

Comparación de Técnicas de Procesamiento de Video en Diferentes Espacios de Color para la Estimación de la Frecuencia Cardíaca

D. González Uranga^{1*}, J. D. Díaz Román¹, J. Cota Ruiz¹, B. Mederos¹, L. Rascón¹, J. Mejía¹

¹Instituto de Ingeniería y Tecnología, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, México

*goudaniela98@gmail.com

Abstract— This study aims to compare different analysis techniques to determine which of these provides better results in estimating heart rate through video processing. The techniques, in general, consisted in the extraction of signals of color intensity variation throughout the video on a region of interest (ROI) of the face, to then determine the dominant frequency of these signals within a possible range of frequencies. Three ROIs were evaluated to find out which region provides more information on cardiac activity: the forehead, the zygomatic region and the area of the face that ranges from the lower part of the eyes to the chin (polygon area). On the other hand, this project also seeks to determine which of the following color spaces allows to obtain better results: RGB, YUV, HSI, HSV and CMY. The best result was obtained when applied the Fourier Transform to the signal extracted from the green channel (of the RGB color space) of the video images, with a root mean square error of 1.5 beats per minutes and a Pearson correlation coefficient of 0.993.

Palabras clave— Espacios de color, frecuencia cardíaca, procesamiento de video, técnicas de procesamiento.

I. INTRODUCCIÓN

Conocer ciertos parámetros fisiológicos del cuerpo humano ayuda al personal de salud a brindar un mejor diagnóstico, por lo que constantemente se busca desarrollar nuevas tecnologías que faciliten estos procedimientos y que, a su vez, generen resultados confiables.

La estimación de la frecuencia cardíaca (FC) es uno de estos parámetros, la cual corresponde al número de contracciones del corazón por minuto [1]. Estimar la frecuencia cardíaca puede llegar a ser incómodo para los usuarios en ciertas ocasiones, por ejemplo, para aquellos que presentan quemaduras o sensibilidad en la piel, debido a que normalmente se ponen sensores sobre la piel del sujeto y estos pueden causar irritación [2].

Gracias a los avances en las técnicas de procesamiento de señales e imágenes, es posible tomar la frecuencia cardíaca por medio del procesamiento de video. Las técnicas que se han propuesto hasta ahora para determinar la FC por medio de video comúnmente sufren de errores consecuentes a distintos factores como el espacio de color seleccionado, la región de interés (ROI, por sus siglas en inglés), algoritmo de identificación o seguimiento facial [3] [4]. En este proyecto se realizan una serie de comparaciones entre distintas técnicas de análisis para estimar la frecuencia cardíaca por medio del procesamiento de video, con el

objetivo de poder determinar cuál de estas brinda mejores resultados y sea posible emplear en aplicaciones clínicas.

II. METODOLOGÍA

En este apartado se describe el procedimiento que se llevó a cabo empezando por: la creación de la base de datos; el procesamiento de las imágenes del video para la detección del rostro, la selección de ROI, seguimiento facial, transformación de imágenes a diferentes espacios de color; y terminando con la obtención de las señales de interés y estimación de la frecuencia cardíaca. En esta sección también se describe el análisis estadístico que se utilizó para analizar y comparar los resultados obtenidos para cada uno de los espacios de color, regiones de interés y procesamiento empleado.

A. Base de datos

Se desarrolló una base de datos que incluyó a 16 sujetos de prueba. De cada uno de los participantes se cuenta con: información del participante como peso, estatura, tono de piel y si tiene barba y/o bigote.

Los videos fueron grabados utilizando la cámara web de una *laptop* empleada en el estudio, tomados a 30 fotogramas por segundo a una resolución de 1280p x 720p y se obtuvo una totalidad de 15 segundos por cada participante. Durante la prueba, los sujetos se mantuvieron sentados en estado de relajación y la cámara fue colocada frente al sujeto a 70 cm (aproximadamente) de los sujetos. Para la toma de la señal ECG se utilizó una tarjeta de adquisición de datos National Instruments USB-6009, un sensor EKG-BTA Vernier y electrodos de espuma desechables con hidrogel adhesivo-conductivo. Se desarrolló un programa en LabVIEW [5] para adquirir simultáneamente el video y la señal de ECG del sensor.

A partir de la señal de ECG se determinó la frecuencia cardíaca de cada participante con la finalidad de servir como patrón al momento de comparar con el valor de FC estimada a partir de los videos. Se empleó el software de programación MATLAB versión R2019b, para procesar las señales de ECG y los videos de la base de datos.

B. Detección del rostro

Puesto que la región de interés se encuentra en la cara, se procedió a detectar el rostro en el primer fotograma del video implementando el algoritmo de Viola-Jones [6]. Así mismo, se aplicó el algoritmo propuesto por A. Asthana el

cual es un algoritmo que consiste en la combinación de un modelo local restringido y un ajuste del mapa de respuesta discriminativa [7]. Este permite identificar automáticamente las coordenadas de 66 puntos de interés en el rostro, incluyendo las cejas, contorno de los ojos, boca nariz y el mentón. En la Fig. 1 se muestran el recuadro del rostro y los 66 puntos detectados.

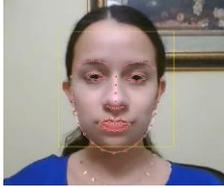


Fig. 1. Detección del rostro y de los 66 puntos de interés para el primer fotograma del video, usando los algoritmos de Viola Jones y Asthana.

C. Regiones de interés a analizar

Para conocer el impacto y diferencias existentes al estimar la frecuencia cardiaca en distintas zonas del rostro. Se definieron tres regiones de interés. Las regiones elegidas fueron la frente, las mejillas y una región poligonal que abarca desde la parte inferior de los ojos hasta la barbilla. Fue posible definir las por medio de las coordenadas obtenidas por los algoritmos anteriores. En la Fig. 2 se muestra la detección de la región poligonal, la frente y la cigomática de la derecha.



Fig. 2. Imágenes de las regiones de interés analizadas: región poligonal (izquierda), la frente (centro) y la cigomática de un lado del rostro (derecha).

D. Seguimiento del rostro

Ya que la detección de los puntos de interés se aplicó únicamente en el primer fotograma del video, se continuó con la aplicación del algoritmo de Lucas–Kanade [8]. Se trata de un método diferencial para realizar el seguimiento de puntos específicos en una secuencia de imágenes. De esta manera, se obtuvo la información en cada región de interés a lo largo del video. En la Fig. 3 se muestra el seguimiento de los puntos pertenecientes a la región de interés de la frente con el sujeto de prueba en movimiento.



Fig. 3. Seguimiento de la región de la frente con el sujeto en movimiento.

E. Obtención de señal en el espacio de color RGB

Para este paso se promedió la información de todos los píxeles ubicados dentro de cada región de interés (por separado) para cada uno de los canales del espacio de color RGB en cada fotograma del video. De esta manera se construyó una señal cuya duración es la misma que la del video (15 s), donde cada muestra representa el valor promedio de la intensidad de los píxeles de la región de interés ubicada en cada fotograma del video. El tiempo entre cada muestra viene dado por el tiempo entre los fotogramas del video, es decir, 1/30 s. En la Fig. 4 se presenta un ejemplo de las señales de los tres canales obtenidos de la región de la frente de un video.

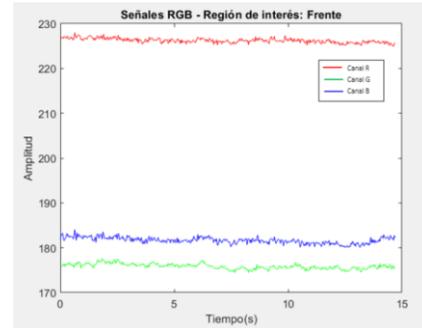


Fig. 4. Señal R, G y B para la región de interés frente en el sujeto de prueba.

F. Obtención de las señales de otros espacios de color

Para este caso, se llevó a cabo una transformación del formato de imagen RGB a los formatos YUV, HSI, HSV y CMY, aplicado a cada fotograma del video. Se emplearon ecuaciones específicas para obtener las transformaciones de cada canal y, por medio del procedimiento descrito en la sección E, se obtuvieron las señales de variaciones de intensidad correspondientes a cada canal de los formatos mencionados.

G. Análisis en el dominio de la frecuencia y estimación de la frecuencia cardiaca

Debido a que el objetivo es estimar la frecuencia cardiaca, realizar análisis en el dominio de la frecuencia fue clave para obtener los resultados esperados. Previo a realizar el análisis en frecuencia, las señales se filtraron empleando un filtro pasa-banda tipo Butterworth de orden 10, con frecuencias de corte entre 0.7 Hz y 4 Hz. Se eligieron estas frecuencias de corte ya que este rango abarca el rango de FC fisiológicamente posible en un ser humano.

Para la estimación de la FC se aplicaron tres técnicas distintas. La primera consistió en la Transformada Discreta de Fourier (DFT) a cada una de las señales de variación de intensidad de color (canal) de cada formato; en la segunda se aplicó el Análisis de Componentes Principales (PCA) a las señales y posterior a ellos se aplicó la DFT a cada componente resultante; para la tercera técnica se procedió de la misma forma que en la segunda, pero se aplicó el Análisis

de Componentes Independientes (ICA) en lugar de PCA a las señales respectivas.

Para estimar la FC se localizó el componente de frecuencia con mayor amplitud en el espectro obtenido por la aplicación de la DFT a cada señal o componente de éstas. Este valor de frecuencia (en Hz) se multiplicó por 60 para obtener el valor en latidos por minuto (LPM).

En la Fig. 5 se muestra un ejemplo donde se aplicó la DFT a las señales de los canales del espacio RGB obtenidas en la región de la frente de un sujeto de prueba. Se observan tres valores de frecuencia (X) cuya amplitud de la componente espectral (Y) es máxima para cada canal del formato RGB.

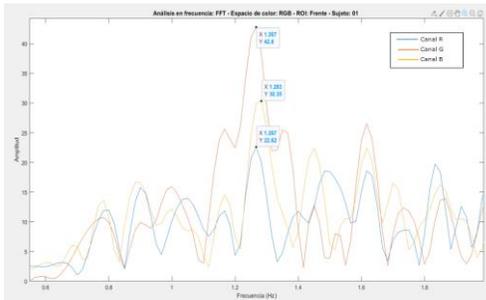


Fig. 5. Análisis en frecuencia de las señales correspondientes a los canales R, G, B de la región de la frente de un sujeto. Se señalan los valores de frecuencia (X) cuya amplitud (Y) es la máxima de cada análisis.

Para los casos de la estimación empleando PCA e ICA, la FC se obtuvo determinando solo el valor de frecuencia con mayor amplitud entre las tres nuevas componentes que resultaran del análisis de PCA e ICA respectivamente. Esto debido a que se espera que solo una de las componentes esté correlacionada con la actividad cardíaca.

H. Análisis estadístico

Al combinar las tres técnicas, con los cinco espacios de color y sus respectivos canales, y las tres regiones de interés, se obtuvieron 75 resultados de estimación para cada participante. Para evaluar los resultados obtenidos, se calculó la raíz del error cuadrático medio (RMSE) y el coeficiente de correlación de Pearson. Los datos de frecuencia cardíaca obtenidos para todos los escenarios fueron comparados con el valor patrón del ECG. Se evaluaron los resultados por métodos gráficos, empleando gráficos de dispersión y gráficos de Bland-Altman, para observar la posible relación de la discrepancia entre las estimaciones y el valor esperado.

III. RESULTADOS

En la Tabla 1 se presentan los mejores resultados obtenidos sobre todas las combinaciones, donde se especifica la técnica aplicada, la ROI, el espacio de color implementado, el canal de color (de ser el caso), el error RMSE (en LPM) y, el coeficiente de Pearson. Entre menor sea el RMSE, mejor será el método por evaluar. Así mismo, cuando el valor del coeficiente de correlación es

exactamente 1.0 significa que la correlación entre ambos métodos (el propuesto vs. el valor patrón) es perfecta. Tomando esto en cuenta, se puede observar que los coeficientes mostrados en la Tabla 1, producto de las estimaciones para las distintas combinaciones, presentan confiables.

TABLA I
ERROR RMSE Y C. DE PEARSON PARA LOS MEJORES RESULTADOS

Técnica	ROI	Espacio de color	Canal	RMSE (LPM)	Coefficiente Pearson
DFT	Frente	YUV	V	3.85	0.951
		RGB	G	3.85	0.951
	Mejillas	YUV	V	3.35	0.958
		RGB	G	1.5	0.993
ICA/ DFT	Polígono	YUV	Y	2.71	0.97
		RGB	-	3.79	0.94
PCA/ DFT	Polígono	YUV	-	3.79	0.94
		YUV	-	2.69	0.97

La combinación que obtuvo el menor error con un valor de 1.5 LPM, fue la aplicación de la DFT a la señal extraída de la región poligonal sobre el canal G del espacio RGB. De igual forma, esta combinación fue la que obtuvo un coeficiente de Pearson mayor, con un valor de 0.993, obteniendo una casi perfecta correlación respecto al valor esperado. Se analizaron y compararon los gráficos de dispersión y de Bland-Altman para las combinaciones mostradas en la Tabla 1 con la finalidad de poder llevar un análisis visual y corroborar los resultados brindados por ecuaciones. En la Fig. 6 se muestra el gráfico de correlación para la combinación con coeficiente de 0.993, donde se puede apreciar visualmente la relación prácticamente lineal entre ambos métodos. Así mismo, en la Fig. 7 se muestra el gráfico Bland-Altman para la misma combinación.

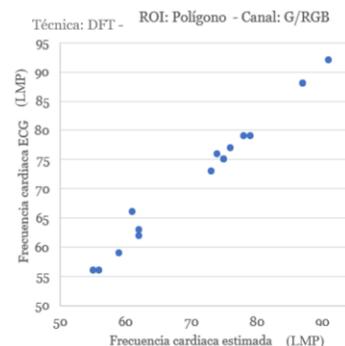


Fig. 6. Gráfico de dispersión para la mejor combinación.

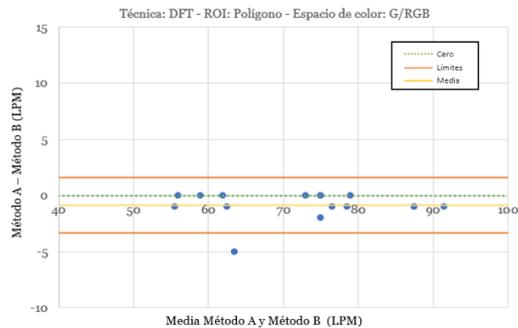


Fig. 7. Gráfico de Bland-Altman para la mejor combinación.

En un gráfico de Bland-Altman, el eje de las Y representa la diferencia entre los métodos a comparar: el resultado patrón brindado por el ECG y el resultado estimado por una combinación empleada. Mientras que el eje de las X muestra el valor de la media de ambos valores. Los puntos azules representan los resultados estimados para cada uno de los sujetos. Así mismo, la recta amarilla representa la media para todos los valores en el eje Y. En la Fig. 7 se puede observar la mayoría de los datos concentrados alrededor del cero y la media con un valor de -0.87 LPM. Este gráfico se caracteriza por establecer límites superior e inferior los cuales abarcan el 95% de los resultados, en consecuencia, entre más cercanos al cero se encuentren los límites, mejor será la estimación obtenida por el método evaluado. En este caso, estos límites se encuentran en el rango [-3.34 LPM, 1.6 LPM], lo que indica que la diferencia entre los valores estimados y los verdaderos son muy pequeñas, otorgándole al método evaluado con la gráfica una gran confiabilidad.

IV. DISCUSIÓN

Con base a los resultados presentados es posible determinar que los coeficientes de mayor valor para cada técnica pertenecen a la región de interés del área poligonal, indicando que de las tres ROI esta es la mejor. Esto se debe principalmente a que es la región que toma más información del rostro. Por otra parte, los espacios de color que se destacaron en la Tabla 1 fueron el RGB y el YUV.

Se analizaron los datos que presentaron una mayor desviación en los gráficos de dispersión para las mejores combinaciones, con la finalidad de comprender las causas de dichas desviaciones. Se encontró que los sujetos que presentan mayor desvío al usar el espacio RGB son aquellos que tienen un tono de piel más oscuro comparados con el resto de los participantes. Sin embargo, las desviaciones presentes al evaluar el canal V de YUV, se obtuvieron excelentes resultados para dichos sujetos de tonos oscuros, y por el contrario, fue el sujeto con tez más clara el que presentó mayor desviación.

Adicionalmente, se revisaron los resultados para los sujetos con bigote o barba, con el propósito de poder determinar si este factor pudiera afectar la estimación de

manera significativa. Sorprendentemente, no fueron estos sujetos los que presentaron mayor desviación. Esto nos indica que al usar la región poligonal se tienen los suficientes datos en el resto del rostro que estos factores no son un impedimento para obtener una buena estimación de la FC. Cabe destacar que la barba y bigote en los sujetos no fue abundante por lo que podría realizarse un futuro estudio. Así mismo, se recomienda realizar un estudio con *wavelets* para un distinto análisis frecuencial.

V. CONCLUSIONES

Estimar la FC por medio del procesamiento de video abre las puertas a nuevas tecnologías médicas. Establecer una metodología sencilla y efectiva para estimar este parámetro es punto clave, razón por la cual se ha desarrollado este proyecto. Como bien se mencionó, la combinación del uso de la DFT, con la ROI del polígono y el canal G perteneciente al espacio de color RGB, arrojó los mejores resultados sobre los 75 escenarios evaluados para cada uno participante. Esta afirmación se puede reforzar analizando los coeficientes de correlación de Pearson, el error RMSE y ambos gráficos.

Es importante mencionar que alcanzar este nivel de correlación y efectividad no fue tan sencillo, ya que, de las 75 combinaciones, únicamente el 16% de estas obtuvo un coeficiente de correlación mayor a 0.9, y solo una alcanzó un coeficiente de 0.99. Finalmente, este proyecto contribuye al desarrollo de futuras aplicaciones médicas donde se busque estimar la frecuencia cardiaca por medio de video. Con base a los resultados finales, es posible elegir una combinación de técnicas, regiones y espacios de color que brinde buenos resultados y pueda ser aplicada según sea conveniente en distintos dispositivos.

REFERENCIAS

- [1] J. E. Hall and A. C. Guyton, "El corazón", Tratado de fisiología médica, 12a ed. Barcelona: Elsevier, 2016, pp. 100-113.
- [2] J. L. Varela-Benítez, J. O. Rivera-Delgado, J. H. Espina-Hernández, and J. M. De La Rosa-Vázquez, "Electrodo capacitivo de alta sensibilidad para la detección de biopotenciales eléctricos," Rev. Mex. Ing. Biomed., vol. 36, no. 2, p. S183S, 2015, doi: 10.17488/RMIB.36.2.1
- [3] M. Poh, D. McDuff, and R. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Opt. Express, vol. 18, no. 10, p. 10762, 2010, DOI: 10.1364/oe.18.010762.
- [4] T. Pursche, J. Krajewski, and R. Moeller, "Video-based heart rate measurement from human faces," Dig. Tech. Pap. IEEE Int. Conf., no. January, pp. 544-545, 2012, doi: 10.1109/ICCE.2012.6161965.
- [5] National Instruments Corp. "LabVIEW Community Edition", 2020 <https://www.ni.com/en-us/shop/labview/select-edition/labview-community-edition.html>
- [6] M. Jones and P. Viola, "Robust Real-time Object Detection", International journal of computer vision, 4(34-47), 4, 2001.
- [7] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3444-3451, 2013, DOI: 10.1109/CVPR.2013.442.
- [8] F. Abdat, C. Maaoui, and A. Pruski, "Real time facial feature points tracking with Pyramidal Lucas-Kanade algorithm," Proc. 17th IEEE

- Int. Symp. Robot Hum. Interact, pp. 71–76, 2008. DOI: 10.1109/ROMAN.2008.4600645.
- [9] L. Porras, A. Jiménez, and N. Castañeda, “Use of Independent Component Analysis in the extraction of Artifacts of the Mismatch Negativity Response,” vol. 38, no. 2, pp. 420–436, 2017, DOI: 10.17488/RMIB.38.2.1.
- [10] J. Vega and M. Reyes, “Transformaciones lineales y no lineales para espacios de color en procesamiento de imágenes,” *Ing. Rev. Int.*, vol.22, no. 3, pp. 223–240, 2006.
- [11] T. Chai and R. Draxler, “Root mean square error or mean absolute error?” *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, DOI: 10.5194/gmd-7-1247-2014.
- [12] E. Lahura, “El Coeficiente de correlación y correlaciones espúreas,” *Univ. Catol. del Perú*, pp. 1–64, 2003.
- [13] Cristina Gil Martínez, “Análisis gráfico de Bland-Altman,” 2017. https://rstudio-pubs-static.s3.amazonaws.com/326677_ed20688ec0b14e39986fc4afe0144748.html (Accessed Apr. 14, 2021).