

ESTIMACIÓN DEL GRADO DE APERTURA DE LOS LABIOS MEDIANTE TÉCNICAS DE SEGMENTACIÓN DE IMÁGENES

JHORMAN ANDREY LANDAZABAL RODRÍGUEZ
ERIK LEONARDO CONTRERAS ROPERO



UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA
2017

ESTIMACIÓN DEL GRADO DE APERTURA DE LOS
LABIOS MEDIANTE TÉCNICAS DE SEGMENTACIÓN DE
IMÁGENES

JHORMAN ANDREY LANDAZABAL RODRÍGUEZ
ERIK LEONARDO CONTRERAS ROPERO

*Trabajo de grado para optar al título de
Ingeniero Electrónico*

Director
FRANKLIN ALEXANDER SEPULVEDA SEPULVEDA
PhD Ingeniería-Automática

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
BUCARAMANGA
2017

*Dedicado a
mi madre, Zulma Yohana por su apoyo incondicional, cariño y paciencia, los llamados de atención y correcciones que me brinda, las cuales me permiten ser mejor persona.*

mi padre, Orlando por enseñarme a ser constante en el trabajo, la importancia del amor de hogar y forjar mi carácter.

mi hermana, Lesly Yisset por su amistad y apoyo constante. A mis hermanas menores las Trillizas que tanto quiero

esa persona especial, que me brindó su apoyo incondicional, me ayudó a ser mejor en lo cotidiano y en lo académico, hoy solo me queda darte mi más sincero agradecimiento.

Por último, a mi compañero de proyecto por su paciencia, colaboración y las múltiples onces que me gastó, a los Ingenieros Elitistas por sus aportes académicos y las constantes recochas.

Jhorman Andrey Landazabal Rodríguez

Dedicado a

mi padre, Delfin Contreras Contreras y mi madre, Esther Ropero Rojas, por sus enseñanzas, apoyo y amor incondicional durante todos estos años, que me permiten seguir adelante como profesional y persona. Sus lecciones y experiencia de vida me han hecho entender que cualquier meta se logra con trabajo duro.

mis compañeros y amigos, por todas esas arduas horas de trabajo y momentos especiales. Gracias a ustedes pude aprender cosas nuevas, experiencias que nunca se olvidarán.

todas aquellas personas que han estado presentes durante mi formación, brindándome consejos para la vida.

Por último, a mi compañero de proyecto por aguantarme todos estos años y su apoyo durante las carreras, que me permiten entender lo importante del trabajo en equipo.

Erik Leonardo Contreras Ropero

Agradecimientos

Al profesor **Franklin Sepulveda**, por su asesoría, acompañamiento y colaboración; por su tiempo y enseñanza durante estos últimos años y por ser un excelente director.

A los profesores **Mier** y **Jairo** por los conocimientos brindados y las ayudas en los laboratorios.

A la secretaria de centro de estudios, "**Yolandita**", por disposición a ayudarnos en los devenires diarios de la carrera.

A nuestros **familiares**, por su apoyo incondicional en esos momentos difíciles, paciencia y enseñanzas.

A **compañeros y amigos**, quienes han presenciado nuestros buenos y malos momentos, que han permitido crear un lazo de amistad.

A la **Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones (E3T)** y a la **Universidad Industrial de Santander (UIS)**

¡Gracias Infinitas!

Índice general

INTRODUCCIÓN	15
1 MÉTODO	17
1.1 DETECCIÓN DEL ÁREA DE LOS LABIOS	17
1.1.1 Detección de rostros	17
1.1.2 Definición área de interés	19
1.2 PROCESAMIENTO DE COLOR	20
1.3 APLICACIÓN DEL MODELO DE CONTORNOS ACTIVOS	23
1.3.1 Snakes	23
1.3.2 Modelos de contornos activos sin bordes (Chan-Vese)	27
1.4 SEGUIMIENTO DE PUNTOS DE CONTORNO Y APERTURA DE LABIOS	29
1.4.1 Lucas-Kanade	29
1.4.2 Cálculo de la Apertura	31
2 ANÁLISIS DE RESULTADOS Y LIMITANTES	32
2.1 PRUEBAS SOBRE IMÁGENES EN FORMATO JPEG	32
2.1.1 Detección del área de interés	33
2.1.2 Aplicación de modelo de contornos activos	33
2.2 PRUEBAS EN SECUENCIAS DE IMÁGENES	34
2.2.1 Pruebas sobre vídeos de youtubers colombianos	35
2.2.2 Pruebas sobre vídeos de presentadores colombianos.	37
2.2.3 Pruebas con fuente de luz controlada	39
2.3 LIMITACIONES DEL MÉTODO PROPUESTO	41
BIBLIOGRAFÍA	46

Índice de figuras

Figura 1	Funcionamiento en cascada del algoritmo de Viola-Jones.	18
Figura 2	Aplicación del algoritmo de detección de rostros Viola-Jones . . .	19
Figura 3	Fallos en detección de labios usando Viola-Jones	19
Figura 4	Ajuste de la región de interés.	20
Figura 5	Histogramas de representación de color RGB sin procesamiento de color.	21
Figura 6	Aplicación tratamiento de componentes RGB	22
Figura 7	Variación de parámetros	23
Figura 8	Aplicación de modelos de contorno activos (<i>Snake</i>) sin procesa- miento de color	27
Figura 9	Aplicación de modelos de contorno activos (<i>Snake</i>) con procesa- miento de color	27
Figura 10	Segmentación de la imagen en zonas de intensidades constantes U_0	28
Figura 11	Resultado de la aplicación de Chan-Vese y minimización de la energía	29
Figura 12	Seguimiento flujo óptico utilizando Lucas-Kanade sobre la imagen procesada	30
Figura 13	Transferencia de puntos de seguimiento a la imagen original . . .	31
Figura 14	Imágenes de prueba	33
Figura 15	Cálculo de la apertura mediante el seguimiento por flujo óptico utilizando Lucas-Kanade	36
Figura 16	Seguimiento flujo óptico utilizando Lucas-Kanade sobre presen- tadores.	38
Figura 17	Prueba con luz controlada.	39
Figura 18	Resultados con aplicación de luz controlada.	40
Figura 19	Comparación extracción del contorno mediante <i>Snakes</i>	42
Figura 20	Fallo del Seguimiento	42

Índice de tablas

Tabla 1	Representación promedio de los tonos de labios.	22
Tabla 2	Tiempo promedio para la extracción de rostro y de labios	33
Tabla 3	Tiempo promedio aplicación de modelos de contornos activos	34
Tabla 4	Tiempos de procesamiento para vídeos de <i>Youtubers</i>	35
Tabla 5	Error del cálculo de apertura por cada frame del vídeo Youtuber01	36
Tabla 6	Error de la apertura de cada vídeo <i>Youtubers</i>	36
Tabla 7	Tiempos de procesamiento para cada vídeo de <i>Presentadores</i>	37
Tabla 8	Error de la apertura de cada vídeo <i>Presentadores</i>	39
Tabla 9	Tiempos de procesamiento para cada vídeo con luz controlada.	40
Tabla 10	Error de la apertura de cada vídeo	40

RESUMEN

Título: Estimación del grado de apertura de los labios mediante técnicas de segmentación de imágenes.¹

Autores:

Erik Leonardo Contreras Ropero²

Jhorman Andrey Landazabal Rodríguez²

Palabras Clave: Contorno de los Labios; Contornos Deformables; Python; Procesamiento Digital de Imágenes; Espacios de Color.

DESCRIPCIÓN

El interés en el estudio de la región de la boca ha crecido significativamente ante la motivación de desarrollar sistemas capaces de interpretar información a partir del movimiento de los labios, donde en diferentes trabajos se han encontrado problemas relacionados con la presencia de vello facial, tono de piel e intensidad de luz principalmente. Nuestro estudio se basa en la implementación de un programa mediante la combinación de aquellas técnicas conocidas que permiten suplir los problemas mencionados y a su vez desarrollar una base para su aplicación en tiempo real. Para su estudio se planteó en primer lugar el análisis individual de los métodos conocidos a utilizar; en segundo lugar, la implementación del método propuesto y por último, el análisis de resultados mediante el cálculo del error y la comparación en los tiempos de ejecución aplicado a diferentes grupos de vídeos bajo diferentes intensidades de luz. Como resultado de este análisis se determinó la importancia de la fuente de luz sobre los vídeos, debido a que depende tanto del tipo de luz como la dirección en como incide en el rostro, a partir de este aspecto es importante realizar un ajuste de parámetros tanto para la modificación de color en la imagen, como la posición necesaria para realizar la extracción del contorno. Por lo tanto se determina que la tasa de procesamiento es de 18 frames por segundo, siendo el objetivo para estudiar el movimiento de la boca.

¹Trabajo de Grado

²Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Franklin Alexander Sepulveda Sepulveda, PhD.

ABSTRACT

Title: Estimation of lips' degree opening by image segmentation techniques¹

Authors:

Erik Leonardo Contreras Ropero²

Jhorman Andrey Landazabal Rodríguez²

Key Words: Contour of the Lips; Deformable Contours; Python; Digital Image Processing; Color Spaces.

DESCRIPTION

Interest in the study of the mouth's region has grown significantly thanks to the motivation to develop systems capable of interpreting information from the lips' movement, where in different works we have found problems related to the presence of facial hair, skin and light intensity mainly. Our study is based on the implementation of a program using the combination of those known techniques that allow to overcome the problems mentioned and in turn to develop a base for its application in real time. For its study, the individual analysis of the known methods to be used was first proposed; secondly, the implementation of the proposed method and, finally, the analysis of results by calculating the error and the comparison in the execution times applied to different groups of videos under different light intensities. As a result of this analysis was determined the importance of the light source on the videos, because it depends both on the type of light and the direction in which it affects the face, from this aspect it is important to perform a parameter adjustment both For the modification of color in the image, as the position necessary to perform the extraction of the contour. Therefore it is determined that the processing rate is 18 frames per second, being the objective to study the movement of the mouth.

¹Bachelor Thesis

²Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Franklin Alexander Sepulveda Sepulveda, PhD.

OBJETIVO GENERAL

Desarrollar un algoritmo que permita obtener el contorno exterior de los labios mediante métodos basados en medidas de gradiente y modelos de contorno activo, aplicados estos sobre secuencias de imágenes del rostro, con el fin de realizar el seguimiento de los movimientos de los labios.

OBJETIVOS ESPECÍFICOS

- ❖ Implementar, un algoritmo a partir del método de Viola-Jones, que permita detectar y separar de manera automática el área de la boca en imágenes de formato JPEG.
- ❖ Desarrollar un algoritmo a partir de la combinación del modelo de contornos activos y las técnicas de segmentación de imágenes para la detección de bordes que permita representar la forma del contorno exterior de los labios.
- ❖ Implementar un algoritmo que realice el seguimiento de los puntos más relevantes del contorno exterior de los labios en una secuencia de imágenes de formato JPEG, partiendo de la variante propuesta por Lucas - Kanade.

INTRODUCCIÓN

La detección de labios es un problema que ha sido abordado durante las últimas décadas y buena parte de las soluciones propuestas utilizan los modelos de contornos activos, e.g. [2, 3, 7, 12, 15]. Adicionalmente se utilizan metodologías tales como *Splines*, que corresponde a una forma de representación a trozos con polinomios. Este método permite el ajuste de curvas mediante figuras geométricas, y, facilita la realización de ajustes suaves a los contornos [3].

En lo que se refiere al uso de *Splines*, en [14] se extraen los contornos de los labios a partir del uso de *B-Splines*, siendo este un caso particular dentro de los *Splines*. Los resultados que presentan, en relación a sus objetivos, son satisfactorios debido a aprovechan la diferencia de tonalidad que existe entre los labios y la piel del rostro. Sin embargo, en personas con tonalidades de labios parecidas a las del rostro el algoritmo no reconoce de igual forma el área de los labios. Más adelante los autores en [3] toman en cuenta el estudio basado en *B-Splines* mencionado anteriormente para desarrollar un algoritmo capaz de obtener el borde de los labios en videos. Este trabajo mencionado presenta un modelo que muestra ser robusto ante la presencia de barba, tono de piel, iluminación y diferencias en enfoque, siendo capaz de obtener gran parte del contorno de los labios. Pero presenta el inconveniente de que no adquiere automáticamente este contorno, y en lugar de eso debe ir ajustando para diferentes casos en el cual se aplica.

De otra parte, los autores en [12] desarrollaron un algoritmo basado en modelos de contornos activos (*Snakes*), cuyo objetivo principal es obtener un método lo suficientemente rápido para ser aplicado en dispositivos de comunicación e interfaces hombre-máquina. Reportan una mejora en la eficiencia del algoritmo, obteniendo resultados de alrededor del 98% en detección de bordes pero consumiendo menos recursos. A este nuevo método se conoce como *Jumping Snakes*. Adicionalmente dan a conocer los pro-

blemas en obtener el contorno del labio inferior y las esquinas cerca a la boca. Existen trabajos adicionales que hacen uso de *Snakes*. En [2] se obtiene el área de los labios a través del modelo de contornos activos (*Snakes*) realizando modificaciones a los diferentes parámetros que lo componen obteniendo resultados que varían en gran medida en la tasa de detección del contorno.

Los autores en [15] usan formas geométricas y análisis del gradiente para realizar el seguimiento a los labios. La metodología se compone de dos partes, la primera se hace uso de *Jumping Snakes* para encontrar los puntos clave que se adaptan a las deformaciones de la boca, lo cual crea un modelo para cuando la boca se encuentre cerrada y otro para cuando se encuentre abierta. La elección del modelo es elegido automáticamente por un criterio geométrico. La segunda parte es el *Seguimiento*, este proceso utiliza el método de Lucas-Kanade [8] y mediante los puntos iniciales obtenidos por *Jumping Snakes* se realiza el respectivo seguimiento. La combinación de ambos métodos maximiza el flujo de gradiente, lo que permite optimizar la estimación de los parámetros.

Por último, se presenta un trabajo complementario expuesto en [13], en donde se implementan tres algoritmos para la detección de labios. El primero consiste en el uso de los espacios de color para la detección de áreas, el segundo en el uso de *B-splines* y modelos deformables para la detección de contornos y el tercero es una combinación de los dos anteriores. Con esto realizan una comparación en la cual afirman que sus resultados mejoraron considerablemente debido a que aumentó la tasa de detección de los labios. Sin embargo, no realizaron pruebas a imágenes con diferentes características para determinar su robustez.

El presente trabajo busca desarrollar un algoritmo que permita la extracción del contorno de los labios y el cálculo de su grado de apertura. Para ello se plantea la conjunción de técnicas de procesamiento de imágenes como procesamiento de color mediante el análisis de intensidades RGB, contornos activos y flujo óptico con el fin de suplir las limitaciones en los trabajos presentados anteriormente, tales como presencia de bello facial, tono de piel e intensidad y tipo de luz.

Capítulo 1

MÉTODO

Este capítulo aborda el método desarrollado en el presente trabajo de grado de pregrado, el cual toma como base los trabajos reportados en, e.g. [3, 15, 16]. Se plantean cuatro secciones: la detección del área de los labios usando el algoritmo de Viola-Jones; el procesamiento de color que busca segmentar el área de interés en dos regiones (labios y piel); la extracción del contorno mediante la aplicación del modelo de contornos activos y por último, el seguimiento de puntos claves haciendo uso del flujo óptico propuesto por Lucas-Kanade [8].

1.1. DETECCIÓN DEL ÁREA DE LOS LABIOS

La correcta selección del área de trabajo es pieza fundamental en la extracción del contorno exterior de los labios. Con ello se busca disminuir información no útil, tales como el fondo y formas pertenecientes a otras regiones, ya que podrían interferir en el comportamiento del gradiente de la imagen, afectando los resultados de procedimientos posteriores como el procesamiento de color y la aplicación del modelo de contornos activos. En el presente trabajo se plantea usar el algoritmo de detección de objetos propuesto por los autores (Paul Viola & Michael Jones) en [18].

1.1.1 Detección de rostros El algoritmo de Viola-Jones de detección de rostros [18] aprovecha las características similares que comparten todos los rostros, como por ejemplo que la región de la nariz es más brillante que la de los ojos. La extracción de características la realiza utilizando la operación que los mismos autores denominan *integral image*. A modo de función analizante utilizan funciones rectangulares que corresponden a regiones rectangulares adyacentes que procuran ser similares a deter-

minadas regiones importantes del rostro [17]; e.g. cejas-pómulos, ojos. Estas funciones analizantes son similares a las bases de representación de la familia Haar [5]. Los autores reportan que el hecho de trabajar en el dominio de la características en lugar del dominio de los pixeles permite detectar objetos de manera rápida.

Este algoritmo presenta una arquitectura en cascada en donde cada etapa consta de un clasificador que usa desde características generales simples, como que la región del ojo que es más oscura que las mejillas, hasta combinaciones complejas como el limite del rostro con el fondo como se observa en la figura 1. El uso de estas características son planteadas por Papageorgiou en [11], dado que genera sub-ventanas multi-escala del mismo tamaño y forma, los cuales son adyacentes tanto horizontal como verticalmente, esto con el fin de obtener valores que permitan diferenciar si la región donde se aplica es una componente del rostro o no, logrando mejorar la velocidad de detección y disminuir la tasa de falsos positivos.

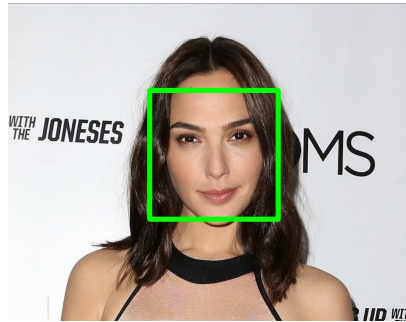
Figura 1: Funcionamiento en cascada del algoritmo de Viola-Jones.



Fuente: <http://www.makemetrics.com/research/viola-jones/>

Los resultados obtenidos de la aplicación del algoritmo Viola-Jones de detección haciendo uso de la librería *OpenCV* en *Python* se pueden observar en la Figura 2.

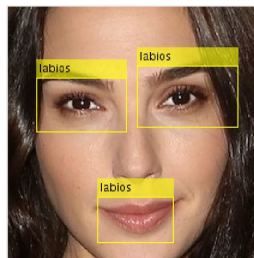
Figura 2: Aplicación del algoritmo de detección de rostros Viola-Jones



1.1.2 Definición área de interés Una vez realizado el proceso de determinación del rostro, lo que sigue de determinar el área de interés (la región de los labios y sus alrededores). Sin embargo, en este caso ya no se usa el algoritmo de Viola-Jones debido al aumento de falsos positivos (ver Figura 3a) causados por las similitudes de forma entre los ojos y labios desde el punto de vista de las funciones analizantes rectangulares. Aun si separase solo la mitad inferior, el algoritmo de Viola-Jones tiene problemas para detectar los labios en los casos en los que el grado de apertura de estos es alto; además, las regiones detectadas cortan puntos relevantes de definición del contorno como las comisuras de los labios, ver Figura (3b).

Figura 3: Fallos en detección de labios usando Viola-Jones

(a) Similitud entre labios y boca.



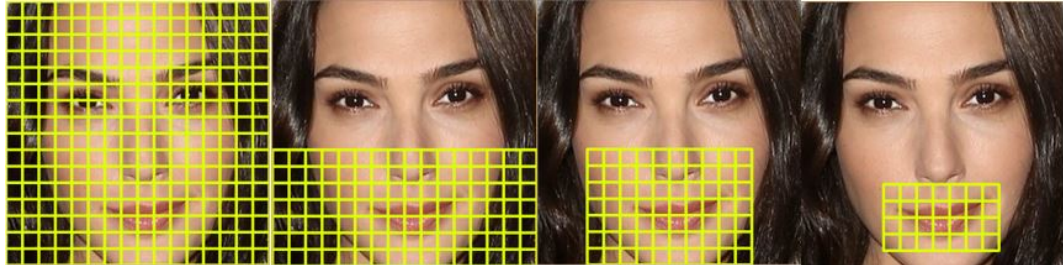
(b) Perdida de puntos de interés



En el presente trabajo, para definir el área de interés se aprovecha la característica multi escala del algoritmo de detección de rostros pues provee un tamaño constante del área de interés sin importar que el tamaño del rostro varíe en relación a la imagen completa. Por tanto, el área de interés se obtiene en dos pasos: primero, seccionamiento del rostro en regiones cuadradas iguales; y segundo, se limita la región a solo aquella

zona donde los labios puedan moverse (Figura 4). Esta zona se obtiene de manera experimental.

Figura 4: Ajuste de la región de interés.

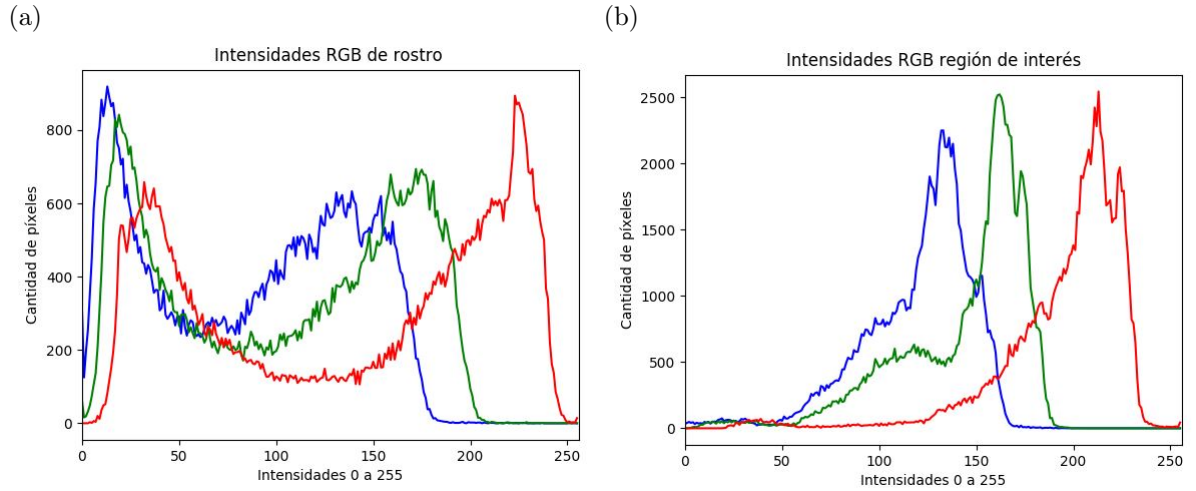


1.2. PROCESAMIENTO DE COLOR

El proceso de segmentación de la región de interés se realiza haciendo uso del espacio de color RGB y la representación en escala de grises. La representación en RGB provee no solo información del color sino también del brillo simulando la capacidad del sistema de visión humana de mantener la percepción del color con diferentes fuentes iluminación, con lo cual es posible plantear una segmentación entre tonos de labios y piel partiendo de la información de intensidades que estos proveen aún variando las condiciones de luz a la que se someta. El objetivo del procesamiento de color es obtener la forma de los labios para facilitar y mejorar los resultados de la aplicación de los modelos de contornos activos.

En los histogramas de las figuras (5a) y (5b) se aprecia que tanto para la interpretación del rostro como para los labios prevalece el rojo, además se tiene una mayor intensidad del componente verde que azul, de acuerdo a lo propuesto por los autores en [4] la piel se representa por la mezcla de un mayor cantidad de cromas verdes que azules (obviando el rojo), mientras que para las tonalidades de labios dichas componentes son semejantes. Por ende, para realizar una depuración de colores que permita identificar adecuadamente los labios se tiene en cuenta la mezcla de intensidades de componentes verde y azul.

Figura 5: Histogramas de representación de color RGB sin procesamiento de color.



De los análisis de histogramas y la correlación encontrada entre la concentración de píxeles de las intensidades verde, azul y rojo resultantes de imágenes con distintas fuentes de luz (Fluorescente y Amarilla) se determina que la componente azul es quien más varía aportando mayor información a la interpretación de la iluminación [19], se propone una configuración de color dada por

$$NewColor(x) = \alpha R(x) - \beta G(x) + \gamma B(x) \quad (1.1)$$

donde inicialmente se busca eliminar los aportes de la cromas verde a la interpretación de la tonalidad de la piel.

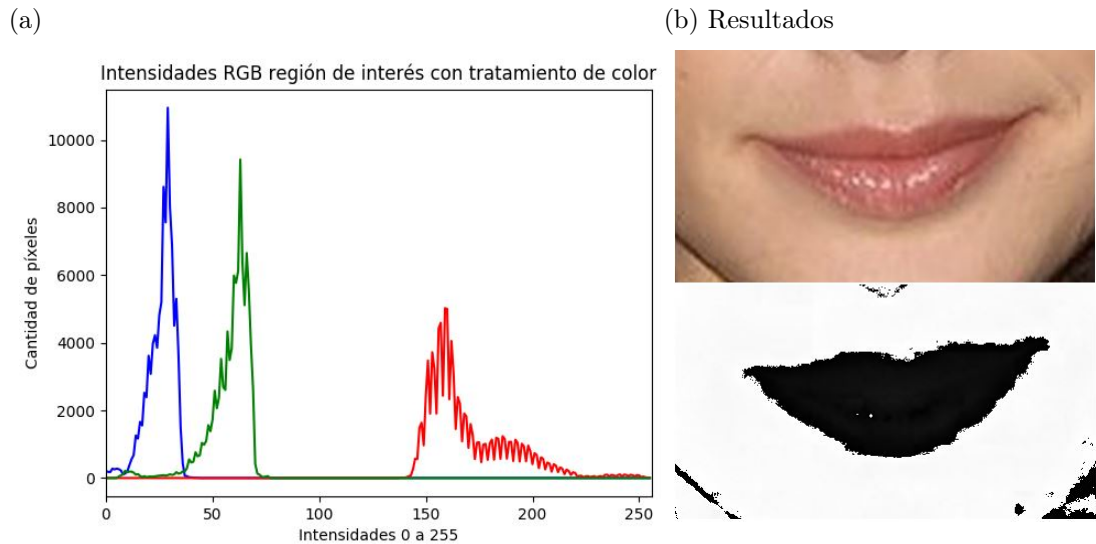
Los valores de las variables escalares α , β y γ buscan que la distribución de intensidades acentúen las cromas rojas que pertenece a la tonalidad de labios donde las componentes verde y azul son similares, del análisis de múltiples imágenes de la región de interés con variaciones en calidad, resolución, tonalidades de piel y labios se determina que las intensidades de rojo que definen los labios varían desde 140 a 200. La tabla 1 muestra la configuración RGB que representa en promedio el color de los labios para los sujetos puestos a prueba.

Tabla 1: Representación promedio de los tonos de labios.

Croma	Intensidad Promedio	Varianza
R	159,4	16,04
G	59,4	13,24
B	25,9	15,29

Para obtener los valores de los parámetros se hace uso de la relación entre las intensidades de cada croma donde se presenta la mayor cantidad de píxeles en la representación del área de interés con el promedio de la representación RGB del tono de los labios, con esto se plantea que los píxeles de la croma azul deben concentrarse sobre una intensidad menor a 50; Además la diferencia entre la intensidad de la croma verde y la roja donde para cada una se acumule la mayor cantidad de píxeles debe ser de 80 puntos o más. A partir de esto se obtiene que $\alpha = 0,22$, $\beta = 0,6$ y $\gamma = 0,3$ modifica la distribución de intensidades ubicando la mayor cantidad de píxeles de cada croma cerca a la configuración de tono de labios planteada, esto se aprecia en la figura 6a. Como resultado de la nueva configuración de color la región de interés se segmenta en dos como se muestra en la figura 6b.

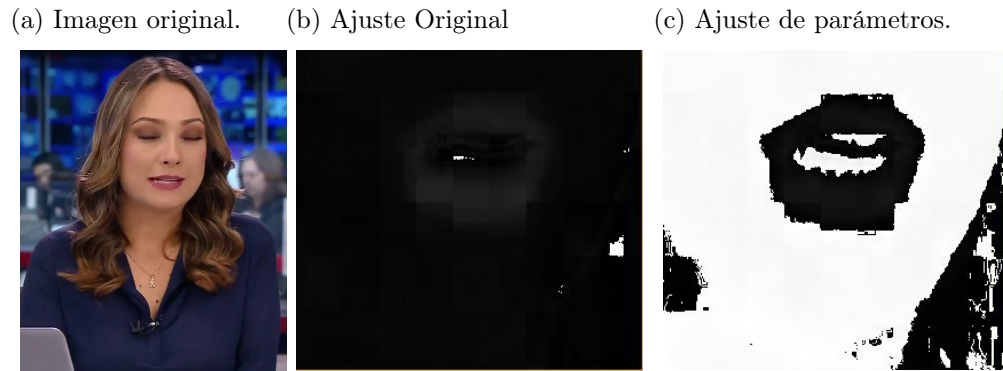
Figura 6: Aplicación tratamiento de componentes RGB



El algoritmo se diseñó con α , β y γ con las condiciones señaladas, pero debido a factores que pueden afectar en la interpretación de color como el tipo luz (Fluorescente o

Amarilla) y el tono de piel, se define un rango donde los parámetros se pueda ajustar, en donde β varia para asimilar una mayor cantidad de tonalidades de piel y γ la variación de luz, como se observa en la figura 7 donde los valores α y β se mantienen fijos, pero γ disminuye a 0.22 pues la luz normalmente usada dentro de un set de noticias es una combinación de fluorescente y amarilla.

Figura 7: Variación de parámetros



La nueva relación se normaliza haciendo uso del valor máximo y mínimo de las intensidades calculadas [4].

$$NewColorLips(x) = \frac{(NewColor - \min(NewColor))}{\max(NewColor) - \min(NewColor)} \quad (1.2)$$

Por último para que esta interpretación de color pueda aplicarse con los demás algoritmos se convierte a formato *uint8* limitando los valores de 0 a 255 como la representación en escala de grises.

1.3. APLICACIÓN DEL MODELO DE CONTORNOS ACTIVOS

1.3.1 Snakes El modelo de contornos activos usa energía de la imagen con la finalidad de que una curva deformable (*Spline*) sea guiada hacia los contornos de un objeto a causa de las fuerza del gradiente de la imagen [6]. La curva deformable utiliza una elipse definida por puntos iniciales, la cual permite definir mejor el área de interés.

El proceso de encontrar el contorno activo que mejor se adecua a la forma del objeto de interés dentro de la imagen se hace mediante el uso de criterios de energía de la forma [6],

$$E_{snake}^* = \int_0^1 E_{snake}(V(s))ds \quad (1.3)$$

$$E_{snake}^* = \int_0^1 (E_{int}(V(s)) + E_{image}(V(s)) + E_{con}(V(s)))ds \quad (1.4)$$

Donde $V(s) = (x(s), y(s))$ corresponde a la representación paramétrica de la posición del *snake*; E_{int} representa la energía interna de la curva deformable, que impone una restricción de suavidad por sección; E_{image} da lugar a las fuerzas de la imagen, las cuales guían al *Snake* a características salientes como líneas, bordes y contornos; E_{con} es la fuerza externa que restringe el movimiento del *Snake* minimizando su energía.

La energía del *Spline* se representa matemáticamente por un término de primer orden $\alpha(s)$ y un término de segundo orden $\beta(s)$ [6],

$$E_{int} = (\alpha(s) | V_s(s) |^2 + \beta(s) | V_{ss}(s) |^2)/2 \quad (1.5)$$

El término de primer orden hace que los *Snakes* actúen como una membrana y el término de segundo orden hace que este actúe como una placa delgada. El ajuste continuo de estos términos permite la adecuación de los *Snakes* al área deseada.

La energía total de la imagen puede ser expresada como la combinación ponderada de tres funciones de energía

$$E_{image} = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \quad (1.6)$$

Donde $E_{line} = I(x, y)$ representa la intensidad de la imagen en si misma ponderada por un valor w_{line} , el *Snake* será atraído tanto por las líneas claras como las oscuras. Por lo tanto el *Snake* se irá alineando el mismo con los contornos claros y oscuros.

Encontrar los bordes en una imagen utiliza una función relacionada a la energía la cual se representa mediante los gradientes de la imagen $E_{edge} = - | \nabla I(x, y) |^2$.

La teoría de detección de bordes de Marr-Hildreth en [9] da a conocer la relación *Space-Scale*, permitiendo experimentar con diferentes bordes funcionales, donde una opción de representación es la siguiente,

$$E_{image} = -(G_{\sigma} * \nabla^2 I)^2. \quad (1.7)$$

G_{σ} es desviación estándar de Gauss.

Con el fin de encontrar términos de segmentos de líneas y esquinas, se utiliza la curvatura de nivel de las líneas en una imagen ligeramente suavizada. Si $C(x,y)$ representa esta imagen suavizada entonces la curvatura puede ser escrita como

$$E_{term} = \frac{\partial \theta}{\partial n_{\perp}} \quad (1.8)$$

$$E_{term} = \frac{\partial^2 C / \partial n_{\perp}^2}{\partial C / \partial n} \quad (1.9)$$

$$E_{term} = \frac{C_{yy}C_x^2 - 2C_{xy}C_xC_y + C_{xx}C_y^2}{(C_x^2 + C_y^2)^{3/2}} \quad (1.10)$$

Donde $\theta = \tan^{-1}(C_y/C_x)$ es el grado de inclinación, $n = (\cos\theta, \sin\theta)$ es un vector unitario y $n_{\perp} = (-\sin\theta, \cos\theta)$ es un vector unitario perpendicular, ambos en la dirección del gradiente. Los términos resultantes C_x , C_y , C_{xy} , C_{xx} y C_{yy} combinados con el término de E_{edge} crean un *Snake* el cual es atraído a bordes y terminaciones.

Morse en [10], plantea la aproximación discreta de E_{snake} haciendo uso de un número discreto de puntos en el *snake* de la forma $\bar{v} = (x_i, y_i)$.

$$E^*_{snake} \approx \sum_1^n E_{snake}(\bar{v}_i) \quad (1.11)$$

Con el fin de minimizar la energía del *Snake*, se aprovecha el comportamiento discreto, donde ∇E_{snake} se define a través de la suma de las derivadas en los puntos de ajuste $\bar{v} = (x_i, y_i)$, esto permite realizar la minimización a partir de iteraciones de \bar{v} .

$$\nabla E_{snake}(\bar{v}_i) = W_{int} \nabla E_{int}(\bar{v}_i) + W_{img} \nabla E_{img}(\bar{v}_i) + W_{con} \nabla E_{con}(\bar{v}_i) \quad (1.12)$$

El autor en [10], afirma que los términos $W_{img} \nabla E_{img}(\bar{v}_i) + W_{con} \nabla E_{con}(\bar{v}_i)$ dependen

solamente de la imagen, por lo que simplemente se calcula su derivada en cada punto, esta función se denomina ∇E_{ext} :

$$\nabla E_{ext} = W_{img} \nabla E_{img}(\bar{v}_i) + W_{con} \nabla E_{con}(\bar{v}_i) \quad (1.13)$$

Mediante la minimización de la energía se obtiene la siguiente ecuación:

$$\bar{v}_i \leftarrow \bar{v}_i - \nabla E_{snake}(\bar{v}_i) \quad (1.14)$$

$$\bar{v}_i \leftarrow \bar{v}_i - \gamma [W_{int} \nabla E_{int}(\bar{v}_i) + \nabla E_{ext}(\bar{v}_i)] \quad (1.15)$$

El cálculo del gradiente de la energía interna se simplifica al hacer los términos $\alpha(s)$ y $\beta(s)$ constantes, como se propone en [6], obteniendo la siguiente ecuación:

$$\nabla E_{int} = \alpha \frac{\partial^2 \bar{v}}{\partial s^2} + \beta \frac{\partial^4 \bar{v}}{\partial s^4} \quad (1.16)$$

Además utilizando diferencias finitas realizan una aproximación, logrando así la siguiente expresión:

$$\bar{v}_i \leftarrow \bar{v}_i - \gamma \{W_{int} [\alpha \frac{\partial^2 \bar{v}}{\partial s^2}(\bar{v}_i) + \beta \frac{\partial^4 \bar{v}}{\partial s^4}(\bar{v}_i)] + \nabla E_{ext}(\bar{v}_i)\} \quad (1.17)$$

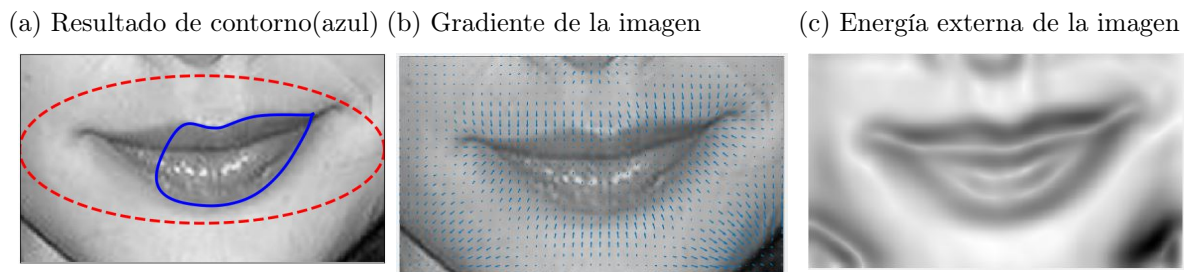
Esta expresión separa las componentes x y y que permitirán realizar el cálculo de las derivadas de la energía interna en cada iteración. Para la energía externa, simplemente es pre-calcular las derivadas de las imágenes usando los nuevos puntos de \bar{v}_i .

La aplicación del modelo de contornos activos se realiza haciendo uso de la librería *Skimage* de *Python*, se basa en [6] y funciona sobre imágenes en escala de grises. Los parámetros de configuración están determinados por α , β y γ de la minimización de energía interna aproximada de la ecuación 1.17. Debido a la falta de definición del contorno de los labios en la imagen, se establece $\alpha = 0,015$ para que las variaciones más pequeñas del gradiente mejoren el comportamiento de membrana de los *Snakes*; Además se fija $\beta = 5$ para segmentar una área grande y por último $\gamma = 0,001$ para definir el paso de las iteraciones durante el ajuste.

La figura 8a se puede distinguir que los puntos iniciales y el resultado final de la aplicación de *Snakes* no se adaptan completamente al contorno de los labios. Aunque

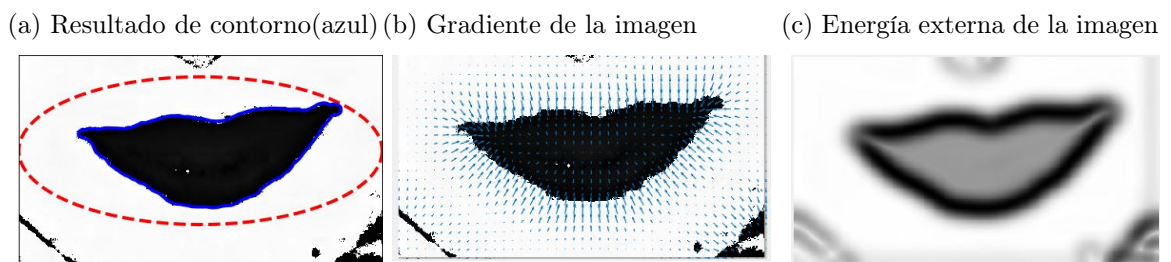
el gradiente de la imagen presentado en la figura 8b muestra el flujo hacia los bordes de los labios, la energía externa de la imagen (Figura 8c) muestra una sección oscura a la derecha inferior donde la fuerza del gradiente es mayor afectando el ajuste del *Spline*.

Figura 8: Aplicación de modelos de contorno activos (*Snake*) sin procesamiento de color



En este punto se puede notar la importancia del procesamiento de color propuesto, cuyo objetivo es aumentar la intensidad del gradiente hacia el contorno de los labios, mejorando el ajuste proporcionado por el modelo de contornos activos. La figura 9a presenta el resultado del ajuste proporcionado por el algoritmo de *Snakes* con la aplicación del procesamiento de color. El flujo de gradiente no presenta variaciones notables (Figura 9b) pero la energía externa de la imagen (Figura 9c) sí presenta cambios significativos y es este el motivo de la mejora presentada en el ajuste al contorno.

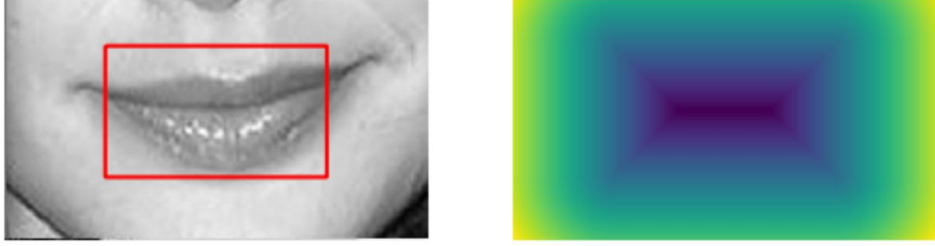
Figura 9: Aplicación de modelos de contorno activos (*Snake*) con procesamiento de color



1.3.2 Modelos de contornos activos sin bordes (Chan-Vese) El método de Chan-Vese en [1], se basa en la minimización de energía partiendo de segmentación. Se asume que se tiene una imagen u_0 formada por dos regiones, u_o^i y u_o^o , cada una con un factor distinto y aproximadamente valores constantes. u_o^i representa

el interior del objeto y u_0^o representa el exterior del objeto, la Figura 10 en rojo presenta la inicialización de u_0 sobre la imagen en escala de grises y la primera distribución energética generada, con esto se define la ecuación 1.18.

Figura 10: Segmentación de la imagen en zonas de intensidades constantes U_0



$$F_1(C) + F_2(C) = \int_{inside(C)} |u_0(x, y) - c_1|^2 dx dy + \int_{outside(C)} |u_0(x, y) - c_2|^2 dx dy \quad (1.18)$$

donde C es cualquier otra curva variable, y las constantes c_1, c_2 dependiendo de C , son los promedios de u_0 dentro de C y respectivamente fuera de C .

$$inf_C \{F_1(C) + F_2(C)\} \approx 0 \approx F_1(C_0) + F_2(C_0) \quad (1.19)$$

Donde C_0 representa el limite del objeto. Si la curva C esta por fuera del objeto, entonces $F_1(C) > 0$ y $F_2(C) \approx 0$. Si la curva C esta dentro del objeto, entonces $F_2(C) > 0$ y $F_1(C) \approx 0$. Si la curva C esta dentro y fuera del objeto, entonces $F_1(C) > 0$ y $F_2(C) > 0$. Finalmente la energía es minimizada si $C = C_0$, es decir, si la curva C esta en el limite del objeto, este planteamiento esta definido por Chan-Vese en [1] y también la energía funcional como se plantea en la siguiente ecuación.

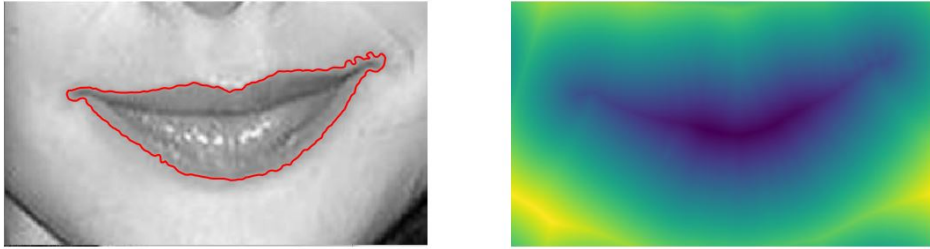
$$F(c_1, c_2, C) = \mu \cdot Length(C) + \nu \cdot Area(inside(C)) \\ + \lambda_1 \int_{inside(C)} |u_0(x, y) - c_1|^2 dx dy + \lambda_2 \int_{outside(C)} |u_0(x, y) - c_2|^2 dx dy$$

donde $\mu \geq 0, \nu \geq 0, \lambda_1, \lambda_2 > 0$ son parámetros fijos. Estos valores son fijados en la mayoría de casos como $\lambda_1 = \lambda_2 = 1$ y $\nu = 0$.

La Figura 11 presenta el contorno detectado por la aplicación de este modelo de

detección y el resultado final de la distribución energética generada por u_0 , a diferencia de la detección realizada por *Snake* el algoritmo propuesto por Chan-Vese [1] puede distinguir el contorno de los labios sin realizar el procesamiento de color y aunque el ajuste realizado no es perfecto la información obtenida es útil para el seguimiento.

Figura 11: Resultado de la aplicación de Chan-Vese y minimización de la energía



1.4. SEGUIMIENTO DE PUNTOS DE CONTORNO Y APERTURA DE LABIOS

1.4.1 Lucas-Kanade La técnica desarrollada por Lucas-Kanade en [8], ha permitido resolver problemas como es el seguimiento de patrones y análisis de movimiento, utilizando la información de los gradientes de intensidad espacial para encontrar las posición que mejor se acopla.

La relación planteada por Lucas-Kanade en el algoritmo se expresa como

$$G(x) = F(xA + h) \quad (1.20)$$

donde A es un matriz expresando una transformación espacial lineal entre $F(x)$ y $G(x)$. Para esto se debe ajustar los valores de A y h , mediante el calculo del error el cual es expresado como

$$E = \sum_x [F(xA + h) - G(x)]^2 \quad (1.21)$$

Para determinar la cantidad de ΔA para ajustar A y cantidad de Δh para justar h , se utiliza la aproximación lineal

$$F(x(A + \Delta A) + (h + \Delta h)) \approx F(xA + h) + (x\Delta A + \Delta h) \frac{\partial}{\partial x} F(x) \quad (1.22)$$

La diferenciación con respecto a estas magnitudes y el establecimiento de los resultados igual a cero, da lugar a un conjunto de ecuaciones lineales que se resuelven simultáneamente.

Esta generalización planteada por los autores es útil en aplicaciones donde dos diferentes vistas del objeto serán vistas diferente, debido a los puntos de vista de la cámara o las diferencias en el procesamiento de las dos imágenes. Con lo cual se modela esta diferencia como una transformación lineal

$$F(x) = \alpha G(x) + \beta \quad (1.23)$$

donde α es el ajuste del contraste y β es el ajuste del brillo.

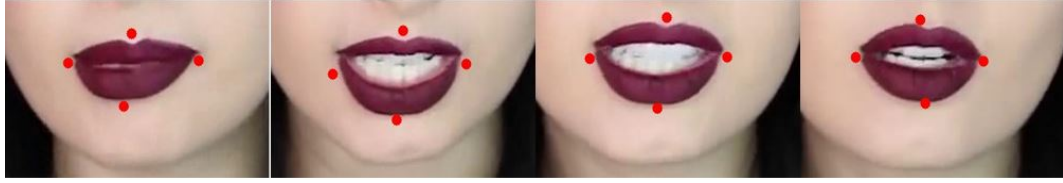
Este algoritmo es conocido por permitir rastrear objetos sin la necesidad de establecer condiciones demasiado específicas de su movimiento, debido a que buscan una descripción general que soporte el seguimiento de regiones presentes en las secuencias de cuadros de vídeo. Aunque la aplicación tanto de *Snakes* como el método de Chan-Vese no son lo suficientemente optimas en tiempo real, se opta por una solución que ha sido aplicada en diferentes trabajos como por ejemplo Silvana Vallejo [16], donde aplica la detección de contornos automática al primer frame del respectivo vídeo y aplica un seguimiento por flujo óptico a los demás frames.

El método de Lucas-Kanade [8], se enfoca en el establecimiento de diferentes ventanas dentro de una misma imagen en escala de grises para realizar el determinado seguimiento de los diferentes puntos. A partir del contorno generado mediante el modelo de contornos activos, se obtienen puntos automáticamente en posiciones deseadas que permiten realizar el respectivo seguimiento a la imagen, estos puntos se pueden apreciar en la figura 12 y 13 sobre una secuencia de imágenes.

Figura 12: Seguimiento flujo óptico utilizando Lucas-Kanade sobre la imagen procesada



Figura 13: Transferencia de puntos de seguimiento a la imagen original



1.4.2 Cálculo de la Apertura Para determinar el grado de apertura de los labios, se utiliza las coordenadas de los puntos posicionados en la parte del labio superior e inferior y mediante un cálculo de distancia entre ellos se obtiene el valor correspondiente, el cálculo se realiza mediante la siguiente ecuación.

$$Apertura = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1.24)$$

Donde *Apertura* se define como la distancia entre dos puntos obtenida a través del teorema de Pitágoras.

Capítulo 2

ANÁLISIS DE RESULTADOS Y LIMITANTES

Se diseñaron pruebas haciendo uso de la herramienta *Pytest* con el fin de comprobar el tiempo exacto de ejecución y el éxito del algoritmo; además se fraccionó el algoritmo en tres etapas para su estudio, centradas en la detección del área de interés, la aplicación de los modelos de contornos activos y seguimiento en secuencia de imágenes. Las pruebas se llevaron a cabo en un computador portátil con procesador Intel Core I5-4210u 1.7Ghz con turbo boost a 2.4Ghz, 6Gb de RAM y sistema operativo Windows 8.1 pro de 64 bits.

2.1. PRUEBAS SOBRE IMÁGENES EN FORMATO JPEG

Los algoritmos de detección del área de interés y modelos de contornos activos, se estudiaron sobre un solo frame, con el objetivo de determinar el tiempo de ejecución de cada sección. Las pruebas se realizaron sobre un total de 20 imágenes en formato *jpeg* con distintas condiciones de luz, resolución de imagen, ambiente y tonalidad de labios.

Se registran los resultados obtenidos sobre 4 de los sujetos puestos a prueba, donde se aprecia baja iluminación y tono de labios con representación cercana al tono de piel (Figura 14a); iluminación frontal (Figura 14b); alto contraste de labios respecto al tono de piel (Figura 14c) y variaciones en el fondo de la imagen (Figura 14d).

Figura 14: Imágenes de prueba



(a): <http://wp.widewallpapers.net/2k/celebs/h/hilary-duff/1280/Hilary-Duff-1280x800-001.jpg>
 (b): <http://tusimagenesde.com/imagenes-de-rostros/>
 (c): <http://eldiariodelanena.com/7-formas-lograr-las-impactantes-cejas-cara-delevingne/>
 (d): <http://www.hollywood.com/general/serious-gal-gadot-turned-down-silly-bond-girl-role-60684265/>

2.1.1 Detección del área de interés La tabla 2 muestra los resultados obtenidos para la extracción automática del rostro a partir del algoritmo de Viola-Jones y la obtención del área de interés mediante el recorte de la imagen. Los tiempos obtenidos en la detección de rostro no varían significativamente, a pesar de los cambios de resolución de los sujetos, a diferencia del tiempo de carga de la imagen, esto debido al funcionamiento en cascada que implementa el algoritmo de Viola-Jones. Por otro lado, el tiempo de obtención del área de interés es constante, pues se aplica sobre el área del rostro detectada, que se estandariza a un tamaño de 640x640 pixeles.

Tabla 2: Tiempo promedio para la extracción de rostro y de labios

	Carga [s]	D. Rostro [s]	A. Interés [s]	Total [s]	Resolución
Sujeto 1	0,14	0,32	0,01	0,47	1024x768
Sujeto 2	0,12	0,32	0,01	0,45	640x634
Sujeto 3	0,18	0,31	0,01	0,5	800x1119
Sujeto 4	0,13	0,32	0,01	0,46	882x688

2.1.2 Aplicación de modelo de contornos activos Con la región de interés definida y normalizada a un tamaño de 480x320 pixeles, se aplica la extracción de contornos mediante el modelo de contornos activos (*Snake*) y el modelo de contornos activos sin bordes propuesto por Chan-Vese [1], los tiempos de ejecución se muestran en la tabla 3. De la comparación de tiempos de ejecución, se evidencia la conveniencia del uso de *Snake* sobre el método de Chan-Vese. Dado que el algoritmo de Chan-Vese

excede por alrededor de 1 a 3 segundos el tiempo de ejecución de *Snakes*.

Tabla 3: Tiempo promedio aplicación de modelos de contornos activos

	Snake [s]	Chan-Vese [s]
Sujeto 1	2,238	5,514
Sujeto 2	3,502	5,272
Sujeto 3	2,252	5,928
Sujeto 4	3,166	5,28

2.2. PRUEBAS EN SECUENCIAS DE IMÁGENES

Adicional a imágenes disponibles en internet, se recopilaron videos de presentadores de noticieros en Colombia y Youtubers colombianos. El motivo de esta selección es que nos interesa evaluar el desempeño del algoritmo desarrollado en personas y recursos multimedia comúnmente encontrados en el entorno colombiano. En este tipo de recursos la persona tiende a estar mirando fijamente una cámara cuando habla, pero entre los dos tipos de video existe dos diferencias que nos permite abarcar más área de análisis: En el caso de los presentadores de televisión al estar dentro de un set, presenta una iluminación que tiende a ser constante, por esta razón el análisis depende del tipo de piel de la persona; por otro lado, los youtubers tienden a utilizar diferentes intensidades de luz y en diferentes direcciones, ya sea por luz natural o artificial.

El algoritmo de *Snakes* se aplica a partir una elipse alrededor del área de interés compuesta por 400 puntos, dichos puntos se desplazan en conjunto lo que les permite adaptarse al contorno, a partir de esto, sí se emplea una menor cantidad de puntos a los propuestos, el algoritmo no se adapta adecuadamente al contorno aproximado de los labios. Las pruebas en seguimiento abarcan 4 de lo 400 puntos que definen el contorno obtenido de la aplicación de *Snakes*. Los cuatro puntos seleccionados están posicionados en las comisuras, parte inferior y arco de cupido de los labios, los cuales se usan para realizar cálculos del grado de apertura y verificación del seguimiento en la secuencia de imágenes.

2.2.1 Pruebas sobre vídeos de youtubers colombianos La tabla 4 presenta los resultados a diferentes capturas de secciones de vídeo de distintos *youtubers*, como se observa, el tiempo que tarda en procesar el vídeo completo tiende a 1,5 veces la duración del vídeo. Por consiguiente, el algoritmo tiene una capacidad de procesamiento de alrededor de 19 frames por segundo.

Por otro lado, el tiempo de extracción obtenido de las imágenes en la tabla 2 con respecto al procesamiento frame a frame de un vídeo en la tabla 4, difieren debido al comando utilizado en *Python*, en el caso de los vídeos se carga el respectivo archivo de vídeo que mediante un comando de lectura es capaz de procesar cada frame en un tiempo significativamente menor.

Tabla 4: Tiempos de procesamiento para vídeos de *Youtubers*.

Youtuber	Tiempo Procesamiento [s]	Duración vídeo [s]	Número de Frames	frames/segundo
01	17,132	11	330	19,26
02	16,016	10	300	18,7
03	7,748	5	150	19,35

La prueba realizada para el análisis de la apertura está compuesta por los 10 primeros frames del vídeo. En la figura 15 se aprecian cuatro de los diez frames propuestos. De otra parte, la tabla 5 muestra el valor promedio de la apertura registrada mediante el algoritmo y la apertura obtenida de manera manual del vídeo del Youtuber01 (Figura 15a), con lo cual se calcula el error de la apertura por frame, para esto se realizó la medida 10 veces por cada frame, y se determina que los resultados obtenidos no difieren pues se realizan siempre bajo los mismos parámetros. Esta prueba se basa en la aplicada por Silvana Vallejo en [16]. Durante estas pruebas se encontró que la mayoría de vídeos de youtubers presentan una iluminación blanca constante que disminuye el ajuste requerido de color, por ende la pérdida de información para realizar el seguimiento es menor.

Tabla 5: Error del cálculo de apertura por cada frame del vídeo Youtuber01

Apertura Teórica	Apertura Experimental	Error [%]
87	83	4,598
89	90	1,124
93	97	4,301
102	101	0,980
105	108	2,857
112	111	0,893
116	106	8,621
107	101	5,607
102	104	1,961
104	107	2,885

La tabla 6 representa el error porcentual del grado de apertura por cada vídeo al que se le realizó la prueba.

Tabla 6: Error de la apertura de cada vídeo *Youtubers*

Youtuber	Error Apertura [%]
01	3,383
02	3,341
03	5,631
04	2,329
05	4,019

Figura 15: Cálculo de la apertura mediante el seguimiento por flujo óptico utilizando Lucas-Kanade

(a) Youtuber01 - Laura Sánchez.





(b) Youtuber02 - Maria Alejandra Ayala.



(c) Youtuber03 - Nicolas Arrieta.

2.2.2 Pruebas sobre vídeos de presentadores colombianos. Las pruebas realizadas sobre los vídeos de presentadores difieren de la de los youtubers. En vista a que cada sujeto perteneciente al grupo de los presentadores requiere de un ajuste en los parámetros del procesamiento de color diferente, para evitar que esto afecte los cálculos de tiempo de ejecución se efectúa un ajuste previo en los parámetros con el fin de que la medida se tome bajo las mismas condiciones para cada presentador. Los resultados de las pruebas en la tabla 7 evidencian una disminución de procesamiento del algoritmo a 18 frames por segundo.

Tabla 7: Tiempos de procesamiento para cada vídeo de *Presentadores*.

Presentador	Tiempo Promedio [s]	Duración vídeo [s]	Número de Frames	Tasa Frames/segundo
01	15,872	9	270	17,011
02	8,4	5	150	17,85
03	11,604	7	210	18,09
04	16,352	10	300	18,34

Las pruebas se realizaron sobre los presentadores, en los cuales se percibe dos factores que afectan el resultado de la aplicación del algoritmo. El primer factor tiene relación con el tamaño del rostro con respecto a la resolución de la imagen. Esto influye pues el área de interés contiene menor información, luego al aplicar el procesamiento

de color se presenta una mayor pérdida del contorno de los labios. En segundo lugar se tiene la incidencia de la luz como se mostró en la figura 7 donde es necesario realizar un ajuste en los parámetros de procesamiento de color, aunque este arreglo es efectivo existe la posibilidad de perder información del contorno de los labios lo cual influye en el resultado del grado de apertura. Como se muestra en la figura 16, aún con estos problemas mencionados el algoritmo permite efectuar el seguimiento al vídeo.

Figura 16: Seguimiento flujo óptico utilizando Lucas-Kanade sobre presentadores.

(a) Presentador01 - Mónica Jaramillo.



(b) Presentador02 -César Chaparro.



(c) Presentador03 - Vaness de la Torra.



Se realiza el mismo cálculo a cada uno de los vídeos de presentadores obteniendo el error porcentual en la tabla 8

Tabla 8: Error de la apertura de cada vídeo *Presentadores*

Presentador	Error Apertura [%]
01	4,763
02	6,293
03	5,682
04	5,639
05	6,88
06	4,648
07	6,356

2.2.3 Pruebas con fuente de luz controlada En las pruebas realizadas a vídeos de *youtubers* y *presentadores* se observaron errores en el cálculo de la apertura debido a la influencia de la luz, pues en algunos casos la posición de la fuente de luz incide en la proyección de sombras, como la producida por la nariz, que genera fallas en el seguimiento del punto ubicado sobre el arco de cupido del labio, en consecuencia aumenta el error del cálculo de apertura.

Para disminuir los errores detectados, se diseñó un ambiente de luz controlada que cuenta con un aro de luz led, que provee luz blanca de 5000K a 7000K (Temperatura de color para lámparas) ubicada frente al rostro del sujeto a 50 cm. Su implementación se observa en la figura 17.

Figura 17: Prueba con luz controlada.



La tabla 9 presenta los resultados con respecto a tiempos de procesamiento promedio por vídeo, con lo cual se afirma que al igual que los vídeos de presentadores se tarda

1,7 veces según la duración del vídeo a procesar. Por último, se afirma que su tasa de frames por segundo es aproximada a la de los presentadores (18 frames/seg).

Tabla 9: Tiempos de procesamiento para cada vídeo con luz controlada.

Persona	Tiempo Promedio [s]	Duración vídeo [s]	Número de Frames	Tasa Frames/segundo
01	13,84	8	258	18,64
02	14,3	8	258	18,04
03	9,37	5	160	17,07

La tabla 10 presenta el error para cada una de las pruebas realizadas a diferentes personas. De esta tabla se muestra que al realizar ajuste a la dirección e intensidad de luz es posible reducir el error en el calculo de la apertura de los labios. La figura 18 presenta algunas de las personas a la cuales se realizaron las pruebas bajo la mismas condiciones de luz.

Tabla 10: Error de la apertura de cada vídeo

Persona	Error Apertura [%]
01	2,204
02	2,584
03	3,445
04	2,343
05	2,99

Figura 18: Resultados con aplicación de luz controlada.

(a) Persona01





(b) Persona02



(c) Persona03

2.3. LIMITACIONES DEL MÉTODO PROPUESTO

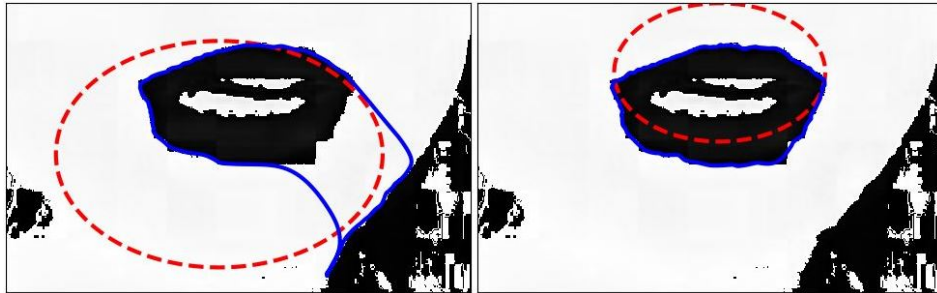
La calidad de la imagen es un factor fundamental en cualquier proceso relacionado con el tratamiento de imágenes. En particular, en el caso del algoritmo de Viola-Jones [18] se planteó una ventana de detección mínima de 220x220 píxeles con el fin de que el área de interés cuente con la información necesaria para lograr un resultado correcto en la aplicación del algoritmo. Esto implica una limitación en la relación del tamaño de la cara con respecto a la imagen completa.

De otra parte, el tipo de luz que se aplica para cada vídeo puede llegar a ser un inconveniente. En algunos casos se presenta una combinación entre luz blanca y amarilla que afecta la tonalidad de la piel y de los labios provocando que se asemejen, por lo que se hace difícil distinguir los labios. Mediante el cambio los parámetros planteados inicialmente en el procesamiento de color es posible solucionar este problema, como se muestra en la figura 7; sin embargo, el algoritmo ya no podría funcionar de manera completamente automática.

El resultado de la extracción de contorno está ligado en este caso a la posición inicial de una elipse definida por puntos, como se observa en la figura 19 la definición es alterada si se define una elipse que abarque o este cercana a áreas diferentes a los labios,

por este motivo es necesario fijar dicha elipse de manera manual pues la posición de los labios no es constante en los distintos vídeos.

Figura 19: Comparación extracción del contorno mediante *Snakes*



Durante las pruebas realizadas se observó que en algunos casos a causa de un fallo en la detección del área de interés se producen saltos durante la aplicación del seguimiento, generando fallos en la ubicación de los puntos de interés como se observa en la figura 20.

Figura 20: Fallo del Seguimiento



CONCLUSIONES

Inicialmente se aborda el algoritmo relacionado con la extracción automática del rostro en imágenes, el cual se decide utilizar el método propuesto por Viola-Jones, debido a su probabilidad de verdaderos positivos del 99,9% y su bajo consumo de recursos computacionales. Como bien se explica en este trabajo, presenta limitaciones con respecto al grado de inclinación lateral de la cabeza en la imagen; aún así, los resultados obtenidos al adaptar el algoritmo de Viola-Jones fueron óptimos.

El procesamiento de color permite segmentar el área de interés en dos regiones, logrando obtener el área de los labios con una mejor definición. Su tiempo de ejecución es similar al tiempo que toma una conversión RGB a escala de grises. Además al ser normalizado para que sus datos estén definidos por intensidades de 0 a 255, posibilita que en su aplicación se usen librerías de procesamiento de imágenes que usan escala de grises. Se observa que, la iluminación constante y frontal es pieza clave en el resultado del mismo.

Para realizar la extracción del contorno de los labios, se tomaron en cuenta métodos como el modelo de contornos activos (*Snakes*) y la adaptación realizada por Chan-Vese. El procesamiento de color facilita la aplicación de ambos métodos generando un porcentaje de éxito similar, por lo cual la comparación se realiza en base al tiempo de ejecución de cada uno bajo las mismas condiciones. A partir de los resultados se determina que para el proyecto es conveniente utilizar *Snakes*.

Por último, se analizó el comportamiento del seguimiento, donde a partir del estado del arte se optó por hacer uso del flujo óptico, donde el que mejor presenta resultados es el propuesto por Lucas-Kanade. Con respecto a los resultados su uso depende del ajuste generado por la extracción del contorno, la correcta definición de los parámetros

usados en el procesamiento de color y la calidad del vídeo a tratar. Se obtuvieron buenos resultados para el cálculo de la apertura de los labios realizando pruebas con diferentes tipos de iluminación; sin embargo, fue necesario hacer ajustes pequeños en los parámetros de configuración de color α , β y γ .

RECOMENDACIONES PARA TRABAJOS FUTUROS

En este proyecto se realizó una aproximación al cálculo de la apertura de los labios a partir de técnicas de segmentación de imágenes. Para ser llevado a cabo en tiempo real se sugiere implementar un modelo diferente al de *Snakes*, debido a que su tiempo de ejecución no es el adecuado.

Se sugiere configurar el tamaño de las ventanas de seguimiento de acuerdo a la posición de los puntos que definen el contorno. Por otro lado, como otra alternativa se plantea la aplicación de un método de extracción de contornos aproximadamente cada 100 frames con el fin de ajustar nuevamente los puntos de seguimiento para mejorar los resultados.

Se recomienda el uso de cámaras que provean una calidad de vídeo de al menos 720p para que los resultados del método propuesto sean adecuados. Aunque el algoritmo tiene la capacidad de trabajar con una menor resolución.

Bibliografía

- [1] CHAN, T. F., AND VESE, L. A. Active contours without edges. *IEEE Transactions on Image Processing* 10, 2 (2001), 266–277.
- [2] CHIN, S. W., SENG, K. P., AND ANG, L. M. Lips contour detection and tracking using watershed region-based active contour model and modified H Infinite. *IEEE Transactions on Circuits and Systems for Video Technology* 22, 6 (2012), 869–874.
- [3] DEL, S., AND EXTERNO, C. Seguimiento del contorno externo de la boca en imágenes de vídeo. *Revista Ingenierías Universidad de Medellín*, 64 (2009), 129–144.
- [4] EVENO, N., CAPLIER, A., AND CODON, P.-Y. A new color transformation for lips segmentation. 2001.
- [5] HAAR, A. Zur theorie der orthogonalen funktionen systeme. *Mathematische Annalen* 69, 3 (1910), 331–371.
- [6] KASS, M., WITKIN, A., AND TERZOPOULOS, D. Snakes: Active contour models. *International Journal of Computer Vision* 1, 4 (1988), 321–331.
- [7] LALITHA, S. D., AND THYAGHARAJAN, K. K. A study on lip localization techniques used for lip reading from a video. *International Journal of Applied Engineering Research* 11, 1 (2016), 611–615.
- [8] LUCAS, B. D., AND KANADE, T. An iterative image registration technique with an application to stereo vision, 1981.
- [9] MARR, D., AND HILDRETH, E. Theory of Edge Detection. *Proceedings of the Royal Society of London B: Biological Sciences* 207, 1167 (1980), 187–217.

- [10] MORSE, B. S. Lecture 21 : Image Understanding. *Image (Rochester, N.Y.)* (2000), 1998–2000.
- [11] PAPAGEORGIOU, C. P., AND OREN, M. A general framework for object detection. *Computer Vision, IEEE International Conference on 0*, January (1998), 555–562.
- [12] PRIETO, F. Automatic Outer Lip Contour Extraction in Facial Images. *IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing* (2010), 336–339.
- [13] SAEED, U., AND DUGELAY, J. L. Combining edge detection and region segmentation for lip contour extraction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6169 LNCS* (2010), 11–20.
- [14] SÁNCHEZ, M. U. R., MATAS, J., AND KITTLER, J. statistical chromaticity models for lip tracking with b-splines. In *Lecture Notes in Computer Science*, vol. 1206. 1997, pp. 69–76.
- [15] STILLITTANO, S., GIRONDEL, V., AND CAPLIER, A. Lip contour segmentation and tracking compliant with lip-reading application constraints. *Machine Vision and Applications 24*, 1 (2013), 1–18.
- [16] VALLEJO CÓRDOBA, S. L. *Extracción de los puntos característicos MPEG-4 de los contornos labiales sobre secuencias de vídeo*. PhD thesis, 2010.
- [17] VIOLA, P., AND JONES, M. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition (CVPR) 1* (2001), I—511—I—518.
- [18] VIOLA, P., AND JONES, M. J. Robust Real-Time Face Detection. *International Journal of Computer Vision 57*, 2 (2004), 137–154.
- [19] YANG, J., LU, W., AND WAIBEL, A. Skin-Color Modeling and Adaptation. *Lecture Notes in Computer Science 1352*, May (1997), 687–694.