



Dyna

ISSN: 0012-7353

dyna@unalmed.edu.co

Universidad Nacional de Colombia

Colombia

SALAZAR, AUGUSTO; PRIETO, FLAVIO

Extracción y clasificación de posturas labiales en niños entre 5 y 10 años de la ciudad de manizales

Dyna, vol. 73, núm. 150, noviembre, 2006, pp. 175-188

Universidad Nacional de Colombia

Medellín, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=49615007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

EXTRACCIÓN Y CLASIFICACIÓN DE POSTURAS LABIALES EN NIÑOS ENTRE 5 Y 10 AÑOS DE LA CIUDAD DE MANIZALES

LIP POSTURES EXTRACTION AND CLASSIFICATION FOR 5 TO 10 YEAR OLD CHILDREN FROM MANIZALES CITY

AUGUSTO SALAZAR

*GTA Percepción y Control Inteligente, Facultad de Ingeniería y Arquitectura, Universidad Nacional de Colombia-Manizales,
aesalazarj@unal.edu.co*

FLAVIO PRIETO

Facultad de Ingeniería y Arquitectura, Universidad Nacional de Colombia, Sede Manizales, faprieto@unal.edu.co

Recibido para revisar 2 de Septiembre de 2005, aceptado 13 de Marzo de 2006, versión final 20 de Abril de 2006.

RESUMEN: Se presentan los diferentes algoritmos y procedimientos utilizados en el desarrollo de un sistema de extracción y clasificación automática de posturas labiales. El sistema se diseñó con el fin de acompañar a los niños con labio y/o paladar hendido corregido, en el proceso de recuperación. Para la detección facial, se trabajan técnicas basadas en el espacio de color YC_bC_r y análisis de conectividad. La detección del contorno de los labios se realiza mediante técnicas de proyecciones, análisis de color (espacio de color HSV y Exclusión de Rojo) y la información de bordes del operador $SUSAN$. La extracción de la información discriminante se hace por diferentes tipos de análisis estadístico a partir de la región descrita por el contorno. La clasificación de las posturas se realiza empleando diferentes tipos de clasificadores.

PALABRAS CLAVE: Contorno, Labios, Labio paladar hendido, Detección, Extracción, Identificación, Automática, Dimensionalidad, Análisis estadístico, Clasificadores.

ABSTRACT: In this work, different algorithms and procedures used in the development of a system for the automatic extraction and classification of the lips postures, are presented. System was designed with the purpose of help to the children with lip and cleft palate in the recovery process after reconstruction surgery. For facial detection, techniques based on the YC_bC_r color space together with connectivity analysis, are used. In order to get a close model of the lip contour for features extraction, projection techniques, color analysis (HSV and red exclusion) and $SUSAN$ edge operator are used. For determining discriminant characteristics from the region described by the contour, different types of statistical analysis, are used. Classification of lips postures is made using different classifiers.

KEYWORDS: Contour, Lips, Cleft lip and palate, Detection, Extraction, Identification, Automatic, Dimensionality, Statistical analysis, Classifiers.

1. INTRODUCCIÓN

El procesamiento de imágenes faciales es una de las áreas de mayor desarrollo dentro de los sistemas de visión artificial. Tiene aplicaciones en áreas como seguridad, seguimiento, interfaces hombre-máquina, medicina, entre otras. Una de las aplicaciones médicas es la antropometría facial

que tiene como objetivo computar las medidas características del rostro. Para el caso de este trabajo se hace énfasis en el área de la boca, específicamente en el contorno labial.

Varias aproximaciones han sido propuestas en el campo de la detección del contorno de los labios. Algunas de ellas con base en el análisis de niveles de gris [1], análisis de color, modelos de plantillas

basados en contornos dinámicos, modelos de forma activa [2], plantillas deformables [3], entre otros. En [4] se presenta un sistema para segmentación y extracción de características orientado al reconocimiento del habla, el problema es que requiere de un hardware específico compuesto por una cámara especial montada sobre el rostro del sujeto, lo que representa una gran desventaja frente a dispositivos donde no se hace ningún contacto con el sujeto.

Este informe muestra el desarrollo de un sistema de extracción y clasificación automática de posturas labiales que funciona bajo condiciones de no invasividad, es decir, no requiere colocar ningún dispositivo sobre el sujeto que se quiera analizar. El proceso está dividido en dos partes: detección automática del contorno exterior de los labios (*DACEL*) y clasificación automática de posturas labiales.

El proceso de *DACEL* esta compuesto de varias etapas. La primera consiste en ubicar el sujeto dentro de la imagen para lo que se utiliza un modelado del tono de la piel basado en la información de la transformación al espacio de color YC_bC_r [5]. La segunda etapa se enfoca en ubicar la posición de la boca, lo cual se hace mediante el realce del tono de los labios [3] y análisis de conectividad [6]. Por último, se detecta el contorno exterior haciendo una preidentificación de la región de análisis para que la búsqueda se enfoque a ciertas regiones de la boca; el procedimiento de detección del contorno se basa en técnicas de proyecciones, análisis de color comparación entre transformación *HSV* y exclusión de rojo [7]) y análisis de distribución espacial de los píxeles de la imagen de bordes *SUSAN* [8].

Para la clasificación automática de las posturas labiales se realiza una preselección de la información mediante: análisis estadístico del espacio de características en clasificación múltiple [9] y análisis discriminante de Fisher [10]. El proceso de clasificación se lleva a cabo mediante la comparación de los resultados de varios tipos de clasificadores: lineal [11], Bayesiano [11], máquinas de vectores de soporte [12] y redes neuronales [13].

El sistema es de mucha utilidad al evaluar la evolución de la cicatrización de los labios de los niños con labio y/o paladar hendido corregido (*LPHC*). La evaluación se hace mediante la comparación de: los registros del paciente antes y después de las intervenciones quirúrgicas y los patrones de niños que no presentan ninguna alteración en la región de la boca. Por lo tanto, los algoritmos fueron probados sobre una base de datos representativa de la población infantil (entre 5 y 10 años) de Manizales [14].

2. MARCO EXPERIMENTAL

Para el desarrollo del sistema de extracción y clasificación de posturas labiales propuesto en este trabajo, se utiliza una parte de la base de datos de imágenes faciales recogida para el estudio antropométrico descrito en [14], el cual fue dirigido a la población de niños y niñas entre 5 y 10 años de la ciudad de Manizales, específicamente todos aquellos inscritos en instituciones oficiales, no oficiales, rurales o urbanas, bajo el supuesto de una cobertura en educación total en todos los conjuntos poblacionales de la ciudad de Manizales.

La información sobre la población objetivo proviene de las bases de datos de la Secretaría de Educación de la ciudad de Manizales, cuyo control esta ejercido a través de los formularios C100 y C600 establecidos por el DANE, en los cuales cada institución, debe entregar los datos precisos de su infraestructura y población (estudiantes matriculados y profesores), respectivamente.

La muestra poblacional utilizada fue de 660 sujetos (mitad niños, mitad niñas) a los cuales se les tomaron dos fotos pronunciando cada uno de los fonemas vocales, resultando un total de 6600 imágenes (1320/fonema). La geometría de adquisición (Figura1) determina que en la escena, solo aparezca un sujeto en posición frontal, pronunciando un fonema vocal. Las imágenes son a color y poseen una dimensión de 2560 x 1920 píxeles y están almacenadas en formato JPEG.



Figura 1. Escena de adquisición.
Figure 1. Acquisition scene.

3. TÉCNICAS DE ANÁLISIS

Con el fin de disminuir el costo computacional, el análisis de las imágenes se hace mediante técnicas de bajo nivel [15], haciendo uso de métodos, tales como, el análisis de imágenes en escala de grises [1], la extracción de bordes horizontales [3] y los bordes obtenidos con el operador SUSAN [8].

3.1 Segmentación

La identificación del área de la imagen en la que se encuentra el sujeto, se realiza mediante el detector del tono de piel (*DTP*) propuesto en [5]. Para la extracción de características en el área de la boca, se comparan dos técnicas:

3.1.1 Exclusión de rojo

Debido a que la región de los labios es predominantemente roja, se resalta el contraste con las otras componentes para lograr una exclusión de rojo. Los valores de verde y azul son combinados como se indica en la Ecuación 1 [7].

$$f(G, B) = \log(G/B) \quad (1)$$

En [7] el umbral β toma un valor fijo, alternativamente, en este trabajo, el valor β se determina promediando el máximo y el mínimo de la fila donde la suma de los valores de gris es mínima.

3.1.2 Filtrado de tono

Esta técnica utiliza la información del tono la transformación HSV (tono, saturación y valor).

Este método se centra en resaltar un valor de tono definido h_0 que en este caso corresponde al tono de los labios, para esto se utiliza el filtro que se define en la Ecuación 2 [3].

$$f(h) = \begin{cases} 1 - \frac{(h-h_0)^2}{w^2} & |h-h_0| \leq w \\ 0 & \text{en otro caso} \end{cases} \quad (2)$$

Donde h representa el valor del tono y w controla la distancia en la que el valor del tono cae alrededor de cero.

4. LOCALIZACIÓN DE LA BOCA

La tarea del reconocimiento automático de posturas labiales, comienza con la detección del sujeto dentro de la escena, posteriormente, se deben encontrar las coordenadas que determinan el área de la boca. Una vez definida la región de interés (imagen boca), se procede a hacer la extracción de las características que serán posteriormente analizadas.

4.1 Detección del Sujeto en la Imagen

Al tener la certeza de que solo existe un sujeto en la escena, se debe definir con precisión en que lugar de la imagen se encuentra. Para esto, se utiliza el *DTP* [5]. Las etapas más relevantes del proceso son:

1. Escalar la imagen para disminuir el costo computacional. Se trabaja con la imagen reducida al veinte por ciento (20%) de sus dimensiones originales.
2. Segmentar la imagen utilizando el *DTP* (Figura 2(b)).
3. Separar las regiones obtenidas en la segmentación, mediante análisis de conectividad (Figura 2(c)) [6].
4. Agrupar las regiones que componen el conjunto facial (*CF*) del sujeto (Figura 2(d)).

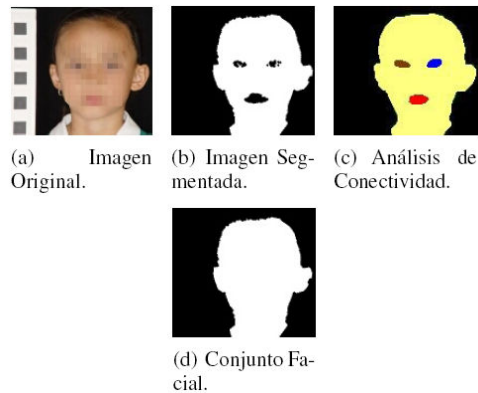


Figura 2. Ubicación del sujeto.
Figure 2. Finding the face.

4.2 Detección de la región de la boca

Una vez se tiene definida la porción de la imagen que contiene al sujeto, se procede a determinar la región donde se halla la boca. El procedimiento de detección por análisis de regiones predominantes (*DARP*) se desarrolló para este trabajo.

Método *DARP*

La detección de la región de la boca mediante de esta técnica consiste:

1. El espacio de búsqueda se reduce a los píxeles que pertenecen al *CF* (Figura 16(a)).
2. Basándose en la altura del *CF*, recortar la tercera parte superior (Figura 16(b)).
3. Erosionar [16] la imagen (Figura 16(c)) para eliminar alguna influencia del fondo que pudo haber quedado luego de la segmentación.
4. Recuperar la información de color, pero solamente en las coordenadas del *CF* (Figura 16(d)).
5. Transformar la imagen a color al espacio HSV y filtrar la componente de tono, utilizando la Ecuación. 2 (Figura 3(e)).
6. Umbralizar la componente de tono filtrada, de tal manera que en la imagen binaria resultante, se resalten los píxeles candidatos para la región de la boca de los demás componentes de la imagen (Figura 3(f)).
7. Mediante análisis de conectividad, encontrar la región más predominante, que para este caso corresponde a la boca (Figura 3(g)).

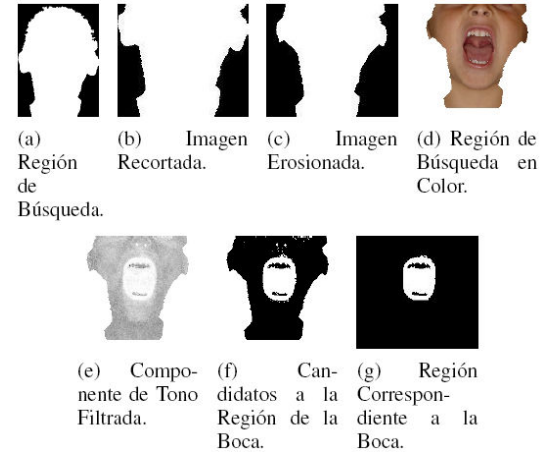


Figura 3. Ubicación de la boca.
Figure 3. Finding the mouth.

Sobre la imagen de la boca, se hará la extracción y el análisis de las características, con las que se intentará identificar el fonema vocal.

5. EXTRACCIÓN DE CARACTERÍSTICAS VISUALES

Existe un conjunto de características visuales, que describen de manera muy aproximada la postura labial para un fonema vocal específico [17]. Debido a que el diagnóstico se hace de forma no invasiva, características tales como la posición del velo del paladar o de la lengua son difíciles de determinar, por lo tanto, se desarrolló un sistema basado en la extracción y análisis del contorno exterior de los labios.

5.1 Aproximación del contorno exterior de los labios

La aproximación del contorno, se basa en la detección de ocho puntos que en conjunto, forman un polígono que describe de manera aproximada la frontera externa de los labios. El proceso comienza con la detección de dos puntos llamados vértices (puntos amarillos Figura 4(a)), posteriormente, se toman las coordenadas de éstos como punto de partida para la obtención de los demás puntos que forman el contorno exterior de los labios (puntos azules Figura 4(a)).

Una vez encontrados los puntos, se trazan líneas que definen el modelo utilizado para la recopilación de las características (Figura 4(b)).

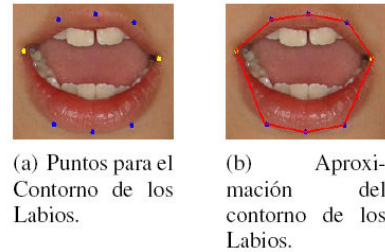


Figura 4. Descripción del modelo del contorno exterior de los labios.

Figure 4. Description of the external lips contour model.

5.2 Extracción de los vértices

Los puntos donde se encuentran la frontera externa superior e inferior de los labios, son llamados vértices. La extracción de estos se realiza utilizando dos alternativas: el método del gradiente vertical (*GV*) propuesto en [3] y el método de reducción del espacio de búsqueda a partir de la segmentación (*REBPS*), este último desarrollado para este trabajo. Se aclara, que los umbrales de trabajo se determinan empíricamente, basándose en resultados obtenidos en pruebas de segmentación.

Método *REBPS*

Este método aprovecha los resultados de la segmentación, para tratar de minimizar el error en la búsqueda. Se trabaja sobre la imagen en escala de grises.

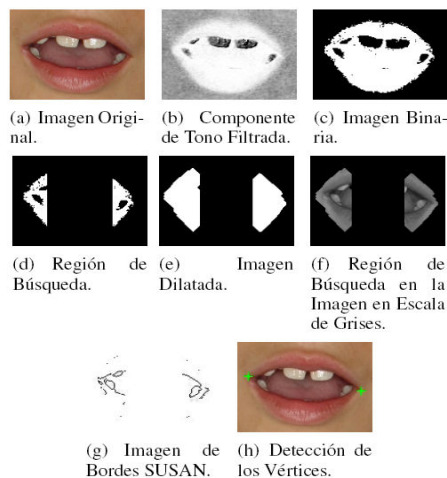


Figura 5. Detección de los vértices de la boca.

Figure 5. Mouth vertex detection.

La detección de los vértices por medio de esta técnica consiste:

1. Transformar la imagen a color (Figura 5(a)) al espacio HSV y filtrar la componente H utilizando la Ecuación. 2 (Figura 5(b)).
2. Convertir la imagen en escala de grises a una imagen binaria (Figura 5(c)). El umbral utilizado es de 240/255.
3. Hacer análisis de la distribución espacial de los píxeles en blanco, para obtener la media y la desviación estándar horizontal.
4. Definir el espacio de búsqueda como las bandas ubicadas a 1.5 y 2.8 veces la desviación estándar, a izquierda y derecha de la media (Figura 5(d)).
5. Dilatar [16] la imagen para aumentar el espacio de búsqueda (Figura 5(e)).
6. Pasar la imagen original a escala de grises y dejar solamente la información de la ventana definida en el paso anterior (Figura 5(f)).
7. Hallar los bordes utilizando el operador *SUSAN* [8] (Figura 5(g)).
8. Determinar la posición vertical de los vértices calculando la media de la distribución de los píxeles en cada una de la bandas de la imagen *SUSAN*.
9. Por último, la posición horizontal de los vértices se toma de la posición de los píxeles que estén más al extremo de cada una de las bandas (Figura 5(h)).

5.3 Detección de los puntos del contorno

Para la ubicación de los puntos del contorno se utiliza la imagen de tono filtrada o la imagen de exclusión de rojo. El proceso de extracción de los puntos se realiza de acuerdo a lo presentado en [3], de la siguiente manera:

1. Se define la región de búsqueda como la banda ubicada en el centro de la distancia entre los vértices (10 columnas a la izquierda y 10 a la derecha) (Figura 6).
2. Se hace la suma de los valores de intensidad de cada fila, obteniéndose una distribución $P[r]$.
3. Se calcula $P[r] - \langle P[r] \rangle$ ($\langle P[r] \rangle$ equivale a la media de $P[r]$), obteniendo $P[r]$ (Figuras 7(a) y 7(b)).

4. La posición vertical de los puntos se determina como el ancho M_W que se muestra la Figura 7.

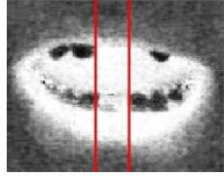


Figura 6. Región de búsqueda de los puntos del contorno.

Figure 6. Contour points search region.

La extracción de los otros cuatro puntos se efectúa ubicando la banda de búsqueda a un cuarto y a tres cuartos de la distancia entre vértices.

6. EVALUACIÓN DE CARACTERÍSTICAS

6.1 Características a evaluar

Con la información de los contornos extraídos, se calculan una serie de parámetros conocidos como descriptores por regiones. La evaluación de las características se hace a un conjunto de descriptores variantes e invariantes a traslación, rotación y escalado [16].

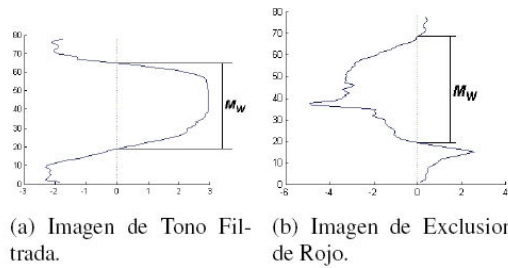


Figura 7. Distribución $P[r]$.
Figure 7. Distribution $P[r]$.

La Figura 8 exhibe las tres imágenes utilizadas para calcular los descriptores. El espacio sobre el cual se realiza el análisis está definido por el contorno y la región encerrada por éste. Los descriptores se calculan en base a función $f(x,y)$, la cuál, se define como la distancia del centro de masa de la región, a cada uno de los píxeles del contorno.

Para tener una mejor definición del perímetro, se debe tener en cuenta que los píxeles de contorno sean vecinos a cuatro [16].

La Tabla 1 muestra el conjunto de características sometidas a evaluación. d representa la densidad de la región, m_{pq} son los momentos bidimensionales de orden pq y ϕ_n son los n momentos invariantes [16]. Además, μ_f y σ_f corresponden a la media y la desviación estándar de $f(x, y)$.

Las características son calculadas para cada una de las tres regiones de análisis, lo que genera un total de 57 características por postura.

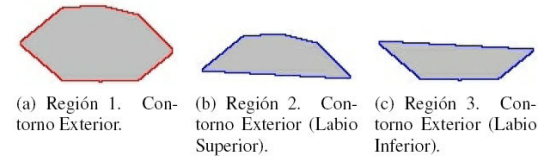


Figura 8. Regiones para el análisis.
Figure 8. Regions to be analyzed.

Tabla 1. Conjunto de características.
Table 1. Feature set.

Notación	Caract.	Notación	Caract.
1	d	11	ϕ_1
2	m_{01}	12	ϕ_2
3	m_{10}	13	ϕ_3
4	m_{11}	14	ϕ_4
5	m_{20}	15	ϕ_5
6	m_{02}	16	ϕ_6
7	m_{30}	17	ϕ_7
8	m_{03}	18	μ_f
9	m_{12}	19	σ_f
10	m_{21}		

6.2 Selección efectiva de características

La selección de los rasgos discriminantes, constituye el aspecto fundamental en un sistema de reconocimiento de patrones. Las propiedades que deben cumplir las características para su selección son las siguientes:

- Capacidad discriminante: es decir, que separen lo más nítidamente posible las clases existentes.

- Confiabilidad: esto es, los objetos de una misma clase deberán presentar la menor dispersión posible.
- Ausencia de correlación: es decir, evitar en lo posible la utilización de características que dependan fuertemente entre si.

Por lo anterior, el proceso de entrenamiento no se debe realizar, mientras no se disminuya al máximo la redundancia en cada una de las características. Para dicha eliminación, se implementaron dos técnicas que son descritas a continuación.

6.2.1 *Análisis estadístico del espacio de características en clasificación múltiple (AEECCM)*

Esta metodología tiene como objetivo encontrar un conjunto de características que evaluadas en grupo generen una mayor discriminancia entre clases, permitiendo la adecuada clasificación automática, y teniendo como criterios la menor probabilidad de error de clasificación. Lo cual puede verse también como la selección adecuada de características, altamente discriminantes y con baja redundancia de información [18].

La metodología en consideración conlleva la aplicación de los siguientes pasos descritos a través de la Figura 9 [9].

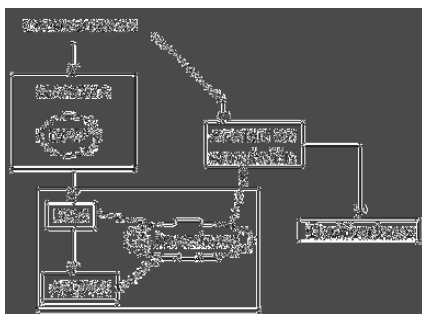


Figura 9. Diagrama de flujo AEECCM.
Figure 9. AEECCM flow-diagram.

6.2.2 *Análisis discriminante de Fisher*

El objetivo del análisis consiste en encontrar una función que retorne valores escalares que permitan una buena discriminación entre diferentes clases de los datos de entrada. Para realizar este análisis se cuenta con:

1. **Índices de Fisher:** estos índices se calculan de la siguiente manera:

$$F_{ij} = \frac{(m_i - m_j)^2}{\sigma_i^2 + \sigma_j^2} \text{ para todo } i > j \quad (3)$$

donde, ij es la combinatoria de clases; m_i , m_j son las medias y σ_i^2 , σ_j^2 son las varianzas de las clases i y j correspondientes. Este proceso se realiza para cada característica y se promedia, obteniéndose un índice de Fisher particular. Entre más alto el índice, más discriminante la característica [10].

2. **Análisis Multivariado de Discriminantes de Fisher (AMDF):** este análisis agrupa patrones de la misma clase y separa patrones de diferentes clases. Los patrones se proyectan de un espacio N-dimensional a un espacio $C-1$ dimensional, donde C es el número de clases de patrones. Por ejemplo, considérense dos conjuntos de puntos en un espacio bidimensional que se proyectan en una sola línea. Dependiendo de la dirección de la línea, los puntos se pueden mezclar o separar. Este análisis encuentra la línea que mejor separa los puntos [10].

7. CLASIFICACIÓN

7.1 Clasificadores implementados

Las metodologías de selección expuestas, crean conjuntos de características discriminantes. Por tal motivo, con el fin de determinar cuál es la mejor alternativa, para el caso tratado en este trabajo, se construyen y evalúan una serie de clasificadores:

1. Lineales [11].
2. Bayesianos [11].
3. Máquinas de Vectores de Soporte [12].
4. Redes Neuronales [13].

Se debe aclarar que los clasificadores evalúan los diferentes conjuntos de características, producto de la selección antes mencionada.

7.2 Evaluación de los clasificadores

7.2.1 Validación

La utilidad de los clasificadores, se evaluó midiendo el porcentaje de observaciones que fueron clasificadas correctamente, generando así, una estimación de la probabilidad de casos correctamente clasificados. Se utilizan tres métodos para estimar dicha probabilidad [19]:

- Prueba de muestras independientes.
- Validación cruzada con k-particiones.
- Leave-one-out (LOO).

7.2.2 Métricas de Desempeño

La evaluación de resultados se realiza a partir de la matriz de confusión. Es una matriz cuadrada de orden n , igual al número de clases. En las filas se representan las clases reales mientras que en las columnas se representan las clases asignadas por el clasificador. Las métricas de desempeño [20] toman valores de 0-1 ó de 0%-100%. Para la efectividad (PVP), la especificidad (PVN) y la precisión (VPP) el valor ideal es del 100%, y para las métricas, error positivo (PFN) y error negativo (PFP) el valor ideal es del 0%.

8. RESULTADOS

8.1 Localización de la boca

El método *DARP*, se pone a prueba sobre todas las imágenes de la base datos (1320/postura).

El criterio de evaluación de desempeño $f_{DM}(x)$, donde x es la imagen resultante del proceso, se muestra en la siguiente expresión.

$$f_{DM}(x) = \begin{cases} 1 & \text{si la boca es detectada en un 100\%} \\ 0 & \text{en otro caso} \end{cases}$$

La función $f_{DM}(x)$ fue evaluada por un experto para las 6600 imágenes resultantes en la prueba. En la Tabla 2 se tiene el resultado de la evaluación del desempeño, además del tiempo promedio requerido para el procesamiento de una imagen utilizando una CPU Pentium 4 a 2.8 GHz.

Tabla 2. Desempeño del algoritmo para la detección de la boca.

Table 2. Mouth detection algorithm performance.

Total Imágenes	Detección Óptima	Desempeño (%)	Tiempo [seg]
6600	6274	95.06	0.73716

La Fig. 10 muestra uno de los resultados de la prueba.

Como se puede ver en la Tabla 2, fueron muy pocas las imágenes en las que la detección no fue óptima, por lo tanto, los casos en donde el algoritmo falló, fueron corregidos manualmente, pues lo que se pretendía desde un principio, era organizar una base de datos de imágenes con las características visuales específicas que se analizan en este trabajo. Se tienen entonces, 6600 imágenes de posturas labiales. El formato de las imágenes es PPM. Las dimensión de las imágenes depende del sujeto y de la postura que se haya extraído.

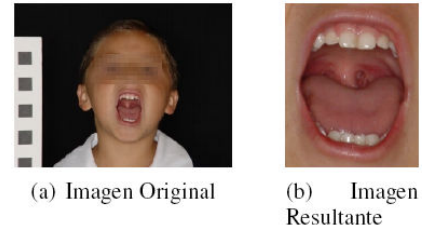


Figura 10. Resultado de la extracción de la boca.

Figure 10. Detected mouth.

8.2 Extracción de características

8.2.1 Pruebas de Localización de Puntos

Los diferentes algoritmos de extracción de características son probados empleando la base de datos obtenida de la prueba del método *DARP*.

Para el análisis de resultados de las diferentes técnicas, se tienen cuatro pruebas:

1. Localización de vértices (V_t) empleando el método de GV .
2. Localización de V_t empleando el método *REBPS*.
3. Localización de los puntos del contorno (CP) empleando la Imagen de Tono Filtrado.

4. Localización de los CP empleando la Exclusión de Rojo.

En cada una de las pruebas se obtiene la localización de los 8 puntos característicos, con sus coordenadas x e y ($Crd_{x,y}$), distribuidos a lo largo de la boca. Además, se mide el tiempo de computo empleado para el procesamiento de cada imagen. En la Tabla 3 se encuentra la definición de los 8 puntos del contorno exterior de los labios.

Tabla 3. Descripción de los puntos característicos obtenidos durante el procesamiento de la imagen.

Table 3. Notation for the extracted points.

Punto	Descripción
$(P_{0,x}, P_{0,y})$	$Crd_{x,y}$ del Vt izquierdo
$(P_{1,x}, P_{1,y})$	$Crd_{x,y}$ del Vt derecho
$(P_{2,x}, P_{2,y})$	$Crd_{x,y}$ del CP superior izquierdo
$(P_{3,x}, P_{3,y})$	$Crd_{x,y}$ del Vt superior central
$(P_{4,x}, P_{4,y})$	$Crd_{x,y}$ del Vt superior derecho
$(P_{5,x}, P_{5,y})$	$Crd_{x,y}$ del Vt inferior izquierdo
$(P_{6,x}, P_{6,y})$	$Crd_{x,y}$ del Vt inferior central
$(P_{7,x}, P_{7,y})$	$Crd_{x,y}$ del Vt inferior derecho

8.2.2 Desempeño de los algoritmos

Los resultados se evalúan a través de la comparación de los resultados obtenidos de manera automática, con los resultados del etiquetado manual de 1030 imágenes de prueba. A partir del error de cada muestra (Ecuación 4), se obtiene el valor del error promedio en píxeles y el error cuadrático medio.

$$\varepsilon_i = p_m - p_a \quad (4)$$

con p_m y p_a como coordenadas de los puntos manual y automático respectivamente.

Para la localización de los Vt , se obtuvieron mejores resultados con el método *REBPS* como se puede observar en la Figura 11. Los tiempos de proceso se muestran en la Tabla 4.

Tabla 4. Tiempos de proceso para la detección de los vértices.

Table 4. Vertex detection processing time.

	Tiempo [mseg]
<i>GV</i>	62.45
<i>REBPS</i>	63.06

La Figura 12 exhibe uno de los resultados de la localización de Vt .

En cuanto a la localización de los CP, se obtuvo un mejor desempeño con la extracción a partir de la información de tono filtrado, como lo ilustra la Figura 13.

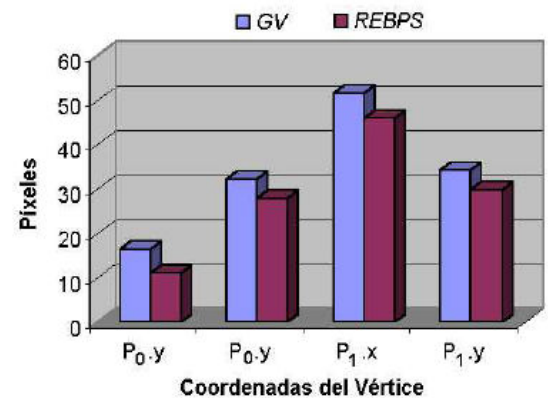
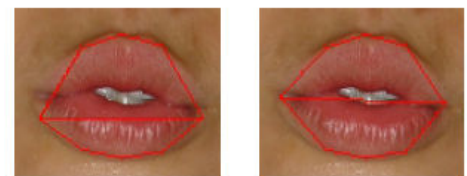


Figura 11. Error en la ubicación automática de los vértices.

Figure 11. Error of the vertex detection algorithms.



(a) Método del *GV*. (b) Método *REBPS*.

Figura 12. Localización de vértices.

Figure 12. Detected vertex.

La detección de los puntos que definen el contorno exterior de los labios, se puede observar en la Figura 14.

8.2.3 Pruebas en pacientes con labio y paladar hendido corregido (LPHC)

El objetivo principal de este trabajo, es su aplicación en el tratamiento en pacientes con *LPHC*. Por esta razón, se realizaron pruebas de detección del contorno exterior de los labios, las Figuras 15 y 16 exhiben algunos de los resultados.

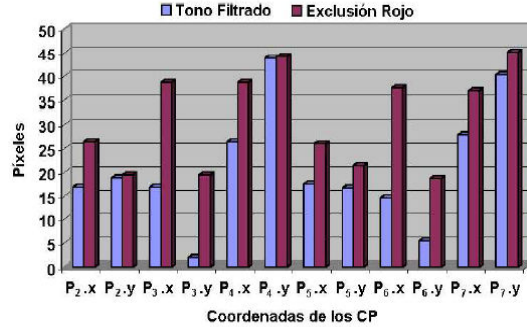


Figura 13. Error en la ubicación automática de los puntos del contorno.

Figure 13. Error of the contour points detection.

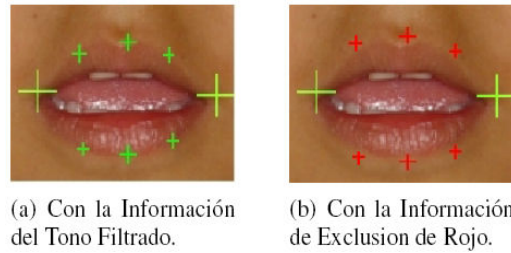


Figura 14. Localización de los puntos del contorno.

Figure 14. Detected contour points.

Estas imágenes fueron tomadas sin control de las condiciones de adquisición, lo que lleva a pensar que los algoritmos pueden llegar a funcionar adecuadamente indistintamente de la morfología de la boca y condiciones de adquisición.

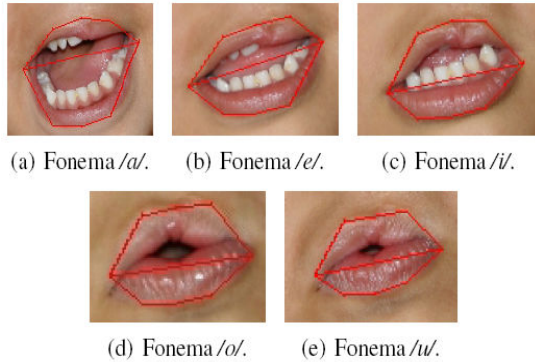


Figura 15. Detección del contorno en un paciente con *LPHC* pronunciando los cinco fonemas vocales.

Figure 15. Detected contour of CLP patient articulating the five vowel phonemes of Spanish.

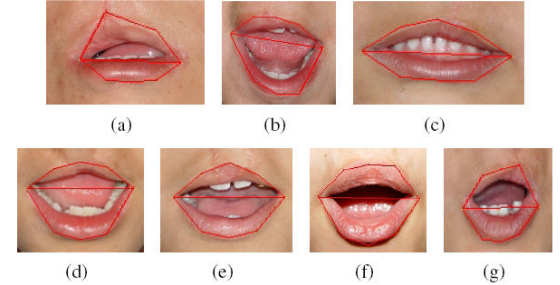


Figura 16. Detección del contorno en varios pacientes con *LPHC*.

Figure 16. Detected contour of different CLP patients.

8.3 Selección de características

8.3.1 *AEECCM*

El método de *MANOVA* y *ANOVA* realiza una observación de la comparación de las características clase a clase con el fin de obtener aquellas que son más discriminantes al momento de realizar la clasificación. En la Tabla 5, se pueden observar las características discriminantes después de realizar el análisis.

La notación se definió en la Tabla 1, el subíndice indica la región de la cuál se extrajo la característica según lo mostrado en la Figura 8.

Tabla 5. Características escogidas luego del *AEECCM*.
Table 5. Selected features using *AEECCM*.

	/a/	/e/	/i/	/o/	/u/
/a/		11 ₁ ,11 ₂	11 ₂ ,12 ₃	2 ₂ ,18 ₂	9 ₃ ,18 ₃
/e/	11 ₁ ,11 ₂		6 ₁ ,12 ₂	3 ₂ ,18 ₂	7 ₂ ,18 ₂
/i/	11 ₂ ,12 ₃	6 ₁ ,12 ₂		3 ₂ ,18 ₂	9 ₂ ,18 ₂
/o/	2 ₂ ,18 ₂	3 ₂ ,18 ₂	3 ₂ ,18 ₂		3 ₃ ,11 ₂
/u/	9 ₃ ,18 ₃	7 ₂ ,18 ₂	9 ₂ ,18 ₂	3 ₃ ,11 ₂	

8.3.2 *Análisis Discriminante*

En la Tabla 6 se pueden observar las 13 características cuyos índices de Fisher fueron los más altos. Según lo mostrado en la Tablas 5 y 6, las características más discriminantes fueron las medias (μ_f) de cada una de la regiones, seguidas por los momentos bidimensionales m_{pq} de orden inferior.

Los momentos invariantes (ϕ_n) no arrojaron información discriminante, debido a que en el momento del análisis, presentaban una alta dependencia.

Tabla 6. Características escogidas a partir del discriminante de Fisher.

Table 6. Selected features using Fisher discriminant.

Carac.	Índice	Carac.	Índice	Carac.	Índice
2 ₁	1,2915	2 ₂	1,4224	2 ₃	1,3705
5 ₁	1,0795	3 ₂	1,2474	3 ₃	1,0899
18 ₁	1,6428	4 ₂	1,0982	4 ₃	1,0241
		5 ₂	1,1081	5 ₃	1,119
		18 ₂	2,2	18 ₃	1,8942

Tabla 7. Separación de la observaciones.

Table 7. Selected samples.

	Buenas	Malas	Buenas (%)	Malas (%)
/a/	648	672	49.09	50.91
/e/	426	894	32.27	67.73
/i/	485	835	36.74	63.26
/o/	417	903	31.59	68.41
/u/	633	687	47.95	52.05

8.4 Clasificación

Una vez encontradas las características discriminantes (clase-clase y multiclase), se seleccionó el conjunto de datos que iban a ser utilizados para el entrenamiento. Para esto se partió de una base de 6600 observaciones (1320 para cada postura). Es de notar que no todas las observaciones corresponden a una postura correcta, por lo tanto, se realizó un proceso adicional de separación manual, del cuál se derivan los resultados mostrados en la Tabla 7.

Debido a que el diseño de los clasificadores se hizo tomando la misma probabilidad de aparición para todas las clases, cada conjunto de datos debe tener igual número de muestras. Teniendo entonces, 417 observaciones (para cada clase) de posturas correctas.

Ésta cantidad se tomó del menor número de muestras correctas, en este caso la observaciones correspondientes al fonema /o/. Se tiene entonces un 70% (291 muestras) para entrenamiento y el restante 30% (126 muestras) para la validación. Finalmente se forman varios conjuntos de datos, basándose en la selección de características discriminantes:

- *CAE (Conjunto del Análisis Estadístico)*: compuesto por las características mostradas en la Tabla 5 producto del AEECCM.
- *CAF (Conjunto del Análisis de Fisher)*: derivado del AMDF, con las 57 características.
- *CADF (Conjunto del Análisis Discriminante de Fisher)*: producto del AMDF, con las 13 características discriminantes mostradas en la Tabla 6.

Con estos tres conjuntos de datos fueron probados los diferentes clasificadores descritos a continuación.

8.4.1 Clasificador lineal

Con MANOVA y ANOVA se obtuvieron conjuntos de características discriminantes entre pares de clases (discriminancia biclase). Por tal motivo, se construyeron tantos clasificadores lineales como parejas de clases hay (en este caso 10 clasificadores).

Para medir el desempeño, cada muestra desconocida es evaluada sobre todos los clasificadores conformados. Para decidir a que clase pertenece dicha muestra, se asigna a la clase que más veces fue señalada por los clasificadores. En los casos en que se presenten igualdades en la cantidad de asignaciones de clase para una misma muestra, se reevalúa directamente sobre los clasificadores correspondientes a las parejas que presentaron dicha igualdad. Obteniéndose entonces, un error global del 33%.

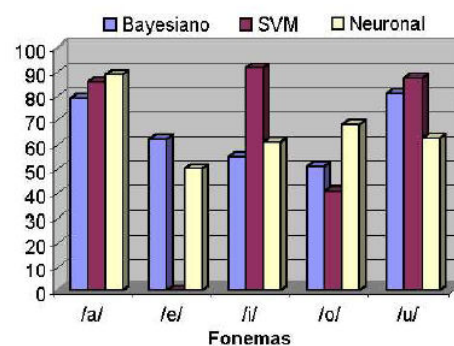


Figura 17. Efectividad (PVP) utilizando el CAF.
Figure 17. Effectiveness using the CAF.

8.4.2 Clasificador Bayesiano

Se construyeron 10 clasificadores Bayesianos biclase entrenados con el *CAE*. La evaluación se realizó de igual forma que con el clasificador lineal. Teniendo como resultado un error global del 32%. Adicionalmente se construyó un clasificador multiclase entrenado con el *CAF* y otro entrenado con el *CADF*.

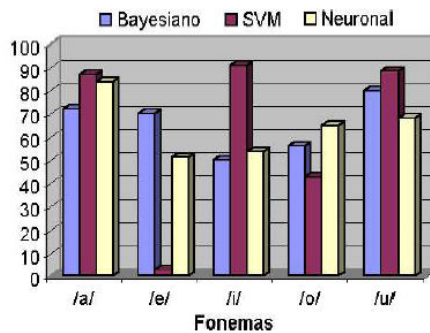


Figura 18. Efectividad (PVP) utilizando el *CADF*.

Figure 18. Effectiveness using the *CADF*.

8.4.3 Clasificador por SVM

La máquina de vectores de soporte fue implementada bajo la consideración de que los datos no son separables, ya que esto determina la constante de penalización de la *SVM* cuando las separaciones no son lineales. El kernel de la máquina es una función de base radial. Se entrenaron dos *SVM*, una con el *CAF* y otra con el *CADF*.

8.4.4 Red neuronal artificial

Se construyó una red multicapa con realimentación hacia adelante, teniendo una capa de entrada, una capa oculta y una capa de salida. En la capa de entrada se tienen tantas neuronas como características seleccionadas. En la capa oculta se trabajó con una rango de 4 a 10 neuronas y en la capa de salida se encuentran 5 neuronas, correspondientes a las clases determinadas por las posturas labiales de los 5 fonemas vocales. Se entrenaron dos redes, una con el *CAF* y otra con el *CADF*.

La efectividad (PVP) de los tres últimos clasificadores mencionados, es comparada en las Figuras 17 y 18.

Como puede verse, los clasificadores tienen un buen desempeño al separar los fonemas /a/ y /u/ y existe confusión al intentar separar las demás clases, esto debido a que como se mostró en la etapa de selección de características, aunque existe discriminancia, los índices son bajos. Es destacar el bajo desempeño de la *SVM*, en el momento de clasificar los fonemas /e/ y /o/, una posible causa de esto, es que la *SVM* hace un análisis geométrico de la distribución de los datos, por lo tanto en el momento en que se presente un traslape entre los datos de dos o mas clases, la *SVM* solo verá una sola. Esto se refleja en un alto desempeño en la clasificación de los fonemas /i/ y /u/, que se traslapan con los fonemas /e/ e /i/ respectivamente. Para probar esto, se agruparon los fonemas /e/ /i/ y los fonemas /o/ /u/, resultando en una mejora del desempeño trabajando únicamente con tres clases, como se ve en la Figura 19.

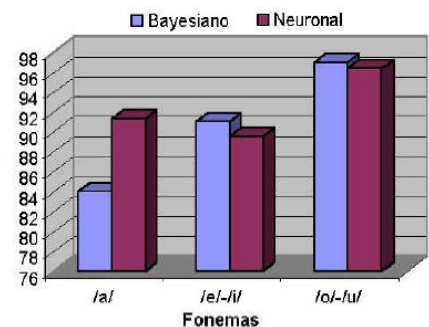


Figura 19. Efectividad (PVP) para las clases agrupadas. Utilizando el *CADF*.

Figure 19. Effectiveness for joint classes using the *CADF*.

9. CONCLUSIONES Y TRABAJO FUTURO

Las técnicas empleadas, basadas en la información de color y sus transformaciones en los diferentes espacios de color, la técnica de proyecciones, los detectores de bordes y el análisis de regiones, son invariantes a la rotación y traslación de la boca. De igual manera a si la boca se encuentra abierta o cerrada.

La evaluación de las características empleando el Análisis Estadístico y el Análisis Multivariado de Fisher permite reducir el costo computacional, sin

disminuir la eficiencia en la clasificación, garantizando, la integridad de los resultados.

Los errores de los diferentes clasificadores implementados permiten concluir que se puede trabajar con un clasificador muy sencillo como el Bayesiano para separar los cinco fonemas vocales. Los resultados obtenidos durante el desarrollo de este trabajo empleando imágenes de niños control pueden ser aplicados en el diagnóstico a niños con labio y/o paladar hendido corregido.

Utilizar el método *REBPS* para la detección de los puntos del contorno y combinar los resultados con técnicas que utilizan contornos deformables para con esto establecer una plantilla que describa la forma estándar de la boca de los niños en las diferentes edades.

Probar con imágenes en las que se tenga una buena diferenciación fonemas /e/-/i/ y /o/-/u/.

Obtener un conjunto de características adicionales (segmentar los dientes, evaluar su visibilidad y separación).

REFERENCIAS

- [1] MERSEREAU R., RAO R. Lip modelling for visual speech recognition. 28th Annual Asimular Conference on Signals, Systems, and Computer, IEEE Computer Society, 2, 1994.
- [2] LUETTIN J. Visual Speech and Speaker Recognition. PhD thesis. Department of Computer Science. University of Sheffield. 1997.
- [3] TORRESANI L. CAPRILE B. COIANIZ T. 2d deformable models for visual speech analysis. Istituto per la Ricerca Scientifica e Tecnologica. 1996.
- [4] LUTHON F. LIÉVIN, M. Lip features automatic extraction. Signal and Image Laboratory, Grenoble National Polytechnical Institute. 1998.
- [5] ABDEL-MOTTALEB M. JAIN A. HSU, R. Face detection in color images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 2002.
- [6] BESSLICH P., BASSMANN H. ADOCULOS. Digital Image Processing. International Thompson Computer Press.
- [7] POWERS D. M. W., LEWIS T.W. Audio visual speech recognition using red exclusion and neural networks. School of Informatics and Engineering, 2003.
- [8] SMITH S. M., BRADY J. M. SUSAN - a new approach to low level image processing. Int. Journal of Computer Vision, 23(1):45–78, May 1997.
- [9] DAZA G., SÁNCHEZ L. G., SUÁREZ J. F, CASTELLANOS G. Pca, kpca y manova sobre señales de voz en imágenes de posturas labiales y audio. Technical report, Grupo de Procesamiento Digital de Señales. Universidad Nacional de Colombia, 2004.
- [10] YAMBOR W. Analysis of pca-based and fisher discriminant-based image recognition algorithms. Technical Report. Fort Collins (Colorado). Colorado State University, 2000.
- [11] DUDA R., HART O., STORE D. G. Pattern Classification. Wiley-Interscience, New York, second edition, 2001.
- [12] VAPNIK V. The Nature of Statistical Learning Theory. Springer, NY, 1995.
- [13] MATHWORKS. Neural Network Toolbox: User's Guide - Version 4. For Use with MATLAB. The MathWorks, Inc., Natick, MA, 2000.
- [14] MEJÍA I. Extracción automática de características faciales para el estudio antropométrico en niños entre 5 y 10 años de la ciudad de manizales. Technical report, Grupo de Percepción y Control Inteligente. Universidad Nacional de Colombia, 2004.
- [15] WARK T., SRIDHARAN S., CHANDRAN V. An approach to statistical lip modeling for speaker identification via chromatic feature extraction. In Proceedings of the IEEE International Conference on Pattern Recognition, pages 123–125, August, 1998.

- [16] WOODS R., GONZÁLEZ R. Tratamiento Digital de Imágenes. Addison - Wesley Iberoamerica, S.A., 1996.
- [17] CORREDERA T. Defectos en la Dicción Infantil. Procedimientos para su Corrección. Editorial Kapelusz, 1949.
- [18] CARREIRA M. A. A Review of Dimension Reduction Techniques. Technical Report CS-96-09, Dept. of Computer Science, University of Sheffield, January 1997.
- [19] MARTINEZ W. L., MARTINEZ A. R. Computational Statistics Handbook with MATLAB. Chapman & Hall/CRC, 2002.
- [20] FERRI C., HERNÁNDEZ J., SALIDO M. A. Volume under the roc surface for multiclass problems. Exact computation and evaluation of approximations. Dep. Sistemes Informàtics i Computació, Univ. Politècnica de Valencia (Spain), April 2003.