SINTONIZACIÓN DE PARÁMETROS DE MODELOS POR MEDIO DE GA, EN EL RECONOCIMIENTO DE POSTURAS LABIALES

MODEL PARAMETERS TUNING USING GA, ON LIP POSTURES RECOGNITION

JORGE A. JARAMILLO
Universidad Nacional de Colombia, Manizales, Grupo Control y Procesamiento Digital de Señales, jajaramillog@gmail.com

JUAN R. GONZÁLEZ
Universidad Nacional de Colombia, Manizales, Grupo Control y Procesamiento Digital de Señales, emarshall311@hotmail.com

GERMÁN CASTELLANOS
Universidad Nacional de Colombia, Manizales, Grupo Control y Procesamiento Digital de Señales, gcastell@telesat.com.co

Recibido para revisar 15 de Diciembre de 2005, aceptado 6 de Abril de 2006, versión final 23 de Junio de 2006

RESUMEN: Se presenta un sistema de segmentación y reconocimiento de posturas labiales en imágenes por medio de la sintonización de modelos con parámetros ajustables: Para la segmentación se usa una plantilla deformable y para la caracterización se usa un modelo de variación de los contornos labiales. Ambos modelos son sintonizados por medio de algoritmos genéticos del tipo canónico, alcanzando desempeños de hasta el 86.5% y 76.6% respectivamente.

PALABRAS CLAVE: Algoritmos genéticos, Segmentación de imágenes, Caracterización, Clasificación de patrones, Análisis de componentes principales, Descomposición en valores singulares.

ABSTRACT: Here, a lip posture recognition and segmentation system by means of tuning adjustable parameter models, is shown. For segmentation, deformable templates are used, and for feature extraction a variational model of lip contours is implemented. Both models are tuned by canonical genetic algorithms yielding performances up to 86.5% and 76.6% respectively.

KEYWORDS: Genetic Algorithms, Image segmentation, Feature extraction, Pattern classification, Principal component analysis, Singular values descomposition.

1. INTRODUCCIÓN

En los últimos años se ha propuesto el empleo del reconocimiento de posturas labiales (*speechreading*) usando técnicas de procesamiento digital de imágenes, como complemento al análisis acústico para el reconocimiento de patologías asociadas a problemas o alteraciones de la emisión de voz [1].

La aplicación más difundida consiste en la creación de un modelo de los labios, cuyos parámetros sean ajustables para obtener versatilidad en las formas a reconocer. Sin embargo, una vez construido el modelo, la

sintonización de estos parámetros ha sido sometida a diversas restricciones prácticas. Por ejemplo, en [2] se utiliza el método de plantillas deformables descrito en [3]. El contorno de los labios es modelado por un conjunto de polinomios codificados que se ajustan al gradiente de la imagen considerada. La búsqueda en la imagen es llevada a cabo bajo la suposición de que se tienen bordes gruesos en los contornos de los labios. Esta suposición es violada en muchos casos reales, y además, la dependencia

sobre la elección inicial de los polinomios no permite modelar contornos con detalles finos. En [4], se presenta una aplicación usando *splines* [5] y filtros de Kalman en la que se impusieron restricciones de deformación a la plantilla limitando el número de grados de libertad. En [6] se describió un método basado en *snakes* [7], las cuales tienen la habilidad de resolver detalles finos en los contornos. Sin embargo, los contornos fueron restringidos a un subespacio aprendido del conjunto de entrenamiento.

Los algoritmos genéticos (GA) han sido empleados en la optimización de tareas con espacios de búsqueda muy amplios (en particular, en el procesamiento de imágenes), gracias a su capacidad de búsqueda en paralelo. Por ejemplo, en [8] se utiliza un modelo con seis parámetros variables, para detectar el ventrículo izquierdo en imágenes de ultrasonido del corazón. Un algoritmo genético lleva a cabo la búsqueda a través de la variación de los seis parámetros y evalúa qué tan bien se acopla el modelo a la imagen, mediante una función de evaluación que depende de la intensidad de gris de unos vectores perpendiculares y calculados sobre contorno modelado.

En [9], se presenta una aplicación para reconocimiento de rostros, el algoritmo genético es utilizado para buscar dentro de todas las posibles subimágenes del cuadro original en la que se quiere detectar un objeto determinado. Por esta razón, los parámetros a codificar corresponden a las posiciones de la esquina superior izquierda e inferior derecha del rectángulo que definirá la subimagen. En la fase de entrenamiento, un conjunto de imágenes previamente segmentadas y escaladas, conteniendo exclusivamente rostros, son entregadas al algoritmo para extraer las *características más expresivas* (MEF's) y las *características más discriminantes* (MDF's) [10]. Luego, con base en una organización jerárquica de estas características, el algoritmo genético busca cuál subimagen es la que más probable contiene un rostro humano.

Un método similar al anterior es el usado en [11] para la detección y verificación de rostros. En este caso, ya que el objetivo es la verificación, el entrenamiento se hace con varias imágenes del mismo sujeto y se pretende hallar la mayor concordancia de la subimágen con el modelo definido en el entrenamiento.
En [12], en lugar de variar directamente los parámetros que dan la forma a la imagen modelo, se utilizan *transformaciones afines* [13] para reconocer objetos en dos dimensiones. El algoritmo genético genera contornos a través de estas transformaciones y calcula las distancias entre cada punto del contorno modelado y el punto más cercano a éste en la imagen original, para calcular el error cuadrático medio y optimizar la ubicación de los contornos.

Teniendo en cuenta estos antecedentes, en este trabajo se plantea un método de segmentación y reconocimiento de posturas labiales mediante el uso de una plantilla deformable y un modelo de variación de los labios, los cuales son sintonizados por un algoritmo genético de dos etapas, específicamente, del tipo canónico. Este método de sintonización de los modelos, podría superar los problemas de implementación de las aplicaciones de *speechreading*, surgidos por el espacio de búsqueda resultante.

2. MODELOS Y SINTONIZACIÓN

2.1 Etapa de segmentación

Una plantilla deformable es un modelo matemático que se utiliza para rastrear los movimientos que un objeto específico [14]. Típicamente la plantilla se encuentra descrita por uno o varios polinomios que definen la intensidad de los píxeles en ésta.

Construcción de la plantilla deformable. Inicialmente, es necesario realizar la segmentación manual del conjunto de imágenes de entrenamiento (este caso se usaron 20) para cada vocal. Se construye una superficie promedio como primera aproximación a la plantilla deformable, mediante el submuestreo posterior normalización del tamaño de todas las imágenes (30×30 píxels). Luego se superponen las intensidades en las componentes R, G y B en la forma:

$$\bar{I} = \frac{\sum_{i=1}^{N} In_i}{N} \qquad (1)$$

donde $\bar{I}$ es la imagen promedio, N es el número de imágenes de entrenamiento, y In_i es el promedio de intensidad de la i imagen. En la Figura 1 se observa un ejemplo de las superficies promedio de las imágenes de posturas labiales para todas vocales, además de la boca cerrada.

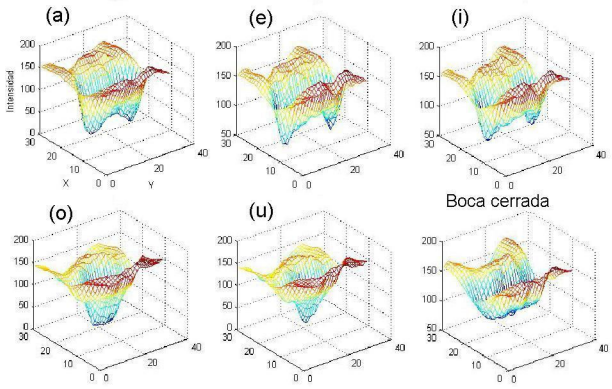


Figura 1. Superficies promedio de las posturas labiales.
Figure 1. Average surfaces for lip postures.

Seguidamente se realiza la aproximación de cada superficie promedio mediante el respectivo polinomio, que permita la deformación realizando la variación de sus coeficientes, y la cual corresponde a la solución de la ecuación normal [15]:

$$\mathbf{Ax} = \mathbf{b} \qquad (2)$$

La solución exacta, si $\mathbf{A}$ es una matriz cuadrada de rango completo, está definida como

$$x = \mathbf{A}^{-1}\mathbf{b} \qquad (3)$$

Cuando A no es cuadrada o no se cuenta con una solución exacta, se tiene la solución dada por:

$$x = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T b \qquad (4)$$

Para evitar el cálculo de $\left(A^T A\right)^{-1}$ se utiliza la aproximación por descomposición en valores singulares:

$$\mathbf{U\Sigma V}^T = \mathbf{A} \qquad (5)$$

$$x = \mathbf{V\Sigma}^{+}\mathbf{U}^T\mathbf{b} \qquad (6)$$

donde $\mathbf{\Sigma}$ y $\mathbf{V}$ son matrices ortogonales, por lo tanto sus inversas son iguales a sus transpuestas, $\mathbf{U}$ es una matriz diagonal y $\mathbf{\Sigma}^{+}$ es la transpuesta de $\mathbf{\Sigma}$, con sus elementos diagonales no nulos invertidos [15].

Teniendo en cuenta la ecuación normal de aproximación para las superficies polinomiales:

$$f(x,y) = \begin{cases} a_1x^n + \ldots + a_nx + a_{n+1}y^n + \ldots + a_{2n}y + \\ +a_{2n+1}x^{n-1}y + x^{n-2}(a_{2n+2}y^2 a_{2n+3}y) + \\ +x^{n-3}(a_{2n+4}y^3 + \ldots + a_{2n+7}y) + \ldots + \\ +x(a_{\frac{(n+1)(n+2)}{2}-n}y^{n-1} + \ldots + \\ +a_{\frac{(n+1)(n+2)}{2}-1}y) + a_{\frac{(n+1)(n+2)}{2}} \end{cases} \qquad (7)$$

entonces, la notación de $f(x,y)$ en forma matricial será:

$$\mathbf{A}_{k+\frac{(n+1)(n+2)}{2}} = \begin{pmatrix} x_1^n & \ldots & 1 \\ \vdots & \vdots & \vdots \\ x_k^n & \ldots & 1 \end{pmatrix} \qquad (8)$$

$$x = \begin{pmatrix} a_1 \\ \vdots \\ \frac{(n+1)(n+2)}{2} \end{pmatrix} \qquad (9)$$

De (8), se obtiene la ecuación de la forma $Ax = b$, donde b son los valores de intensidad de la imagen a ajustar. Los resultados concretos obtenidos para ajustar cada imagen promedio a una función de orden 11 se representan en la Fig. 2.

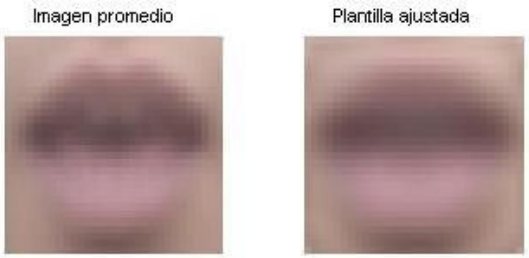


Figura 2. Imagen promedio y plantilla construida
Figure 2. Average image and constructed template.

Finalmente, los puntos evaluados x e y corresponden a los índices de los píxels de la imagen promedio. Por tanto, la plantilla puede ser deformada calculando la función en

intervalos menores a 1, si se quiere aumentar el número de píxels en una dirección, y mayores a 1 si se requiere disminuirlos. En la Figura 3 se muestra como se deforma la plantilla en la dirección x o escala $s1$ y en y o escala $s2$.

Figura 3. Ejemplos de deformación de la plantilla
Figure 3. Template variations examples.

Algoritmo de Segmentación. Para realizar la segmentación como tal, se construyó una función de error que consta de cuatro variables independientes correspondientes a la posición $(x1, x2)$ del punto central de la plantilla y las escalas $(s1, s2)$ correspondientes a la dimensión de la plantilla en las direcciones x e y respectivamente:

$$g(x1, x2, s1, s2) = \begin{cases} g(x1, x2, s1, s2), & cond = 1 \\ \infty, & cond = 0 \end{cases} \quad (10)$$

donde $cond = 1$ si:

$$g(x1, x2, s1, s2) = \begin{cases} 0 < (x1 - \frac{s1}{2}) \,\&\, (x1 + \frac{s1}{2}) < f \,\& \\ 0 < (x2 - \frac{s2}{2}) \,\&\, (x2 + \frac{s2}{2}) < c \,\& \end{cases} \quad (11)$$

Siendo f y c las filas y las columnas de la imagen a segmentar. En caso contrario, $cond = 0$. A fin de reducir el espacio de búsqueda, de acuerdo a la observación de la base de datos para las imágenes de 72×86, se tuvo en cuenta las siguientes consideraciones, siguiendo la cuales se agregaron condiciones al cálculo de la función de error, de forma que la nueva condición para el cálculo de la función de error es:

$$g(x1, x2, s1, s2) = \begin{cases} (0 < (x1 - \frac{s1}{2})) \,\&\, ((x1 + \frac{s1}{2}) < f) \,\& \\ (0 < (x2 - \frac{s2}{2})) \,\&\, ((x2 + \frac{s2}{2}) < c) \,\& \\ (s1 \le 28) \,\&\, (abs(s1 - s2) > 22) \,\& \\ (s1 < s2) \,\&\, (s1 > 0.4 s2) \,\&\, (x1 \ge 25) \,\& \\ (x2 \ge 25)(x1 \le 52 \,\&\, x2 \le 55) \end{cases} \quad (12)$$

La función $g1(x1, x2, s1, s2)$ está definida como:

$$\frac{1}{s1 s2} \sum_{i=x1-\frac{s1}{2}}^{x1+\frac{s1}{2}} \sum_{j=x2-\frac{s2}{2}}^{x2+\frac{s2}{2}} \left(I_{(i,j)} - P_{\left((i-x1+\frac{s1}{2}),(j-x2+\frac{s2}{2})\right)} \right) \quad (13)$$

Siendo I la imagen original que está siendo comparada con la plantilla deformada P.
La ecuación 13 depende de cuatro parámetros, dos de posición y dos de escalamiento, los que serán codificados y servirán como entrada al algoritmo genético que se describe en la sección 2.3, con el que se buscará minimizar esta ecuación, ya que cuando 13 sea mínimo, se sospecha de la existencia de un segmento correspondiente a la postura labial. La codificación se realiza transformando los parámetros en decimales a cadenas binarias independientes de 6 bits, esto por la reducción del espacio de búsqueda hecha en el algoritmo de segmentación.

2.2 Etapa de reconocimiento

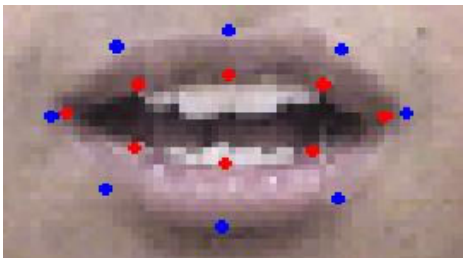
Construcción del modelo de variación de los contornos labiales

Sobre las imágenes segmentadas en la etapa anterior, se realiza una normalización automática para eliminar los efectos de rotación, traslación y escala, es decir, se busca que las bocas quedaran horizontales, centradas y con una longitud estándar de comisura a comisura, como se muestra en la Figura 4.

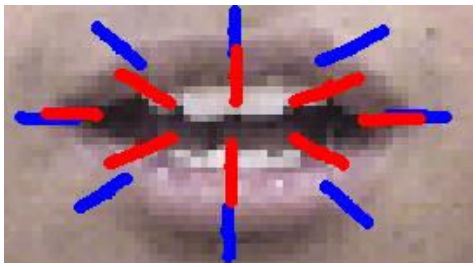
Figura 4. Imagen original (izquierda) y la imagen preprocesada (derecha)
Figure 4. Original image (left side) and preprocesed image (right side).

El sistema de caracterización que se implementó es propuesto en [16] y [17], el cual se basa en la construcción de modelos que tengan las principales formas de variación de la boca, tanto en forma como en la intensidad de la imagen en las cercanías de los contornos labiales. La construcción de estos modelos constituye la fase de entrenamiento del sistema de caracterización. Para construir el modelo de forma de la boca, a cada una de las fotos de entrenamiento del sistema le fueron señalados manualmente 16 puntos, que describen los contornos exterior e interior de los labios, como se muestra en la Figura 4.

Estos puntos son la referencia para la extracción de los vectores de intensidad alrededor de los contornos labiales. Dichos vectores son radiales al centro de la imagen y de una longitud de 10 pixels, siendo el pixel número 5 uno de los puntos previamente señalados, como se muestra en la Figura 4. Los vectores son extraídos de la matriz de *tono* de la imagen, que corresponde a la primera matriz del sistema *HSV*, por lo que la imagen se lleva antes a este espacio de color, en lugar del *RGB*.



(a) Puntos de referencia
(a) Referente points



(b) Vectores de intensidad extraídos
(b) Extracted intensity vectors

Figura 5. Modelo de forma de la boca
Figure 5. Mouth shape model

El objetivo de construir modelos de variación de los contornos labiales, es restringir el uso de suposiciones heurísticas acerca de las formas legales de variación de la boca, a la hora de realizar la búsqueda en una imagen [17].

En este caso, para obtener una *forma promedio*, se señalaron los contornos de un *conjunto de entrenamiento representativo*, conformado por 20 imágenes de la postura labial correspondiente a cada una de las vocales: /*a*/,/*e*/,/*i*/,/*o*/,/*u*/ y boca cerrada.

La *i-ésima* forma del conjunto de entrenamiento ($i = 1...N$) es representada por el vector v_i, de la forma:

$$\mathbf{v}_i = [\mathbf{x}_{i0}, \mathbf{y}_{i0}, \mathbf{x}_{i1}, \mathbf{y}_{i1}, ..., \mathbf{x}_{iN-1}, \mathbf{y}_{iN-1}]^T \quad (14)$$

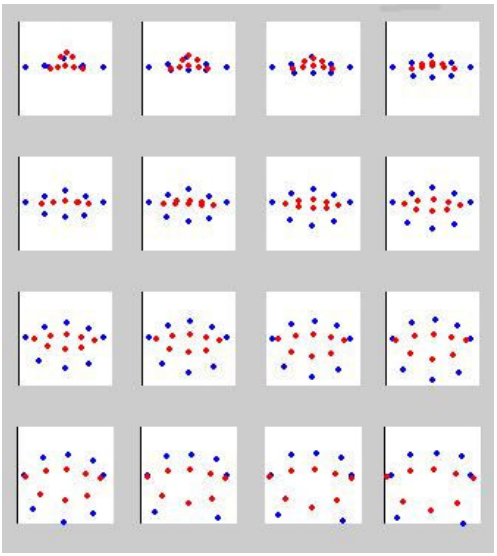
donde $\{x_{ij}, y_{ij}\}$ son las coordenadas del *j-ésimo* punto ($j = 1...32$) correspondiente a la *i-ésima* forma.

Teniendo las 120 formas de entrenamiento, se calcula la forma promedio $\bar{x}$ y la matriz de covarianza S.

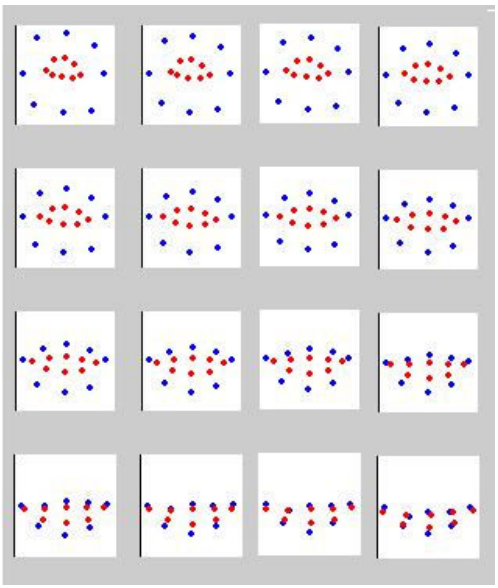
En calidad de características representativas del modelo se obtienen los vectores propios y valores propios de la matriz de covarianza usando *análisis de componentes principales* (PCA) [18]:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb} \quad (15)$$

donde $P = [p_1, p_2, p_3, ..., p_N]$ es una matriz con los N vectores propios correspondientes a los N valores propios más altos y b es un vector con los coeficientes correspondientes a cada eigenvector. La variación de estos coeficientes será la que dé el espacio de búsqueda para el algoritmo genético. Las Figuras 15 y 15 muestran las formas producidas por la variación de los dos primeros coeficientes, desde -3 desviaciones estándar hasta $+3$ desviaciones estándar, respectivamente.



(a) Primer modo de variación
(a) First variation mode



(b) Segundo modo de variación
(b) Second variation mode

Figura 6. Modos de variación del modelo de forma
Figure 6. Variation modes of the shape model

El modelo completo contiene 6 modos de variación que generan diversos detalles y asimetrías de las formas de variación del conjunto de entrenamiento y serán utilizados como características para la clasificación.
La búsqueda en cada imagen realiza mediante el modelo de las variaciones de intensidad alrededor de los contornos labiales, para lo cual se extraen los vectores de intensidad representados en la Figura 4, siendo g_{ij} el *j-ésimo* vector de la *i-ésima* imagen de entrenamiento, con los cuales se conforma la matriz respectiva:

$$\mathbf{h_i} = [\mathbf{g_{i0}}, \mathbf{g_{i1}}, ..., \mathbf{giN}]^{\mathbf{T}} \quad (16)$$

Luego se calcula nuevamente la media $\bar{h}$ y la matriz de covarianza S_g para representar las formas de variación de intensidad por:

$$\mathbf{h} = \bar{\mathbf{h}} + \mathbf{P_g b_g} \quad (17)$$

siendo P_g la matriz con los primeros M vectores propios de S_g, de forma que se tenga el 90% de varianza acumulada y b_g es el vector de los coeficientes correspondientes a cada vector propio.

2.2.1 *Algoritmo de reconocimiento de posturas*

El algoritmo genético implementado es el encargado de localizar los contornos labiales dentro de la imagen. Los cromosomas representarán al vector de coeficientes b para formar diferentes posturas labiales utilizando la ecuación (15), en la cual cada uno de los coeficientes puede tomar uno de 16 valores discretos, desde -3 desviaciones estándar hasta $+3$ desviaciones estándar con incrementos de 0.375, de tal manera que cada uno se representa con 4 bits. El cromosoma total, tendrá entonces 24 bits (Ver Figura 7), lo que genera un espacio de búsqueda de más de dieciséis millones de posibles posturas. La solución de esta tarea de búsqueda sobre espacios tan amplios, genera la necesidad de emplear los algoritmos genéticos.

Figura 7. Codificación del vector b en un cromosoma

Figure 7. *b* vector encodig onto a chromosome

El algoritmo genético va extrayendo los vectores de intensidad de la matriz de rojo de la imagen. Dado que los vectores propios P_g de la ecuación (17) son ortogonales, el vector de parámetros b_g puede ser calculado como:

$$\mathbf{b_g} = \mathbf{P_g^T}(\mathbf{h} - \bar{\mathbf{h}}) \qquad (18)$$

Teniendo este valor, se puede calcular el error entre la descripción de intensidad de la imagen y la descripción del modelo alineado, de la siguiente forma:

$$E = (\mathbf{h} - \bar{\mathbf{h}})^{\mathbf{T}}(\mathbf{h} - \bar{\mathbf{h}}) - \mathbf{b_g^T b_g} \quad (19)$$

El algoritmo genético busca minimizar este error, por lo tanto la función de evaluación será inversamente proporcional:

$$E_{val} = \frac{1}{E} \qquad (20)$$

En la sección 2.3 se especifican los parámetros del GA implementado.

Para la clasificación se tomaron los 6 coeficientes del vector b en (15), que son descriptores de la forma del modelo, y los M coeficientes del vector b_g calculado en (18).

Sobre este conjunto de 38 características, se aplicó la metodología de selección efectiva de características por medio de PCA, KPCA y MANOVA descrita en [19].

La validación del clasificador se llevó a cabo utilizando *validación cruzada*. Para este caso, se usó tres subconjutos de validación k=3, o sea, que se dividió la base de datos en grupos de 84 imágenes (28 de cada clase) y se validó 3 veces. El valor de k fue escogido de tal manera que los grupos de validación representaran el 30% de la cantidad total de imágenes, entrenando así con el 70% cada vez.

2.3 El Algoritmo Genético

En el presente trabajo se emplea un algoritmo genético de estructura simple (algoritmo canónico), el cual se corre en dos etapas. Primero, se utiliza como población de entrada la codificación de los parámetros de la ecuación 13 para segmentar la imagen, como se describe en la sección 3. Cuando el algoritmo converge, inmediatamente comienza a correr de nuevo, utilizando esta vez la codificación descrita en la sección 2.2.1 para caracterizar la postura labial.
El algoritmo se implementó con los operadores indicados por [20] para un algoritmo genético canónico:

- Selección con muestreado universal estocástico sin elitismo.
- Recombinación con cruce de un sólo punto.
- Mutación por complementación del bit.

con los parámetros planteados en [21]:

- Tamaño de la población entre 50 a 100 individuos
- Probabilidad de recombinación del 60%
- Probabilidad de mutación del 0.1%

Las condiciones de parada se establecieron experimentalmente, llegando a que el algoritmo debe detenerse bajo una de las siguientes condiciones:

- Si la población tiene más del 60% de convergencia.
- Si el algoritmo supera las 100 generaciones.

3. EXPERIMENTOS Y RESULTADOS

La base de datos con que se probó el método de segmentación y reconocimiento de posturas labiales, consta de 252 imágenes de posturas labiales extraídas de una población de personas con edades en el rango de 17 a 25 años. Las imágenes se encuentran divididas en seis clases correspondientes a las 5 vocales del idioma español, además de la boca cerrada en relajación. Las muestras fueron tomadas con resolución de imágenes 144×192,formato *bmp* de 24 bits sin compresión.

3.1 Segmentación

Los resultados concretos de segmentación de posturas labiales de las vocales */a/*, */e/*, */i/*, */o/*, */u/* y la boca cerrada se muestran en la Figura 8. Las muestras se clasificaron de la forma en que se muestra en la Tabla. 1, de acuerdo a la imagen que resulta después de aplicada la segmentación.

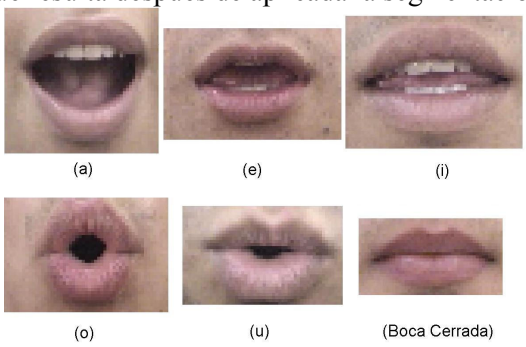


Figura 8. Imágenes segmentadas de todas las vocales y boca cerrada
Figure 8. Segmented images for all vowels and closed mouth

Los parámetros para determinar la correcta segmentación de las imágenes son una valoración subjetiva; con base en información a priori que posee el sujeto evaluador, como por ejemplo el reconocimiento de patrones de posturas labiales en multiples sujetos, posición del área de la boca en el encuadre de la escena y la plantilla de comparación entre otros. Los resultados para cada una de las categorías anteriormente mencionadas se presentan en la Tabla 2.

Tabla 1. Clasificación cualitativa del segmentador.
Table 1. Qualitative classification

1	Los labios están completos
2	Contiene aproximadamente el 90% de los labios
3	Contiene más de la mitad de los labios
4	Contiene los labios y características faciales
5	Contiene el labio inferior y características faciales
6	Contiene el labio superior y características faciales
7	Tiene otras características faciales (diferentes a los labios)

Tabla 2. Resultados del segmentador propuesto
Table 2. Results of the proposed segmentation.

Tipo	Muestras	Rendimiento (%)
1	125	73.53
2	24	12.94
3	6	3.53
4	2	1.77
5	9	5.29
6	2	0.59
7	5	2.94

El método de segmentación propuesto mostró un rendimiento de 86.47%, haciendo del segmentador por error minimizado mediante algoritmos genéticos un buen método, a pesar de la carga computacional que implica debido a su carácter iterativo, el cual converge en un rango de 50 a 100 generaciones con un tiempo de cálculo de $3.54\,s$ por generación, haciendo que el tiempo máximo de procesamiento por imagen sea $354\,s$ ($aprox. 6\,min$), que es un tiempo considerablemente alto para un procesador pentium(R) 4 de 2.4 GHz y 512Mb de Ram.

3.2 Caracterización

La figura 9 muestra, algunos de los resultados del sistema de caracterización.

Tabla 3. Porcentaje de clasificación
Table 3. Classification rates.

Aciertos (%)	Error por mala caracterización (%)	Error de clasificación con buena caracterización (%)
76.6	18.2	5.2

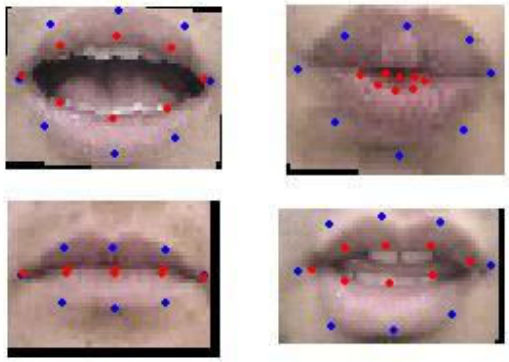


Figura 9. Resultados del sistema de caracterización
Figure 9. Feature extraction results.

El sistema de caracterización mostró un buen rendimiento, convergiendo en un promedio de 50 generaciones, lo que equivale a un tiempo de aproximadamente 34 segundos.

Clasificación

Para la clasificación se tomaron los 6 coeficientes del vector b en 15, que son descriptores de la forma del modelo, y los M coeficientes del vector b_g calculado en 18. En el experimento se obtuvo que para tener un 90% de varianza acumulada, $M = 32$.

Sobre este conjunto de 38 características, se aplicó la metodología de selección efectiva de características por medio de PCA, KPCA y MANOVA descrita en [?]. Los resultados de esta clasificación se encuentran en la Tabla 3.

Para analizar las causas de error, se construyó la matriz de confusión de clases mostrada en la Tabla 4. Las filas representan la clase verdaderas y las columnas representan la clase reconocida.

Tabla 4. Matriz de confusión en la clasificación
Table 4. Confussion matriz

	/a/	/e/	/i/	/o/	/u/	cerrada
/a/	37	2	3	0	0	0
/e/	4	31	5	2	0	0
/i/	2	4	34	2	0	0
/o/	3	3	4	25	5	2
/u/	0	2	2	6	27	5
cerrada	0	0	0	0	3	39

Puede observarse que la mayor confusión de clases se encuentra entre la /*o*/ y la /*u*/. Esto se debe a que en muchas ocasiones hay sombra alrededor del contorno interior de la boca, y esta sombra es interpretada por el sistema como parte de la apertura de la boca, caracterizándola más grande de lo que es en realidad. Otros casos en los que se presenta confusión es en las personas con dientes muy grandes, ya que el clasificador reconoce los dientes como un rasgo distintivo de la letra /*i*/; además se encuentra el caso de la pronunciación exagerada, que puede hacer que el clasificador identifique la apertura de la boca como característica de una /*a*/, cuando la vocal pronunciada es una /*o*/ o una /*e*/.

El desempeño total del sistema se encuentra resumido en la tabla 5.

Tabla 5. Error total del sistema
Table 5. Total error of the system

Aciertos totales	Error por mala segmentación	Error por mala caracterización
66.23%	13.53%	20.23%

Se observa que el porcentaje de error crece debido a que el sistema funciona serialmente, por lo tanto el error de la primera fase se acumula para la segunda.

4. CONCLUSIONES

Se ha descrito un método de segmentación de posturas labiales basado en un modelo de aproximación polinomial a la imagen promedio. Al combinar este modelo con el algoritmo genético simple, se obtuvo un sistema con un buen desempeño en la ubicación de posturas labiales, teniendo en cuenta que la plantilla de comparación contiene las características de iluminación de la base de datos y la aproximación polinomial(SVD) contiene las características de forma más generales, por lo cual el algoritmo no depende de condiciones de iluminación ni de detalles finos de las imágenes.

La utilización de algoritmos genéticos en la minimización de la función de error mostró ser una herramienta computacional de optimización que lleva a cabo la búsqueda de soluciones en paralelo, lo que los hace muy eficientes en espacios de búsqueda extensos, permitiendo obtener el mínimo de la función de error en un máximo de 100 generaciones teniendo en cuanta que habían $16.777.216$ posibilidades.

Se ha descrito un método de caracterización y clasificación de posturas labiales basado en un modelo de las variaciones de forma de los contornos labiales. Al combinar este modelo con el algoritmo genético canónico [20], se obtuvo un sistema con un buen desempeño en la caracterización y reconocimiento de posturas labiales en imágenes.

En la clasificación de posturas labiales, no es suficiente con utilizar características de forma, dado que hay posturas que son prácticamente iguales en cuanto a forma y sólo se diferencian por la posición de la lengua. Para solucionar éste inconveniente, es necesario añadir características de intensidad de la imagen.

Debido a la naturaleza serial del proceso, el error se vuelve acumulativo, lo que genera un error total bastante alto. Sin embargo, teniendo en cuenta las 2 etapas por separado, presentan buenos desempeños en sus respectivas tareas.

REFERENCIAS

[1] G. Potamianos, C. Neti, G. Iyengar, and E. Helmuth, "Large vocabulary audio-visual speech recognition by machines and human" 2001.
[2] M. Hennecke, K. Prasad, and D. Stork, "Using deformable templates to infer visual speech dynamics," 1994.
[3] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vision*, vol. 8, no. 2, pp. 99–111, 1992.
[4] R. Kaucic, B. Dalton, and A. Blake, "Real-time lip tracking for audio-visual speech recognition applications," in *ECCV (2)*, 1996, pp. 376–387.
[5] R. H. Bartels, J. C. Beatty, K. S. Booth, E. G. Bosch, and P. Jolicoeur, "Experimental comparison of splines using the shape-matching paradigm," *ACM Trans. Graph.*, vol. 12, no. 3, pp. 179–208, 1993.
[6] C. Bregler and S. M. Omohundro, "Surface learning with applications to lipreading," in *Advances in Neural Information Processing Systems*, J. D. Cowan, G. Tesauro, and J. Alspector, Eds., vol. 6. Morgan Kaufmann Publishers, Inc., 1994, pp. 43–50.
[7] M. Kass, A. Witkin, and D. Tersopuolos, "Snakes: Active contour models." *International Journal of Computer Vision*, vol. 1, no. 4, p.321–331, 1988.
[8] A. Hill and C. J. Taylor, "Model-based image interpretation using genetic algorithms," *Image and Vision Computing*, vol. 10, no. 5, pp. 295–300, 1992.
[9] D. L. Swets, B. Punch, and J. Weng, "Genetic algorithms for object recognition in a complex scene," in *ICIP*, 1995, pp. 2595–2598.
[10] Y. Cui, D. L. Swets, and J. J. Weng, "Learning-based hand sign recognition using shoslif-m," in *Proc. International conference on computer vision*, 1995.
[11] G. Bebis, S. Uthiram, and M. Georgiopoulos, "Face detection and verification using genetic search," *International Journal on Artificial Intelligence Tools*, vol. 9, no. 2, pp. 225 – 246, 2000.
[12] S. J. Louis, G. Bebis, S. Uthiram, and Y. Varol, "Genetic search for object identification," in *Evolutionary Programming VII*, V. W. Porto, N. Saravanan, D. Waagen, and A. E. Eiben, Eds. Berlin: Springer, 1998, pp. 199–208.
[13] G. Bebis, M. Georgiopoulos, N. DaVitoria Lobo, and M. Shah, "Learning affine transformation of the plane for model-based object recognition," *Proceedings of the 13th Intenational Conference on Pattern Recognition*, pp. 60–64, 1996.
[14] I. Matthews, T. Cootes, S. Cox, R. Harvey, and J. Bangham, "Lipreading using shape, shading and scale," 1998.
[15] H. Gene and F. Charles, "Matrix computations," in *John Hopkins*, 1997.
[16] J. Luettin and N. A. Thacker., "Active shape models for visual speech feature extraction," *Electronic Systems Group Report, University of Sheffield*, 1995.
[17] J. Lüttin and N. A. Thacker, "Speechreading using probabilistic methods," 1997.
[18] L. Smith, "A tutorial on principal component analysis," 2002.
[19] G. Daza and L. G. Sánchez, "Pca, kpca y manova sobre señales de voz en imágenes de posturas labiales y audio," *Trabajo de grado. Universidad Nacional de Colombia, sede Manizales*, 2004.
[20] D. Whitley, "An overview of evolutionary algorithms: practical issues and common pitfalls," *Information and Software Technology*, vol. 43, no. 14, pp. 817–831, 2001.
[21] M. Mitchell, *An Introduction to Genetic Algorithms*. Massachusetts Institute of Technology, 1996.