecuencias extremas están situadas por fuera del rango de interés y por tanto son atenuadas por el filtro.

El método escogido para el filtrado es de tipo digital mediante un filtro Infinite Impulse Response (IIR). Para la generación de los coeficientes se utilizó una ventana de Butterworth dada su simplicidad y bajo costo computacional [Balakrishnan et al., 2013]. El diseño de su estructura se realiza interpretando un filtro de tipo IIR como una ecuación de diferencias [Proakis and Manolakis, 2007] a través de la cual corre una señal de entrada discretizada. A su paso, las muestras de la señal se multiplican y acumulan por cada uno de los coeficientes constantes definidos por la ventana de filtrado elegida. La forma general de un filtro IIR se muestra en la Ecuación (2.31).

$$y(k) = \sum_{i=0}^n a_i x(k-i) - \sum_{i=1}^n b_i y(k-i) \quad (2.31)$$

donde  $a_i$  y  $b_i$  son coeficientes constantes,  $n$  es el orden del filtro e  $y$  es la respuesta del filtro en la muestra  $k$ .

Una vez que se ha pasado la señal pura por el proceso de eliminación de tendencia no-lineal, se procede a la aplicación del filtro pasa-banda de orden  $n = 5$  en los rangos de interés. Dado que no todas las frecuencias presentes en la curva son requeridas o de hecho útiles para la detección del pulso se considera aplicar un filtro de entre  $[0.83 \ 1.66]$  Hz lo que cubre el rango normal de ritmo cardiaco en adultos sanos o lo que es lo mismo  $[50 \ 100]$  bpm. La Figura 2.18 muestra la salida de una señal PPG después de que se ha filtrado en el rango propuesto.



**Figura 2.18:** Señal filtrada en el rango de interés.

En este caso, las frecuencias por fuera de dicho rango afectan los resultados esperados, dado que en muchas ocasiones los movimientos involuntarios o el ruido producido por el exterior pueden ser dominantes frente a las pequeñas variaciones que se pretenden analizar y que sí corresponden con las frecuencias de pulso. Sin embargo, cuando se trata de tramas de videos tomadas a la población infantil, más exactamente a recién nacidos, el valor de pulso varía en ocasiones a más del doble del que presenta naturalmente un adulto, sin ser esto fuera de lo esperado, pues como se dijo al inicio de este libro, los bebés sanos presentan un rango de ritmo cardiaco aproximadamente en  $[100\ 160]$  bpm lo que corresponde en frecuencia a  $[1.66\ 2.66]$  Hz.

### 2.6.3. Amplificación de la Señal Filtrada

El objetivo de este paso es realzar los puntos que componen la señal para obtener un mejor desempeño en el cálculo de los máximos locales presentes en la curva PPG. La información contenida en la señal unidimensional, se puede proyectar inicialmente como un problema de relación señal a ruido más conocido como SNR, dado que se identifica a la señal de interés envuelta en un ruido de magnitud mucho mayor en comparación con ella misma. Para comprender mejor, la relación se representa brevemente en la siguiente ecuación:

$$S = h + n; \quad h \ll n \tag{2.32}$$

donde  $S$  representa la señal obtenida,  $h$  las frecuencias de interés y  $n$  el ruido presente en la señal.

La amplificación se aplica directamente sobre la señal filtrada y permite encontrar más claramente los máximos locales presentes en la señal, los cuales corresponden a los picos sistólicos en la señal PPG. Un valor de amplificación  $\beta = 80$  define más claramente los puntos máximos y su lugar en la señal extraída, reduciendo considerablemente el nivel de ruido.

## 2.7. Puntos Máximos y Análisis Espectral FFT

En este punto, la señal filtrada es una curva apta para extraer cuidadosamente la información necesaria para detectar la frecuencia de pulso cardiovascular. En el desarrollo del proyecto, se propuso inicialmente la estimación mediante el cálculo único de los puntos máximos de la curva.

Mediante la función *findpeaks* de Matlab<sup>®</sup>, se halla la localización en el tiempo de los puntos máximos en la señal. Además, se calcula la diferencia entre las distancias a las que se encuentra un pico de otro y se extrae el promedio total, así:

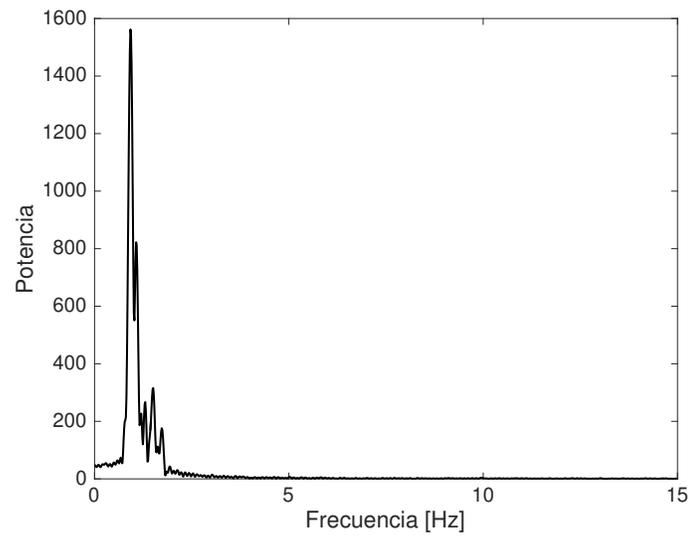
$$y_p = [y_p(2) - y_p(1) \dots y_p(m) - y_p(m - 1)] \quad (2.33)$$

$$P_{y_p} = \frac{\sum y_p}{m} \quad (2.34)$$

Donde  $y_p$  representa la trama de puntos en donde se ubican los máximos locales en el tiempo,  $m$  el tamaño del vector y  $P_{y_p}$  el valor promedio de los elementos del vector.

De la Ecuación 2.34, el valor promedio es calculado en segundos. Para encontrar el valor del ritmo cardiaco en latidos/min, simplemente se realiza un cálculo adicional, así:  $\frac{60}{P_{y_p}}$ , de esta manera se obtiene su equivalente en latidos por minuto.

Esta forma de análisis de los datos para estimar un valor de ritmo cardiaco, se apoya bastante bien cuando a la misma curva se le extrae la Transformada Rápida de Fourier FFT, quedando evaluada no sólo en su forma temporal sino también en frecuencia, por medio de un análisis espectral.



**Figura 2.19:** *Espectro de potencia de la señal PPG.*

La FFT estima la distribución de potencia de la señal PPG y se obtiene el punto de máxima frecuencia como el valor de la frecuencia cardiaca  $f_{HR}$  para la curva en estudio.

# Capítulo 3

## Evaluación del Sistema

En este capítulo se presenta el protocolo de pruebas diseñado para verificar y evaluar los resultados obtenidos mediante el uso de datos que contienen una verdad de referencia. El sistema de estimación de ritmo cardiaco presentado, utiliza como entrada videos cortos previamente grabados en diferentes escenarios para probar su desempeño.

En primer lugar se presentan los resultados de una secuencia de imágenes adquiridas en un video que toma exclusivamente la cara del sujeto. Este caso se examina como la condición más favorable para el sistema y sirve como referencia inicial de prueba. Posteriormente se evalúa el sistema para una secuencia de imágenes que presentan un fondo y en donde la cara del sujeto no lo es todo en la imagen. Además, esta prueba se realiza para un recién nacido, lo que permite extender el rango de valores del sistema, es decir, se debe calcular un valor que es incluso el doble del estimado para el primer sujeto.

Seguidamente, se realizan pruebas para videos en condiciones menos favorables, es decir con ruido de fondo y en ambientes reales, como son la habitación de una casa que presenta excesiva luminosidad del exterior, lo que dificulta la adquisición de las características de color presentes en la piel del sujeto. Otra condición de prueba real se da para un video grabado en el espacio del laboratorio sin controlar el fondo presente. En este caso las pruebas son evaluadas en videos tomados de cámaras de muy bajo costo.

## 3.1. Protocolo de Pruebas

Un protocolo de pruebas se ha diseñado para la aplicación del marco de trabajo y la verificación del mismo. Esto permite construir una idea clara para probar el sistema, el cual dado un video de entrada, debe proporcionar una medida real y aproximada del ritmo cardiaco del sujeto en estudio en ese instante de tiempo.

### 1. Adquisición de los datos

Fragmentos de video de 10 segundos, en diferentes escenarios y personas con edades y tonos de piel distintos, fueron utilizados como entrada del sistema. Se usaron varios tipos de videos para poner a prueba el sistema, la primera parte resulta de videos hechos en condiciones de entorno controladas, es decir en todo el video solo se observa el rostro del sujeto. Para ello, se descargaron 2 grabaciones presentadas por Wu *et al.* [Wu et al., 2012] y Balakrishnan *et al.* [Balakrishnan et al., 2013] para probar su trabajo mediante los cambios de color y movimiento, respectivamente. El tercer video usado es para medir el ritmo cardiaco a un bebé, en este caso la toma del video no es exclusiva del rostro sino que existe un fondo. En este caso se prueba la funcionalidad de la metodología, ya que los recién nacidos poseen un ritmo cardiaco mucho mayor al de un adulto sano. El cuarto video es tomado desde un celular, con ruido de la luz del día. Fue usado para probar la metodología expuesta por Bush [Bush, 2016] y grabado desde un celular a una velocidad de 15 fps.

Por último, se toma un video bajo una condición de prueba real, sin modificar el fondo. Este entorno es el lugar de trabajo en el Laboratorio de Visión, Robótica e Inteligencia Artificial LaViRIA. La velocidad de reproducción estándar es de 30 fps. En este trabajo, la mayoría de los videos de prueba, se reproducen a dicha velocidad.

### 2. Verificación

En este trabajo se ha realizado un análisis de varios espacios de color para determinar su desempeño en la detección de las sutiles variaciones del color en la piel, características naturales presentes en las personas, pero que no son visibles

por el ojo humano. Particularmente, el mejor espacio será aquel que presente un menor error en la medida del ritmo cardiaco para cada video de prueba.

El sistema debe estimar correctamente el valor de ritmo cardiaco que entrega la referencia. Una vez se obtienen los resultados, éstos son comparados con los valores reales y se calcula el error entre la verdad de referencia denominada  $HR_r$  y el valor medido a través del sistema propuesto  $HR_m$  de la siguiente manera:

$$Error(\%) = \left| \frac{HR_r - HR_m}{HR_r} \right| \times 100 \quad (3.1)$$

Los valores de referencia fueron obtenidos en cada caso, de los trabajos anteriores a este y se usan para medir el error de la metodología presentada.

### 3. Significancia de los resultados

Como se explicó al principio, en la sección de conceptos básicos, los adultos sanos poseen un ritmo cardiaco regular entre 50 y 100 bmp en reposo. Aunque estos rangos pueden variar según la edad, sirven como referencia para ubicarse y tener una idea de los valores esperados. Cuando se calcula el error presente en los resultados obtenidos, se espera que sea lo más cercano a cero, es decir, que el valor medido sea igual o aproximadamente igual al valor de la referencia.

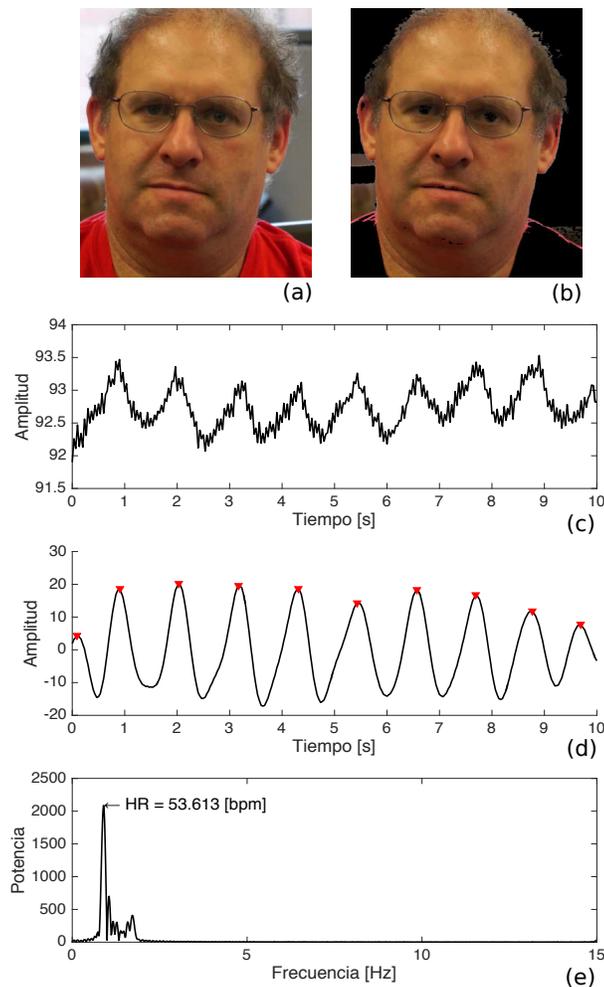
Los resultados están organizados de manera que para cada video de prueba, se estima un valor de ritmo cardiaco proveniente del promedio de picos hallados en la señal filtrada y además el que se puede estimar a partir del espectro de potencia de la señal. Esto con el fin de corroborar los datos obtenidos junto con los aportados por la referencia.

## 3.2. Pruebas y sus Resultados

En esta sección se presentan los resultados obtenidos a partir de una serie de grabaciones cortas. Cada secuencia de video analizada cumple un rol importante que pone a prueba la funcionalidad del sistema propuesto. Para cada video, se muestra una figura representativa del proceso, para una única prueba. Los demás

resultados serán expuestos más adelante sólo de forma numérica para el cálculo del error promedio.

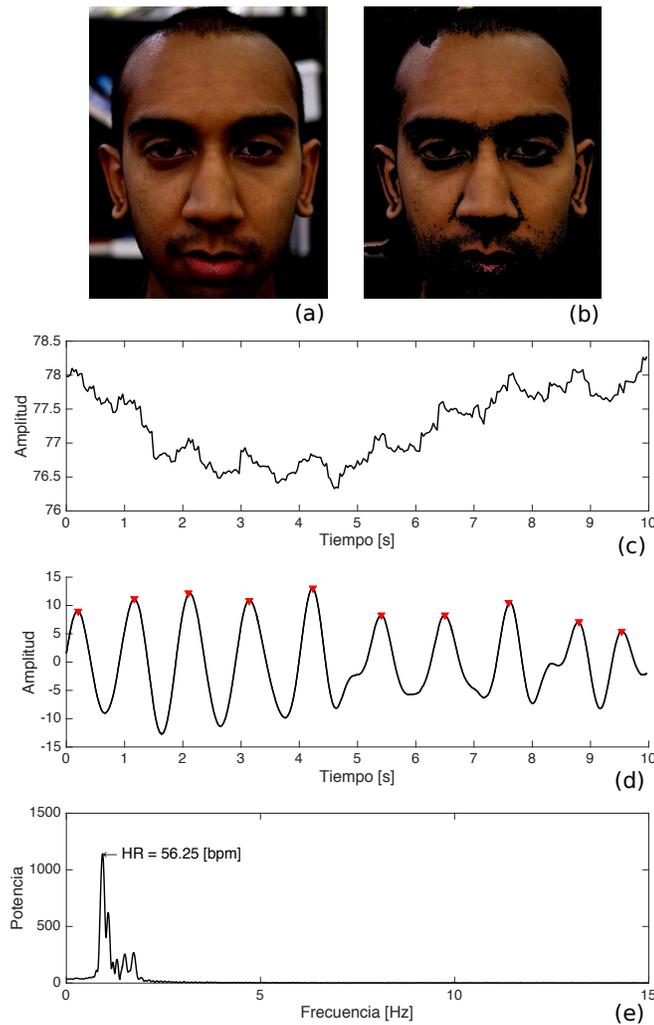
El primer video *face1*, es un video grabado en condiciones de entorno controlado. Fue presentado por Wu *et al.* [Wu et al., 2012] y sirve como base para probar el sistema en su condición más favorable. El video fue grabado a 30 fps y tiene una longitud única de 10 segundos. Para este video en particular, la curva proveniente de los promedios de intensidades de la región de interés no presenta ruido excesivo y facilita la estimación del ritmo cardiaco en cada espacio de color.



**Figura 3.1:** Resultados del video *face1*. El ritmo cardiaco de referencia es de 54 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR.

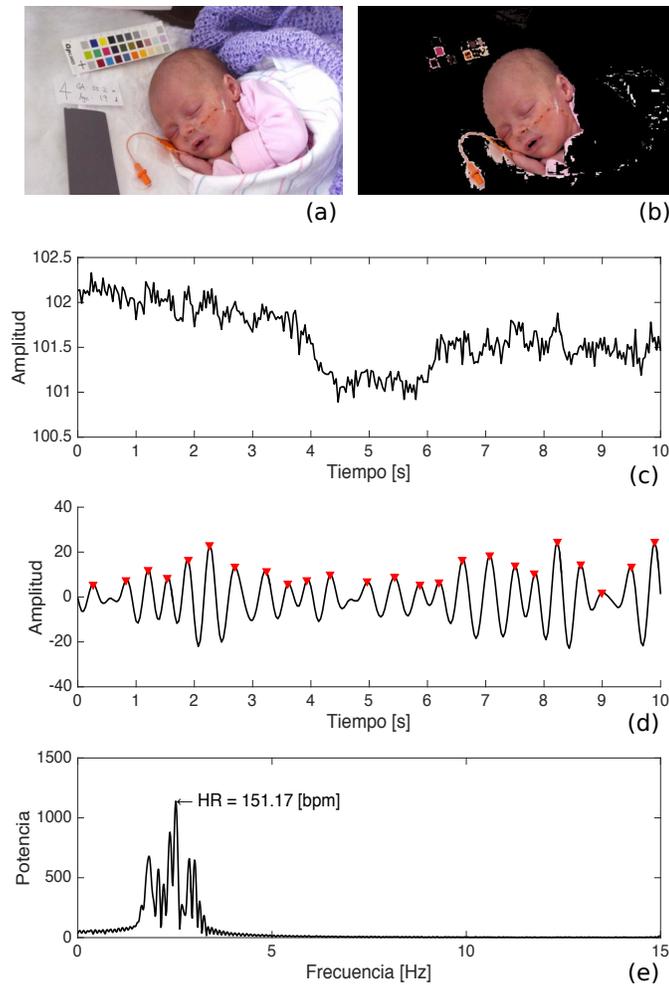
Como se observa en la Figura 3.1, el proceso seguido evalúa los cuadros extraídos del video y los segmenta para facilitar la búsqueda de las características de interés.

El segundo video *face2* que pone a prueba el sistema, se elige dado que el sujeto grabado posee un tono de piel mucho más oscura, lo que podría dificultar la etapa de segmentación. En la Figura 3.2 se muestran la señal pura extraída junto con su equivalente filtrada. La estimación obtenida, muestra que el sistema es robusto frente a diferentes tonos de piel.



**Figura 3.2:** Resultados del video *face2*. El ritmo cardiaco de referencia es de 55 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR.

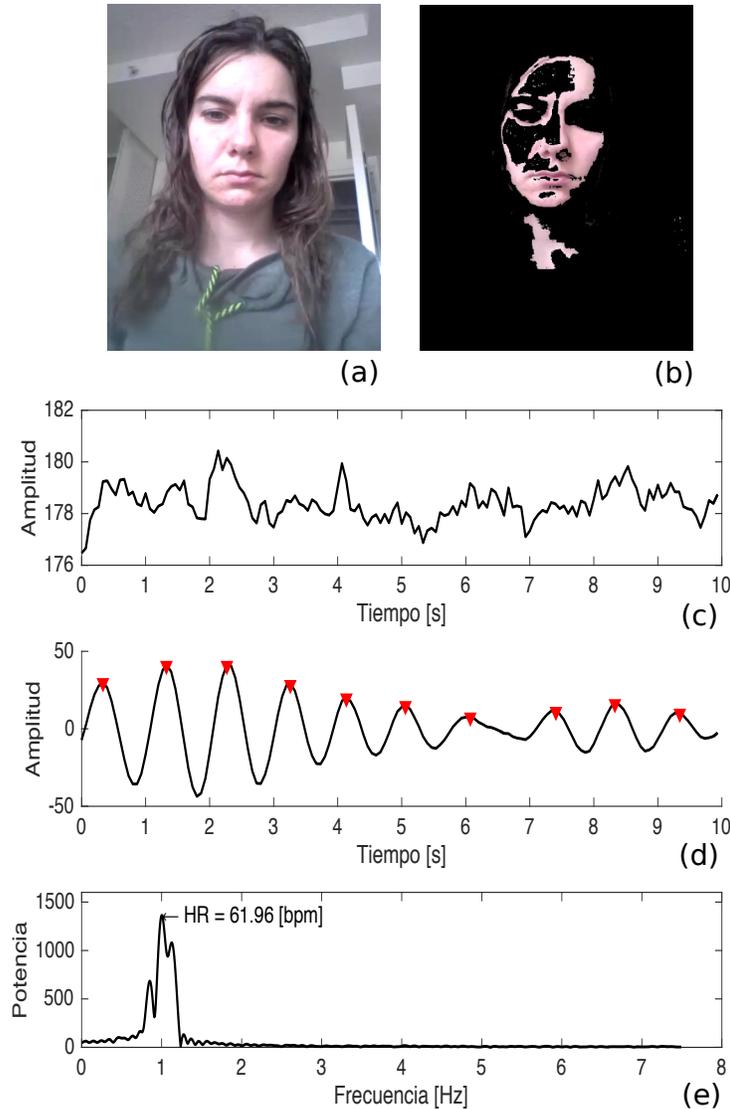
Un tercer video *baby*, es analizado con el fin de ampliar lentamente el espacio de posibilidades e inclinar la estimación a una población que puede ser ampliamente favorecida con técnicas de medición de este tipo. Se trata de la estimación de ritmo cardiaco en bebés o recién nacidos. La Figura 3.3 muestra las gráficas obtenidas para los resultados de esta secuencia en la que se ha grabado a un bebé mientras duerme.



**Figura 3.3:** Resultados del video *baby*. El ritmo cardiaco de referencia es de 152 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR.

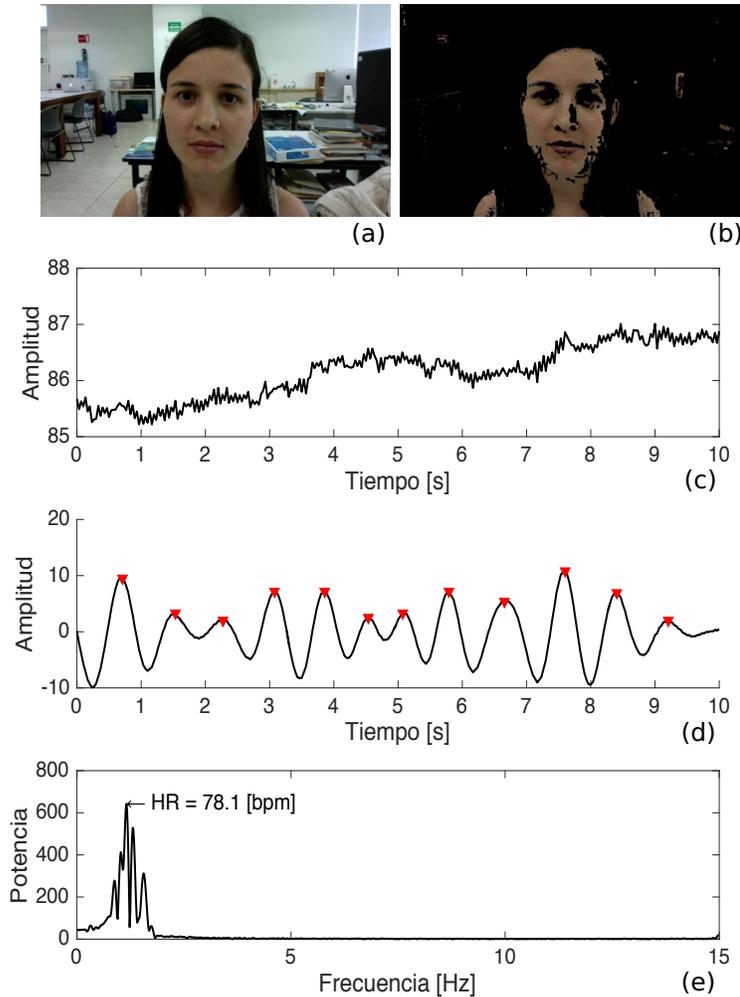
Hasta este punto, los videos probados han mostrado el comportamiento esperado en la aplicación del sistema, aun cuando las entradas pertenezcan al grupo de videos en los que no existen mayores variaciones de luz y de fondo.

El cuarto video usado *Bush*, es presentado bajo una condición de ruido aun mayor, particularmente por el exceso de iluminación en la escena. Además, la resolución del video siendo de 15 fps limita al sistema en la cantidad de información disponible por unidad de tiempo. Es decir, para construir la señal proveniente de los promedios de las intensidades de color, se tiene un menor número de muestras en el tiempo.



**Figura 3.4:** Resultados del video *Bush*. El ritmo cardiaco de referencia es de 60 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR.

El último video usado, corresponde a una grabación realizada en LaViRIA. El escenario dispuesto representa el lugar de trabajo habitual, lo que constituye un ambiente de prueba real. Para este caso se utilizó como verdad de referencia un medidor de ritmo cardiaco proporcionado por el centro de acondicionamiento físico de la DICIS-UG, el cual corresponde a un Monitor RC FT1 de la marca Polar.



**Figura 3.5:** Resultados del video Gesley. El ritmo cardiaco de referencia es de 80 bpm. (a) Entrada de video. (b) Detección de piel en la secuencia de imágenes. (c) Señal unidimensional pura  $Y_G$ . (d) Señal PPG después del proceso de filtrado y amplificación, con detección de picos. (e) Espectro de potencia de la señal y su valor de HR.

Todos los resultados obtenidos para cada video son organizados en la Tabla 3.1, y corresponden a la medida de pulso cardiovascular para cada espacio de color, calculado a partir del número de picos detectados en la curva, así como su

respectivo porcentaje de error.

Video	Referencia [bpm]	HLS (%e)	HSV (%e)	Munsell (%e)
face1	54	56(3.7)	55(1.8)	55(1.8)
face2	55	58(5.4)	57(5.6)	58(5.4)
baby	152	148(2.6)	144(5.3)	150(1.3)
Bush	60	64(6.7)	66(10)	65(8.3)
Gesley	80	78(2.5)	83(3.7)	85(6.2)

**Tabla 3.1:** *Estimación de ritmo cardiaco por número de picos detectados y su porcentaje de error.*

Dado que también se analizó el espectro de potencia de las señales, se presenta un conjunto adicional de datos que contiene la estimación del ritmo cardiaco debido al pico de frecuencia máxima. Los resultados se muestran en la Tabla 3.2 junto con su respectivo porcentaje de error.

Video	Referencia [bpm]	HLS (%e)	HSV (%e)	Munsell (%e)
face1	54	54(0)	54(0)	54(0)
face2	55	56(1.8)	56(1.8)	56(1.8)
baby	152	142(6.6)	151(0.6)	152(0)
Bush	60	63(5)	61(1.7)	62(3.3)
Gesley	80	78(2.5)	82(2.5)	83(3.7)

**Tabla 3.2:** *Estimación de ritmo cardiaco mediante la FFT y su porcentaje de error.*

# Conclusiones

En este trabajo se presentó un sistema de estimación de ritmo cardiaco sin contacto a partir del análisis de secuencias de imágenes. Los procesos que constituyen el pilar central de la metodología están basados en algoritmos de bajo costo computacional, permitiendo que sean utilizados en el procesamiento de imágenes extraídas a partir de secuencias de video. Adicionalmente, el sistema completo puede ser adaptado para distintos espacios de color dependiendo del caso de estudio, comprendiendo HLS, HSV y Munsell.

La metodología propuesta tiene la capacidad de identificar y cuantificar de forma confiable el HR de un sujeto a partir de la generación de señales PPG, con un margen de error menor 6.6 % para el caso de la estimación a partir de la FFT. El banco de pruebas usado permite validar el correcto funcionamiento del sistema para la medición del HR en un ambiente no controlado y con el uso de cámaras de video de bajo costo, las cuales comprenden el nivel de consumo doméstico de la mayoría de usuarios potenciales.

El uso de conceptos sencillos y algoritmos de bajo costo computacional, hace que el sistema aquí propuesto difiera positivamente en complejidad y utilización de recursos con respecto a muchas de las metodologías reportadas actualmente. De esta manera, se ofrece una alternativa que, sin sacrificar confiabilidad y funcionalidad, aborda de manera más directa y simple el problema de la medición de HR sin contacto, de forma remota y con un costo mínimo.

Dado que el sistema se probó con distintos espacios de color, la metodología propuesta ofrece una alta flexibilidad al momento de especificar las características del tono de piel de los sujetos bajo observación. Además el tiempo necesario para la configuración inicial del sistema es mínimo en comparación a procesos de entrenamiento y/o protocolos de calibración presentes en otras alternativas. Sin

embargo, el uso de diferentes espacios de color no ofrece un único mejor resultado en las pruebas realizadas. Esto puede deberse a que los videos no son grabados con el mismo equipo y la calidad de la grabación varía.

## Trabajo Futuro

El trabajo desarrollado provee ciertas área de oportunidad sobre las cuales se puede expandir y profundizar el alcance del sistema desarrollado:

- *Optimizar los algoritmos para su uso en tiempo real:* Dada la naturaleza de las aplicaciones que podría tener el estudio aquí realizado, sería conveniente optimizar su código fuente para que el sistema completo pueda acoplarse a esquemas de medición y monitoreo en tiempo real. Actualmente, el sistema propuesto no cumple con las restricciones necesarias para poder trabajar de manera ininterrumpida sobre una señal de video y registrar (de manera sincronizada y simultánea) las mediciones de HR, que serían necesarias para ofrecer un monitoreo continuo del sujeto bajo observación. Sin embargo, las partes componentes del diseño pueden ser reestructuradas de manera más eficiente para lograr que los datos registrados sean procesados *online* y retornen una lectura inmediata y continua en tiempo real.
- *Validar y expandir los resultados obtenidos a partir de bancos de prueba con bases de datos reales:* Acceder a un mayor número de secuencias de video, en donde varíen la naturaleza del sujeto en observación y su entorno, así como realizar estudios a partir de otros espacios de color utilizados en el campo de procesamiento de imágenes, serviría para hacer aun más robusto el diseño propuesto y proporcionar un análisis estadístico más completo acerca del desempeño del sistema propuesto. La información recolectada de estas pruebas adicionales sería útil para ajustar de manera más precisa los parámetros del sistema e identificar posibles mejoras sobre el diseño evaluado aquí.
- *Incrementar el número de índices útiles que el sistema puede proporcionar:* Dada la naturaleza modular del sistema propuesto y la cantidad de información recolectada por el mismo que no se procesa en su totalidad, es posible añadir etapas extra de procesamiento y cálculos sencillos que utilicen

las señales internas del proceso para hacer el cálculo de otros indicadores vitales útiles para quien eventualmente realice un diagnóstico médico sobre el sujeto bajo observación. Mientras dichos cálculos no supondrían un uso significativo de recursos adicionales, índices como la presión arterial o la identificación de patrones extraños en las curvas de PPG aportarían información adicional que favorecería una mejor evaluación sobre el estado de salud actual del usuario.

# Bibliografía

- [Allen, 2007] Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1.
- [Balakrishnan et al., 2013] Balakrishnan, G., Durand, F., and Guttag, J. (2013). Detecting pulse from head motions in video. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3430–3437.
- [Bernacchia et al., 2014] Bernacchia, N., Scalise, L., Casacanditella, L., Ercoli, I., Marchionni, P., and Tomasini, E. P. (2014). Non contact measurement of heart and respiration rates based on kinect<sup>TM</sup>. In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, pages 1–5. IEEE.
- [Bush, 2016] Bush, I. (2016). Measuring heart rate from video. Stanford Computer Science.
- [Casati et al., 2013] Casati, J. P. B., Moraes, D. R., and Rodrigues, E. L. L. (2013). Sfa: A human skin image database based on feret and ar facial images. In *IX workshop de Visao Computational, Rio de Janeiro*.
- [Fairchild, 2013] Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.
- [Farreras and Rozman, 2000] Farreras, P. and Rozman, C. (2000). *Medicina interna*, vol ii.
- [Garbey et al., 2007] Garbey, M., Sun, N., Merla, A., and Pavlidis, I. (2007). Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418–1426.

- [Hertzman and Dillon, 1940] Hertzman, A. B. and Dillon, J. B. (1940). Applications of photoelectric plethysmography in peripheral vascular disease. *American Heart Journal*, 20(6):750–761.
- [Humphreys et al., 2007] Humphreys, K., Ward, T., and Markham, C. (2007). Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry. *Review of scientific instruments*, 78(4):044304.
- [Hyun et al., 2015] Hyun, D., Trahey, G. E., and Dahl, J. J. (2015). Real-time high-framerate in vivo cardiac slsc imaging with a gpu-based beamformer. In *Ultrasonics Symposium (IUS), 2015 IEEE International*, pages 1–4. IEEE.
- [Kakumanu et al., 2007] Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106 – 1122.
- [Kumar et al., 2015] Kumar, M., Veeraraghavan, A., and Sabharwal, A. (2015). DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomedical Optics Express*, 6(5):1565.
- [Kurylyak et al., 2013] Kurylyak, Y., Lamonaca, F., and Grimaldi, D. (2013). A neural network-based method for continuous blood pressure estimation from a ppg signal. In *Instrumentation and Measurement Technology Conference (I2MTC), 2013 IEEE International*, pages 280–283. IEEE.
- [Kwon et al., 2012] Kwon, S., Kim, H., and Park, K. S. (2012). Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2174–2177. IEEE.
- [Lam and Kuno, 2016] Lam, A. and Kuno, Y. (2016). Robust heart rate measurement from video using select random patches. *Proceedings of the IEEE International Conference on Computer Vision*, 11-18-Dece:3640–3648.
- [Li et al., 2014] Li, X., Chen, J., Zhao, G., and Pietikäinen, M. (2014). Remote heart rate measurement from face videos under realistic situations. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 4264–4271.

- [McCamy, 1992] McCamy, C. (1992). Munsell value as explicit functions of cie luminance factor. *Color Research & Application*, 17(3):205–207.
- [Osman et al., 2012] Osman, G., Hitam, M. S., and Ismail, M. N. (2012). Enhanced skin colour classifier using rgb ratio model. *International Journal on Soft Computing*, 3(4):1–14.
- [PHILIPS, 2011] PHILIPS, P. V. S. C. (2011). <http://www.vitalsignscamera.com>.
- [Poh et al., 2010a] Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2010a). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774.
- [Poh et al., 2011] Poh, M.-Z., McDuff, D. J., and Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11.
- [Poh et al., 2010b] Poh, M.-Z., Swenson, N. C., and Picard, R. W. (2010b). Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography. *IEEE Transactions on Information Technology in Biomedicine*, 14(3):786–794.
- [Proakis and Manolakis, 2007] Proakis, J. and Manolakis, D. (2007). *Tratamiento digital de señales*. Pearson Educación.
- [Rhee et al., 2001] Rhee, S., Yang, B.-H., and Asada, H. H. (2001). Artifact-resistant power-efficient design of finger-ring plethysmographic sensors. *IEEE Transactions on Biomedical Engineering*, 48(7):795–805.
- [Verkruysse et al., 2008] Verkruysse, W., Svaasand, L. O., and Nelson, J. S. (2008). Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445.
- [Vezhnevets et al., 2003] Vezhnevets, V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon*, volume 3, pages 85–92. Moscow, Russia.
- [Viola and Jones, 2004] Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, 57(2):137–154.

- [Weinman et al., 1977] Weinman, J., Hayat, A., and Raviv, G. (1977). Reflection photoplethysmography of arterial-blood-volume pulses. *Medical and Biological Engineering and Computing*, 15(1):22–31.
- [Wu et al., 2012] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, 31(4).
- [Yu et al., 2015] Yu, Y.-P., Raveendran, P., and Lim, C.-L. (2015). Dynamic heart rate measurements from video sequences. *Biomedical optics express*, 6(7):2466–2480.
- [Zeng et al., 2015] Zeng, W., Zhang, Q., Zhou, Y., Xu, G., and Liang, G. (2015). Infrared video based non-invasive heart rate measurement. In *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1041–1046. IEEE.
- [Zhang et al., 1998] Zhang, J., Sokhansanj, S., Wu, S., Fang, R., Yang, W., and Winter, P. (1998). A transformation technique from rgb signals to the munsell system for color analysis of tobacco leaves. *Computers and electronics in agriculture*, 19(2):155–166.