Rev. Fac. Ing. Univ. Antioquia N.º 47 pp. 103-116. Marzo, 2009

Segmentación de la región de la boca en imágenes faciales: Revisión bibliográfica

Mouth Segmentation in Images: A Review

Juan Bernardo Gómez[1,2], Flavio Prieto[1], Tanneguy Redarce[2]*

[1]Departamento de Ingeniería Eléctrica, Electrónica y Computación, Universidad Nacional de Colombia Sede Manizales, Carrera 27 N.º 64-60, Manizales, Caldas, Colombia

[2]Laboratoire Ampère, Cnrs, Umr 5005, 25 avenue Jean Capelle, INSA de Lyon, Antoine de Saint-Exupery, 69621 Villeurbanne Cedex, France

(Recibido el 15 de marzo de 2008. Aceptado el 6 de noviembre de 2008)

Resumen

En este artículo presentamos una revisión bibliográfica de las técnicas de segmentación de la región de la boca en imágenes faciales. Nos concentramos especialmente en los avances hechos en la última década. Para que la interpretación y comparación de las técnicas sea sencilla, éstas se presentan en forma taxonómica. Diferentes etapas del proceso de segmentación son tratadas, que abarcan desde la representación de color hasta la parametrización de la región de interés. Se realiza una comparación de algunos de los métodos revisados. Finalmente, se presenta una discusión de cada etapa de la segmentación.

------ *Palabras clave*: Boca, segmentación, imágenes faciales, labios

Abstract

This article presents a review on lip segmentation techniques, focusing in the advances of the last decade. The methods are introduced in a taxonomic manner, making it easier for interpretation and comparison. Each stage in lip segmentation process is highlighted, from the prior color representation study until the later mouth parameterization. A comparison between different methods is presented, when available. Finally, a discussion on each stage in lip segmentation is presented.

------ *Keywords*: Mouth, segmentation, facial images, lips.

* Autor de correspondencia: teléfono: + 57 + 6 + 887 94 98, correo electrónico: jbgomezm@unal.edu.co (J. Gómez).

Introducción

El permanente avance en los sistemas de cómputo, en cuanto a su capacidad de cálculo, a precios cada vez más accesibles, ha impulsado en la última década, el estudio y desarrollo de diferentes tareas como la detección automática de la región de la boca en imágenes. La segmentación y caracterización de los labios es una tarea común en aplicaciones como: la lectura audio visual del habla, lectura automática de los labios, antropometría de los labios, interacción hombre máquina a través de gestos, etc [1 - 4]. La detección de los labios es un proceso compuesto por tres grandes etapas, las cuales se encuentran casi en todo sistema de visión artificial. La primera etapa está relacionada con la selección de la región de interés dentro de la imagen completa. La siguiente etapa es la segmentación de los elementos de la imagen, en esta se deben separar los labios del resto de la imagen en la región de interés. Finalmente, una etapa que en ocasiones es omitida, es la parametrización de la región de la boca, la cual se obtiene generalmente mediante la extracción de los contornos internos y externos. La parametrización de la boca se utiliza con frecuencia como paso inicial en la detección de gestos y en el reconocimiento de rostros, por lo que el proceso es orientado a la selección de puntos de referencia y a la extracción de características. Esta revisión está organizada siguiendo las siguientes etapas: Después de la introducción, discutimos como algunas representaciones de color han sido ajustadas para la segmentación de los labios. Luego se presentan las tendencias en segmentación automática de los labios y un resumen de las diferentes técnicas empleadas en el modelado paramétrico de la región de la boca. Posteriormente, se presentan las técnicas más comunes para medir la calidad de la segmentación y parametrización de los labios. Finalmente, se presentan las conclusiones.

Realce de la boca mediante transformaciones del color

En muchas aplicaciones en visión artificial es posible encontrar transformaciones de color lineal, las cuales permiten resaltar las regiones de interés del resto de la imagen. En tales aplicaciones, una vez que la imagen ha sido resaltada o transformada, una simple umbralización es suficiente para realizar la segmentación. Por ejemplo, en [5] se presenta una interesante revisión en la representación en los espacios de color para detección de la piel.

Es conocido que el problema de clasificar la piel y el color de los labios no es un problema linealmente separable [6 - 8]. Sin embargo, aun cuando el Tono y sus variantes, han mostrado que realzan la región de los labios, ellas no están desacopladas de la luminancia. Para manejar estas limitaciones, se han propuesto dos técnicas diferentes. La primera, que puede ser llamada basada en píxel, se fundamenta en la realización de una separación no lineal entre las clases. En este caso, cada color representa una variable en el espacio de características. La segunda, llamada basada en imagen, depende de los parámetros intrínsecos que pueden ser ajustados para cada imagen diferente. Algunas transformaciones basadas en píxel son presentadas en [9 - 11]. Combinaciones de transformaciones basadas en píxel y basadas en imagen se pueden encontrar en [3, 12, 13].

Transformaciones basadas en píxel

La región de los labios es muy similar, en cuanto al color, al resto de la piel. Por esta razón diferentes transformaciones de color han sido desarrolladas. Como ejemplo, la transformación de semitono, propuesta en el trabajo de Hurlbert y Poggio [14], exhibe las diferencias entre labios y piel bajo condiciones de iluminación controladas.

Una versión normalizada de la transformación de semitono puede ser encontrada en [12]. El semitono lleva a un resultado muy similar al que se logra utilizando la transformada de tono. Sin embargo, el semitono se concentra en la relación entre la información de rojo y verde de cada píxel. Un ejemplo en imágenes faciales, representado en semitono, puede verse en la Figura 1(b). Esta transformación es usada junto con el canal de luminancia, para obtener la transformación conocida como *curve map* [12, 15] (ver la Figura 1(c)).

Algunas transformaciones lineales del espacio de color RGB han permitido lograr buenos resultados, en términos de separabilidad de color entre labios y piel. Guan [11], y Morán y Pinto [16] hacen uso de la transformada discreta de Hartley (DHT), para mejorar la representación de color en la región de los labios. La componente C_3 de la transformación DHT resalta correctamente el área de los labios en sujetos con piel clara y sin barba, como se muestra la Figura 1(d).

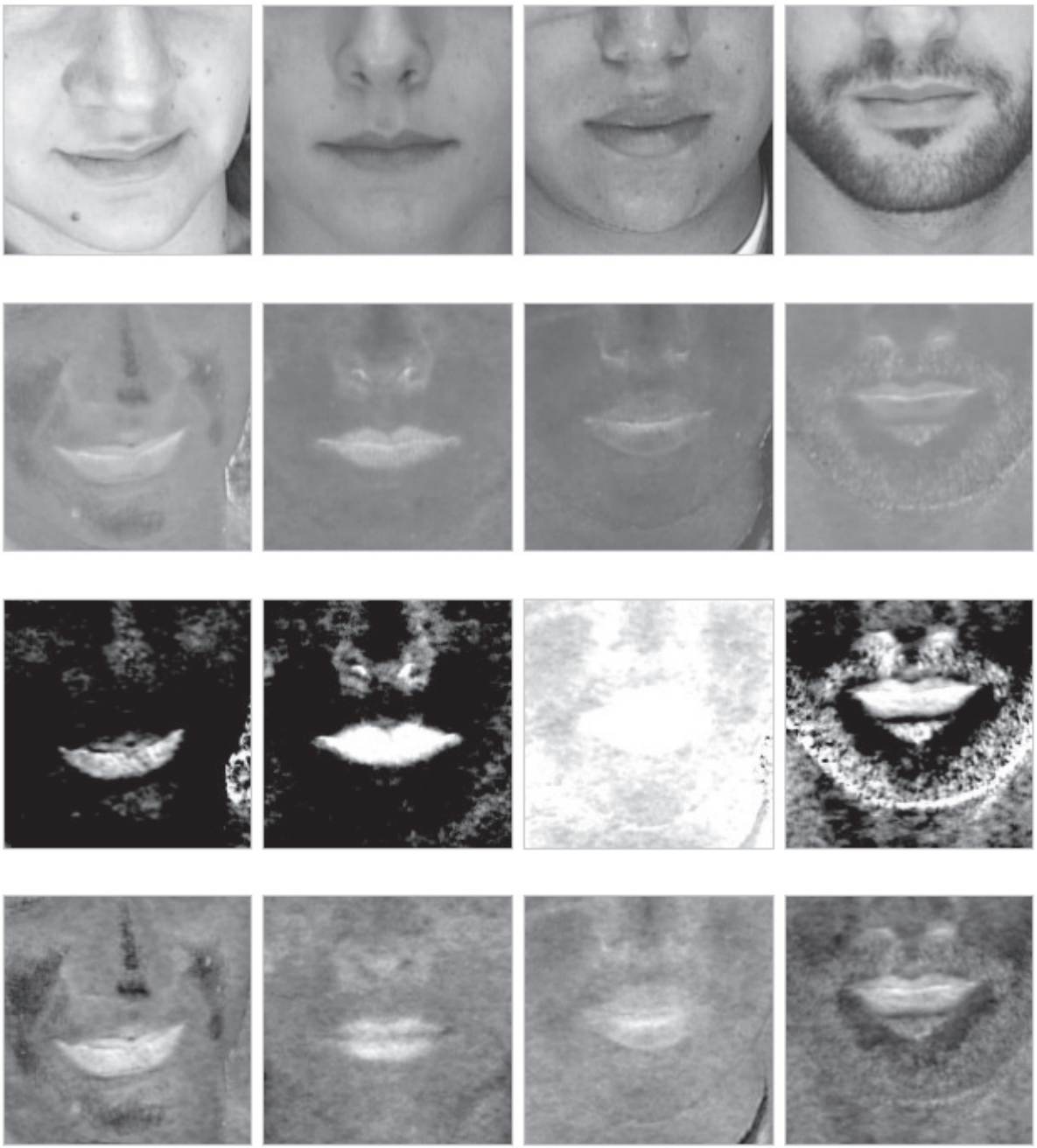

Figura 1 Efectos de las trasformaciones de color en imágenes

El uso de las transformaciones perceptuales no lineales basadas en píxel, las cuales presentan mejor constancia de color sobre variaciones de intensidad pequeñas, ha sido una de las grandes tendencias desde finales de los años 90. Dos transformaciones perceptuales bien conocidas, presentadas por la comisión internacional de iluminación (*Comission Internationale de l'Eclairage*, CIE), son el CIELAB y el CIELUV. El principio detrás de estas transformaciones es la compensación del comportamiento logarítmico del sensor. Trabajos como los presentados en [2, 17] hacen uso de estas representaciones de color, con el objeto de facilitar el proceso de segmentación de los labios. Salazar *et. al.* [18] utilizan YCbCr e información de la transformación de tono para esta segmentación.

Algunas veces, hay condiciones no controladas en el proceso de adquisición que producen cambios inesperados en la imagen de color o luminancia. Para sobrellevar este problema, un conjunto de transformaciones de color implícitas ha sido desarrollado [19]. Una transformación remarcable, orientada a la segmentación del labio, es presentada en el trabajo de Hsu *et. al.* [20]. La transformación, llamada *Mouth Map*, ajusta la compensación global de color en función de los valores de color en la imagen completa. Con esto se evita tener que calcular un umbral diferente para cada imagen, pero el proceso se hace más sensible frente a artefactos (tales como presencia/ausencia de barba, dientes, etc.). En [12], una forma normalizada de semitono es presentada, esta considera los valores máximo y mínimo del semitono, con el objeto de compensar disparidades debidas a efectos de iluminación. En [3], la componente de color verde es normalizada contra los valores máximo y mínimo de intensidad, para mejorar la estabilidad del umbral. Sin embargo, la transformación es sensible a la presencia de dientes.

Segmentación basada en operaciones de píxel y de región

Las técnicas basadas en píxel son las más simples y en general, la alternativa más rápida para realizar la segmentación de la imagen. Ellas utilizan comparaciones lógicas entre un conjunto de umbrales y los valores de color en los píxeles de la imagen. El valor de la comparación lógica define si el píxel pertenece o no a una región específica en la imagen. Ejemplos de segmentación basada en píxel se encuentran en [3, 21]. En Gómez *et. al.* [3], una mezcla de tres diferentes espacios de color es usada como paso precedente a un proceso de umbralización. La mezcla de los espacios de color (componente verde del espacio RGB, la componente de tono y el *Mouth Map* [20]), hace que el algoritmo sea más selectivo, llevando a un decremento de las regiones espurias en la segmentación. Los autores realizan un recorte de la región de interés (ROI), con el objeto de descartar la región de los orificios de la nariz. Un ejemplo de una imagen segmentada usando esta técnica se observa en la Figura 2.

La principal desventaja de las técnicas basada en píxel es la ausencia de restricciones de conectividad en el método. Por esto, trabajos como [22] consideran la conectividad en la umbralización. No obstante, las técnicas basadas en color requieren, generalmente, una etapa de postproceso para la eliminación de regiones espurias en la segmentación final [22].

Técnicas basadas en píxel son muy sensibles a cambios pequeños en el color, debidos esencialmente a cambios de la iluminación. Para tratar con esto, se han presentado algunos algoritmos para la selección automática del umbral. Lucey *et. al.* [22] propusieron un algoritmo de segmentación basado en una técnica de umbralización dinámica. El primer paso en este método es presentar la imagen en una versión restringida de la relación R/G. Posteriormente, una función de entropía, que mide la incertidumbre entre clases (fondo y labios), es minimizada con respecto a los parámetros de la función de membresía. Zhang *et. al.* [23] utilizan un análisis discriminante lineal (LDA) de Fisher, para encontrar la transformación lineal que maximiza la diferencia entre el color de la piel y los labios. Seguidamente, desarrollan una selección automática del umbral, de acuerdo a la transformación de color encontrada

en una etapa previa. Rongben *et. al.* [24] también proponen una técnica de LDA de Fisher, orientado a un proceso de selección automática del umbral. Kim *et. al.* [25] sugieren el uso de datos con marcas manuales con el objeto de entrenar un sistema de inferencia difuso, el cual es usado como un índice de confianza para la selección automática del umbral. Máquinas de vectores de soporte (SVM), también han sido utilizadas para modelar la diferencia entre los labios y la piel [26].

Figura 2 Ejemplo de segmentación de la boca presentada en [3]

Modelado estadístico de los labios

Otro método para segmentar la región de la boca es mediante el modelado de la región y/o el color de los labios, y la distinción entre ese modelo y el fondo. Este proceso de modelado puede ser determinista (como el modelado de contorno en la Sección 3), o estocástico, revisado en esta sección. Una técnica común de modelado estocástico para segmentación de imágenes corresponde al modelado por campos aleatorios de Markov (Markov Random Field, MRF). En MRF, la imagen es presentada como una realización de una variable estocástica en la cual cada elemento (píxel) es descrito en términos de relaciones de vecindario. Formalmente, se dice que un objeto aleatorio X en una red S con vecindario ∂s es un MRF si para todo s$\in S$ se cumple la Ecuación 1,

$$p(x_s \mid x_r \text{ for all } r \neq s) = p(x_s \mid x_{\partial r}) \quad (1)$$

Se ha mostrado que, si $P\{X=x\}>0$ para todo x –esto es generalmente asumido–, entonces $P\{X\}$ tiene la forma de una distribución de Gibbs. Una distribución común usada en segmentación de imágenes es la MRF no gaussiana, basada en un par de intervalos y dada por la formulación presentada en la Ecuación 2:

$$p(x) = \frac{1}{Z}\exp\left(\sum_{\{s,r\}\in C} b_{sr}\,\rho\left(\frac{x_s - x_r}{\sigma}\right)\right) \quad (2)$$

Donde Z es la constante de normalización de la densidad, C es el conjunto de intervalos, σ es el término de escala de nivel de gris, y ρ es la función potencial. Información adicional sobre MRF se puede encontrar en [27, 28].

Liévin y Luthon [6] hacen uso de MRF para selección de etiquetas en el proceso de segmentación. En este caso, la MRF es usada para el etiquetado

correcto, dadas las características dinámicas y estáticas de cada píxel. La característica estática corresponde a la etiqueta binaria obtenida a partir del valor del tono rojo (semitono), mientras que la característica dinámica es calculada de las etiquetas obtenidas a partir de la diferencia (sin signo) entre dos imágenes consecutivas en una secuencia de video. La etiqueta en la siguiente iteración depende sólo de la etiqueta del píxel en la iteración precedente y de las etiquetas de sus *n* vecinos. La región de interés (ROI) debe ser recortada antes de la ejecución del algoritmo, con el propósito de reducir el efecto de regiones espurias en la segmentación. Zhang y Mersereau [10] usan una mezcla de tono y el valor en el espacio de color de semitono. Las etiquetas obtenidas en cada imagen son usadas en un MRF, para mejorar la detección del contorno interno y externo de los labios. Algunas restricciones son impuestas en las funciones de energía MRF, con el propósito de restringir el etiquetado final del contorno. Los autores reportaron que, aun en los casos donde las fronteras de los labios no son suaves, es posible calcular con precisión los puntos de referencia sobre la segmentación.

Mezclas de modelos también se han usado para modelar algunas características específicas en las imágenes. En el trabajo de Sadeghi *et. al.* [29], un modelo de mezclas gausianas (Gaussian Mixture Model, GMM) es usado para modelar el color en la región del labio. Un muestreo Sobol es usado para reducir la carga computacional del algoritmo. En Gacon *et. al.* [30], se propone un método para la caracterización dinámica del color del labio y para su segmentación. El método es basado en modelos gaussianos estadísticos del color de la cara. Los modelos son entrenados con imágenes con marcas manuales. El método es capaz de modelar características dinámicas y estáticas en el color del píxel y en la forma de la boca, permitiendo que el algoritmo compense cambios de iluminación y movimientos rápidos de la boca. Los autores también proponen un método para modelado estático de la comisura de los labios. La segmentación global es desarrollada optimizando la posición de un modelo cuadriculado de los labios. Como el modelo tiene varios cientos de puntos, la optimización es un problema de dimensión alta.

En Goswami *et. al.* [31], se presenta un método de segmentación automática del labio basado en estimadores estadísticos. Un estimador del determinante de covarianza mínimo (Minimum Covariance Determinant Estimator MCD), y un estimador no robusto son utilizados para estimar la región de la piel en el espacio de color. La región de los labios es hallada como la mayor región conectada que no sea piel. Los autores reportan una mejora significativa sobre los resultados presentados en [15]. El método asume que la región de la piel puede ser detectada con mayor facilidad que la región de los labios. Otra técnica estadística para la separación del color es presentada en [32]. En Bouvier *et. al.* [33], se presenta un algoritmo para la extracción del contorno externo de los labios. Los autores se centran en la detección de la ROI antes de la parametrización del contorno de los labios. Ellos realizan una estimación del área de los labios usando máxima esperanza. Posteriormente, se halla un mapa de membresía del labio a partir de la distribución de color de la piel. Realizan una optimización del umbral, basada en el gradiente y en información de la máscara. La extracción del contorno de los labios se realiza aproximando un conjunto de curvas de Bézier en un mapa multiescala del contorno de la boca.

Segmentación C-Media difuso (FCM)

Es una técnica común de agrupamiento usada en segmentación de imágenes, introducida en clasificación de patrones a finales de los años 60 [34]. Las bases del FCM son presentadas en el texto de Bezdek [35, 36].

La segmentación FCM está basada en el principio de disimilaridad de características: dado $\mathbf{X}=\{\mathbf{x}_{1,1},\mathbf{x}_{1,2},...,\mathbf{x}_{N,M}\}$ un conjunto de características que corresponden a la imagen I de tamaño $N\times M$, cada $\mathbf{x}_{r,s}\in \mathrm{R}^q$ es el vector de características del píxel correspondiente en I; y C el número de clústeres difusos en la imagen, la meta es encontrar un conjunto $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C\}$ de C dife-

rentes centroides $\mathbf{v}_i \in R^q$, y una matriz **U** con tamaño $N \times M$ la cual es una partición difusa de **X**, tal que minimice la función de costo $J(\mathbf{U}, \mathbf{V})$ en la Ecuación 3, con las restricciones impuestas en la Ecuación 4.

$$J(\mathbf{U},\mathbf{V}) = \sum_{r=1}^{N}\sum_{s=1}^{M}\sum_{i=1}^{C} u_{i,r,s} D_{i,r,s} \qquad (3)$$

Sujeto a

$$\sum_{i=1}^{C} u_{i,r,s} = 1, \text{ para todo } (r,s) \in I \qquad (4)$$

$D_{i,r,s}$ es un término que refleja la distancia entre el vector de características $\mathbf{x}_{r,s}$ y el centroide $\mathbf{v}_i$. El conjunto de características **X** está compuesto generalmente por diferentes representaciones de color, como las descritas en la Sección 1. Generalmente, la partición difusa **U** es comparada con un umbral, con el objeto de obtener un conjunto clásico de regiones disjuntas en la imagen fuente.

Con el objeto de extender la evolución FCM, no sólo en el espacio de características sino también en el de forma, algunas modificaciones han sido aplicadas a la función básica de costo, presentada en la Ecuación 3. En el trabajo de Wang *et. al.* [17, 37] un nuevo conjunto de parámetros fue introducido en el término de distancia $D_{i,r,s}$. Ellos propusieron el uso de función de base espacial con forma elíptica, buscando mejorar la segmentación de los labios. Restricciones en la forma, ayudan a reducir el efecto de algunos píxeles espurios fuera de la región de la boca, los cuales producen un efecto indeseado en la función básica de costo. Liew *et. al.* [38] también usan una técnica basada en FCM en el espacio de características de color con constricciones en la geometría.

Artefactos en las imágenes (como la barba), también han sido tratados con modificaciones de FCM. Wang *et. al.* [8, 39], proponen el uso de FCM con el objeto de modelar la región de los labios en un clúster, mientras que el fondo es modelado con muchos clústeres. Las funciones objetivos del FCM son modificadas para incluir en la optimización una constricción en la forma paramétrica, como en [37]. Su esquema de segmentación, también llamado MS-FCS (multi-class, shape-guided FCM), muestra una gran reducción del error de segmentación en presencia de barba, comparado con el FCM tradicional y con el trabajo de Zhang y Mersereau [10].

Otras técnicas de segmentación

La inteligencia artificial también ha sido utilizada en segmentación de los labios. Mitsukura *et. al.* [40, 41] usan dos Redes Neuronales (RN) de propagación hacia adelante previamente entrenadas para modelar el color de la piel y los labios. Restricciones en la forma son incluidas en los pesos de la RN para detección de los labios. Una vez que los candidatos a boca son detectados, una prueba de piel es desarrollada en su vecindario, usando la RN para detección de piel. Posteriormente, se usa una RN para detección de los labios y así seleccionar la región de la boca. En otro trabajo, los autores presentan un segundo esquema [42], basado en algoritmos evolutivos para el modelado de los labios. Para resaltar el área de la boca Lie *et. al.* [43] usan un conjunto de operaciones morfológicas en imágenes con diferencias temporales.

Segmentación de labios por ajuste a plantillas parametrizables

El modelado de contorno se desarrolla encontrando un conjunto de puntos que controlan el modelo, y el conjunto de funciones de modelado. Una técnica básica que muestra el proceso completo se encuentra en Rao y Mersereau [44]. En este trabajo, los autores usan operadores lineales para encontrar el contorno horizontal de los labios y posteriormente aproximan el contorno con dos parábolas. Este método no es adecuado para aplicaciones antropométricas, ya que una parábola no es suficiente, en general, para modelar el labio superior o inferior [4].

Modelo de forma activa (ASM) y Modelo de apariencia activa (AAM)

Estos dos métodos estadísticos para ajuste de plantillas, surgieron a mediados de los 90. ASM

son modelos estadísticos de la forma del objeto que se deforma iterativamente, para ajustar un objeto determinado en una nueva imagen [45]. Las formas están restringidas por un modelo estadístico de forma, para variar sólo en las formas precisadas en el conjunto de entrenamiento, compuesto por ejemplos etiquetados. Los puntos de referencia están localizados, generalmente, en los contornos. De otra parte, AMM es una generalización de la técnica ASM, pero usa toda la información cubierta por el objeto destino en la región de la imagen y no sólo los bordes [46].

En el trabajo de Caplier [47], se presenta un método para la detección y el seguimiento automático de los labios. El método hace uso de una inicialización automática de puntos de referencia, previamente presentada en [48]. Una vez realizado el proceso de inicialización, las varianzas resultantes de un proceso de ajuste de plantillas son usadas para extraer un conjunto reducido de características. Finalmente, las características son utilizadas para seleccionar y adaptar un ASM que describe el gesto de la boca. El filtrado de Kalman es usado para acelerar la convergencia del algoritmo. Turkmani y Hilton [49] usan AAM en secuencias de habla, para localizar contornos internos y externos de la boca. En Jiang *et. al.* [50], se usa una mezcla de un modelo determinista de filtrado de partículas y un modelo estocástico ASM, con el propósito de mejorar la convergencia y la precisión en el seguimiento de los labios. En [51] se propone un filtrado de partículas guiado por atractores.

Contornos activos

Contornos activos o *snakes*, son curvas generadas por computador que se mueven dentro de las imágenes para encontrar fronteras de objetos, en este caso, el contorno interno o externo de la boca. Un contorno activo puede ser definido como una curva $\mathbf{v}(u,t) = (x(u,t), y(u,t))$, $u \in [0,1]$, con t la posición temporal del punto en la secuencia que se mueve en el espacio de la imagen [52]. La evolución de la curva es controlada por la función de energía presentada en la Ecuación 5.

$$E_{ac} = \int E_{int}(v(u)) + E_{im}(v(u)) + E_{ext}(v(u))du \quad (5)$$

E_{int} representa la energía interna de la curva, y controla las propiedades de estiramiento y doblado de la curva. E_{im} es la imagen energía y está relacionada con las propiedades de la imagen. E_{ext} es una energía externa que generalmente representa constricciones específicas en la evolución de la curva. Un ejemplo de la extracción del contorno externo de los labios, mediante el uso de contornos activos se observa en la Figura 3.

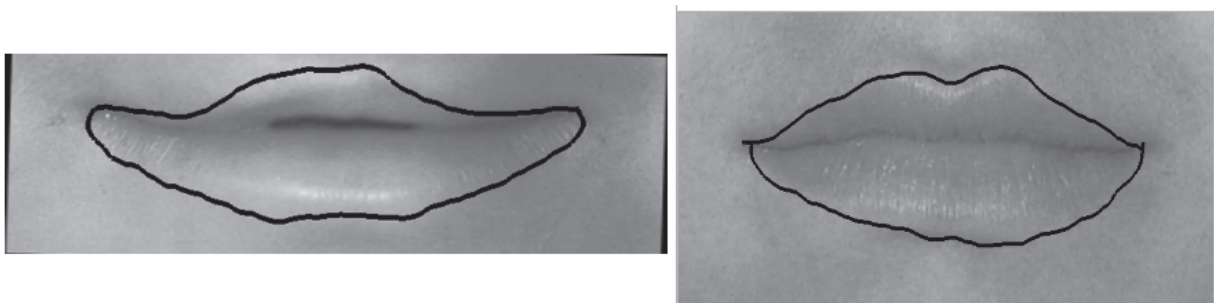

Figura 3 Parametrización del contorno de los labios mediante contornos activos [18], [52]

Una técnica llamada flujo del vector gradiente (Gradient Vector Flow, GVF), ha sido utilizada para mejorar la convergencia y precisión en la representación de fronteras con alta curvatura. Desde la introducción del concepto de GVF en contornos activos [53, 54], han surgido diferentes técnicas de segmentación parametrizada que utilizan este concepto. Morán y Pinto [16] usan GVF para restringir la aproximación de un contorno paramétrico, limitado por un conjunto de

puntos de referencia, conformando un modelo activo de forma. La caja, región frontera de la boca, es encontrada cortando los ejes horizontales y verticales en la proyección perpendicular de cada eje. El GVF es calculado en el espacio de color C_3+U, donde U representa la componente u en el espacio de color CIELUV. Otra aproximación que usa GVF es la presentada en [55]. En este caso, el detector de rostro propuesto por Viola-Jones [56], es usado para detectar la caja envolvente del rostro y la caja envolvente de la boca. Después de esto, una formulación de contornos activos que utiliza un método de selección de nivel sin reinicialización es implementado, con el propósito de ajustar el modelo. Hernández *et. al.* [52] presentan una forma simplificada de GVF para contornos activos y la aplican a la segmentación de la boca. Una parametrización simplificada del contorno externo, el cual utiliza polinomios de cuarto orden después de la convergencia del contorno activo, se puede encontrar en [4]. El contorno externo de la boca puede ser descrito con precisión utilizando esta técnica, pero es altamente dependiente de la segmentación precedente.

En Eveno *et. al.* [15] desarrollaron una manera simple de representación del contorno de los labios, mediante la búsqueda de un conjunto de puntos clave en las proyecciones de intensidad horizontal y vertical de la región de la boca, y luego aproximan un conjunto de polinomios a los puntos encontrados. La búsqueda de los puntos y el ajuste fino es controlada por un gradiente especial de la imagen llamado *hybrid edges*, basado en la luminancia y el semitono. Su trabajo evolucionó a una nueva técnica llamada *jumping snake* [57, 58]. Este método permite una detección del contorno de los labios, con sólo la indicación de un punto arbitrario sobre la región del labio en la imagen. En cada iteración, un nuevo par de nodos es adicionado a las esquinas de cada modelo. Seyedarabi *et. al.* [59] usan una técnica de contornos activos de dos pasos para aproximar el contorno externo de los labios. Primero, un operador de Canny es utilizado, luego, mediante un umbral alto se extrae el contorno externo del labio superior. Después de la convergencia, un segundo umbral de valor bajo es usado para ajustar un contorno activo, el cual sólo se detiene al hallar el contorno inferior del mismo labio. Beaumesnil y Luthon [1] presentaron una técnica de contorno activo en tiempo real para la segmentación de la boca, en el cual el modelo 3D del rostro es ajustado directamente a la imagen. La precisión no fue una restricción en ese trabajo, ya que la meta es sintetizar la expresión del rostro en un modelo generado por computador.

Otras técnicas paramétricas

En Werda *et. al.* [60], se propone un algoritmo que combina segmentación paramétrica y segmentación basada en píxel. Un conjunto de colores y operaciones morfológicas es aplicado para obtener una segmentación inicial de los labios. Posteriormente, el modelo paramétrico es pegado y el contorno externo de los labios encontrado. La representación final contiene un modelo geométrico de parametrización fuerte, cuyos parámetros permiten que el contorno sea deformado en un conjunto restringido de formas posibles. Otro ejemplo de parametrización fuerte del contorno de los labios se encuentra en [61]. Moghaddam y Safabakhsh [62] presentan un algoritmo rápido para la extracción del contorno externo que utiliza mapas auto-organizados. Xie *et. al.* [63] proponen una segmentación de los labios que combina ASM y proyecciones acumulativas, para mejorar la robustez en la detección del contorno.

Medidas de desempeño

La mayoría de los algoritmos de parametrización son creados para suministrar información a procesos posteriores, como reconocimiento audiovisual del habla [2]. En este sentido, en general, las medidas de calidad están orientadas a mostrar el desempeño específico de la aplicación, más que de la segmentación. Sin embargo, algunas medidas han sido realizadas para medir el desempeño de la segmentación. En la Tabla 1 se muestra un resumen de algunas técnicas utilizadas para la medición del error en la segmentación y la detección de labios en imágenes.

Tabla 1 Medidas de desempeño de la segmentación de los labios

Referencia	*Forma de Medida*	*Datos de Prueba*	*Comparado con*
Wang *et. al.* (FCMS) [17]	• Error de labio interno (ILE). • Error de labio externo (OLE).	Resultados reportados para dos imágenes.	FCM.
Liew *et. al.* [38]	• Error de traslape $OL = \frac{2\lvert A_1 \cap A_2 \rvert}{\lvert A_1 + A_2 \rvert} \times 100\%$ • Error de segmentación $SE = \frac{OLE + ILE}{2TL} \times 100\%$	70 Imágenes tomadas de las bases de datos XM2VTS [64] y AR [65].	- o -
Leung *et. al.* (FCMS) [37]	• Igual que [17]. • $SE=P(O)P(B\|O)+P(B)P(O\|B)$	Resultados reportados para tres imágenes y para 27.	CT [12], FCM, LL [6], ZM [10]
Wang *et. al.* (MS-FCM) [8]	Igual que [37].	Resultados reportados para tres imágenes.	FCM, LL [6], ZM [10]
Xie *et. al.* (RoHiLTA) [63]	Error promedio de ubicación del modelo anotado contra el modelo detectado $M_{\mathrm{e}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \left\| f_i - f_i^{\mathrm{h}} \right\|$	150 imágenes tomadas de la base de datos AVCONDIG; 50 con barba o sombras.	- o -
Eveno *et. al.* (CT) [12]	Error de segmentación $\varepsilon = \frac{S_{\mathrm{in}} + L_{\mathrm{out}}}{L_{\mathrm{total}}}$; equivalente a dos veces *SE*, como es planteado en [38].	Resultados reportados de tres imágenes.	- o -
Eveno *et. al.* [57]	• Error de rastreo medio normalizado (error medio de localización) $\varepsilon_{i,\mathrm{tracking}} = \frac{1}{\mathrm{T}} \sum_{n=1}^{\mathrm{T}} \frac{\lvert \mathbf{p}_{i,\mathrm{ref}}(n) - \mathbf{p}_{i,\mathrm{tracking}}(n) \rvert}{\lvert \mathbf{p}_{1,\mathrm{ref}}(n) - \mathbf{p}_{5,\mathrm{ref}}(n) \rvert}$ • Error humano medio normalizado $\varepsilon_{i,\mathrm{human}} = \frac{1}{\mathrm{T}} \sum_{n=1}^{\mathrm{T}} \left(\frac{1}{\mathrm{K}} \sum_{k=1}^{\mathrm{K}} \frac{\lvert \mathbf{p}_{i,\mathrm{ref}}(n) - \mathbf{p}_{i,\mathrm{human}}(n,k) \rvert}{\lvert \mathbf{p}_{1,\mathrm{ref}}(n) - \mathbf{p}_{5,\mathrm{ref}}(n) \rvert} \right)$	Resultados reportados sobre 300 imágenes, de 11 secuencias distintas, y etiquetadas por diferentes personas.	- o -

Referencia	*Forma de Medida*	*Datos de Prueba*	*Comparado con*
Hammal *et. al.* [58]	Igual que [57].	Resultados reportados sobre 300 imágenes, de 8 secuencias distintas, y etiquetadas por diferentes personas.	- o -
Bouvier *et. al.* [33]	Igual que [57].	Resultados reportados de 450 imágenes, de 12 secuencias distintas.	Eveno [57] & Gacon [30]
Gacon *et. al.* [30]	Igual que [57].	No especificado.	- o -
Guan [11]	*SE* igual que en [38].	Resultados reportados para cuatro y 35 imágenes.	FCM, CGC [66]
Khan *et. al.* [26]	Medida de calidad de segmentación $q(E,G)=\frac{\lvert E\cap G\rvert}{\lvert E\cup G\rvert}$	Resultados reportados para 122 imágenes.	- o -

Conclusiones

Existen bastantes trabajos en modelado del color de la piel y los labios. Sin embargo, se ha observado que la separación entre estas regiones es altamente dependiente de las condiciones de iluminación y, por lo tanto, difícil de predecir. Algunas transformaciones de color muestran un buen realce para labios y/o piel de color específico. Sin embargo, ellos tienen problemas cuando los sujetos poseen características de color diferentes. Las técnicas revisadas, aunque tratan de predecir estos cambios de color inesperados, fallan con pieles oscuras o en presencia de ruido especular o de los dientes en a imagen.

Las técnicas de segmentación que utilizan restricciones geométricas en general están diseñadas para condiciones específicas de la imagen, y aunque permiten rechazar regiones espurias, las restricciones hacen que los métodos no sean aplicables cuando no se tiene información a priori, sobre el tamaño y la forma de la boca en la imagen. Algunos algoritmos han mostrado un buen desempeño en imágenes con presencia de barbas, pero fallan cuando se cambia la iluminación. Pocos trabajos han sido implementados para funcionar en tiempo real.

Un gran número de los métodos de parametrización son muy sensitivos al algoritmo de inicialización. Técnicas como la presentada en [4] son muy precisas en el modelado del contorno externo y el modelo obtenido es fácilmente interpretable. Sin embargo, ellas requieren una buena segmentación inicial de la boca y una parametrización gruesa del contorno del labio. Algoritmos paramétricos como AAM y ASM, también son sensibles al proceso de inicialización. Cuando la inicialización está alejada del objeto destino, ellos pueden converger hacia mínimos locales.

No existe unanimidad, por parte de los autores, en la selección de los datos de entrenamiento y prueba. En general, ellos son seleccionados para aplicaciones específicas. Tampoco existe una estan-

darización en la forma de reportar los resultados. Estas diferencias dificultan los procesos de evaluación de las técnicas de segmentación de los labios.

Referencias

1. B. Beaumesnil, F. Luthon. "Real time tracking for 3D realistic lip animation". *Proceedings of the 18th International Conference on Pattern Recognition.* ICPR 2006. Vol. 1. 2006. pp. 219-222.
2. I. Arsic, R. Vilagut, J. P. Thiran. "Automatic extraction of geometric lip features with application to multimodal speaker identification". *2006 IEEE International Conference on Multimedia and Expo*. 2006. pp. 161-164.
3. J. B. Gómez, J. E. Hernández, F. Prieto, T. Redarce. "Real-time robot manipulation using mouth gestures in facial video sequences". *Lecture Notes in Computer Science*. Vol. 4729. 2007. pp. 224-233.
4. A. E. Salazar, J. E Hernández, F. Prieto. "Automatic quantitative mouth shape analysis". *Lecture Notes in Computer Science*. Vol. 4673. 2007. pp. 416-423.
5. V. Vezhnevets, V. Sazonov, A. Andreeva. "A survey on pixel-based skin color detection techniques". *Proceedings of GraphiCon 2003*. pp. 8. Disponible On Line: http://citeseer.ist.psu.edu/676368.html. Consultada el 1 de marzo de 2008.
6. M. Liévin, F. Luthon. "Unsupervised lip segmentation under natural conditions". *IEEE International Conference on Acoustics, Speech, and Signal Processing*. ICASSP'99: Vol. 6. 1999. pp. 3065-3068.
7. Z. Jian, M. N. Kaynak, A. D. Cheok, K. C. Chung. "Real-time lip tracking for virtual lip implementation in virtualenvironments and computer games". *The 10th IEEE International Conference on Fuzzy Systems.* Vol. 3. 2001. pp. 1359-1362.
8. S. L. Wang, W. H. Lau, A.W.C. Liew, S. H. Leung. "Robust lip region segmentation for lip images with complex background". *Pattern Recognition*. Vol. 40. 2007. pp. 3481-3491.
9. G. I. Chiou, J. N. Hwang. "Lipreading from color video". *IEEE Transactions on Image Processing*. Vol. 6. 1997. pp. 1192-1195.
10. X. Zhang, R. M. Mersereau. "Lip feature extraction towards an automatic speechreading system". *Proceedings of IEEE International Conference on Image Processing*. Vol. 3. 2000. pp. 226-229.
11. Y. P. Guan. "Automatic extraction of lip based on wavelet edge detection". *Eighth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC '06.* 2006. pp. 125-132.
12. N. Eveno, A. Caplier, P. Y. Coulon. "New color transformation for lips segmentation". *IEEE Fourth Workshop on Multimedia Signal Processing*. 2001. pp. 3 - 8.
13. J. Loaiza, J. B. Gómez, A. Ceballos. "Análisis de discriminancia y selección de características de color en im´agenes de labios utilizando redes neuronales". *Memorias del XII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial STSIVA07*. 2007. pp. 4.
14. A. C. Hurlbert, T. A. Poggio. "Synthesizing a color algorithm from examples". *Science*. -1988. Vol. 239. Pp. 447-514.
15. N. Eveno, A. Caplier, P. Y. Coulon. "A parametric model for realistic lip segmentation". *Seventh International Conference on Control, Automation, Robotics And Vision (ICARCV'O2)*. 2002. pp. 1426-1431.
16. L. E. Morán, R. Pinto. "Automatic extraction of the lips shape via statistical lips modelling and chromatic feature". *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*. 2007. pp. 241-246.
17. S. L. Wang, S. H. Leung, W. H. Lau. "Lip segmentation by fuzzy clustering incorporating with shape function". *IEEE International Conference on Acoustics, Speech, and Signal Processing*. ICASSP'02. 2002. Vol. 1. pp. 1077-1080.
18. A. Salazar, F. Prieto. "Extracción y clasificación de posturas labiales en niños entre 5 y 10 años de la ciudad de Manizales". *DYNA*. 2006. Vol. 73. pp. 175-188.
19. R. Collins, Y. Liu, M. Leordeanu. "On-line selection of discriminative tracking features". *IEEE Transaction on Pattern Analysis and Machine Intelligence*. Vol. 27. 2005. pp. 1631-1643.
20. R. L. Hsu, M. Abdel-Mottaleb, A. K. Jain. "Face detection in color images*". IEEE Trans. on Pattern Analysis and Machine Intelligence.* Vol. 24. 2002. pp. 696-706.
21. J. A. Dargham, A. Chekima. "Lips detection in the normalised RGB colour scheme". *Proceedings of 2nd Information and Communication Technologies*. ICTTA. Vol. 1. 2006. pp. 1546-1551.

22. S. Lucey, S. Sridharan, V. Chandran. "Chromatic lip tracking using a connectivity based fuzzy thresholding technique". *Proceedings of the Fifth International Symposium on Signal Processing and its Applications*. ISSPA '99. 1999. pp. 669-672.

23. J. M. Zhang, D. J. Wang, L. M. Niu, Y. Z. Zhan. "Research and implementation of real time approach to lip detection in video sequences". *Proceedings of the Second International Conference on Machine Learning and Cybernetics.* 2003. pp. 2795-2799.

24. W. Rongben, G. Lie, T. Bingliang, J. Lisheng. "Monitoring mouth movement for driver fatigue or distraction with one camera". *Proceedings of he 7th International IEEE Conference on Intelligent Transportation Systems.* 2004. pp. 314-319.

25. J. Y. Kim, S.Y. Na, R. Cole. "Lip detection using confidence-based adaptive thresholding". *Lecture Notes in Computer Science*. Vol. 4291. 2006. pp. 731-740.

26. A. Khan, W. Christmas, J. Kittler. "Lip contour segmentation using kernel methods and level sets". *Lecture Notes in Computer Science*. Vol. 4842. 2007. pp. 86-95.

27. H. Bunke, T. Caelli. "Hidden Markov Models: Applications in Computer Vision". *World Scientific Series In Machine Perception And Artificial Intelligence Series*. Vol. 45. World Scientific Publishing Co. 2001. pp. 244.

28. R. Chellappa, A. K. Jain. *Markov Random Fields: Theory and Application*. Ed.Academic Press. 1993. pp. 581.

29. M. Sadeghi, J. Kittler, K. Messer. "Real time segmentation of lip pixels for lip tracker initialization". *Lecture Notes in Computer Science*. Vol. 2124. 2001. pp. 317-324.

30. P. Gacon, P. Y. Coulon, G. Bailly. "Statistical active model for mouth components segmentation". *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP' 05*. Vol. 2. 2005. pp. 1021-1024.

31. B. Goswami, W. J. Christmas, J. Kittler. "Statistical estimators for use in automatic lip segmentation". *Proceedings of the 3rd European Conference on Visual Media Production (CVMP)*. 2006. pp. 79-86.

32. I. Mpiperis, S. Malassiotis, M. G. Strintzis. "Expression compensation for face recognition using a polar geodesic representation". *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*. 2006. pp. 224-231.

33. C. Bouvier, P. Y. Coulon, X. Maldague. "Unsupervised lips segmentation based on ROI optimisation and parametric model". *IEEE International Conference on Image Processing, ICIP 2007*. Vol. 4. 2007. pp. 301-304.

34. A. K. Jain, M. N. Murty, P. J. Flynn. "Data clustering: A review. ACM Computer Surveys". Vol. 31. 1999. pp. 264-323.

35. J. C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, 1981. pp. 256.

36. A. K. Jain, M. N. Murty, P. J. Flynn. "Data clustering: a review". *ACM Computing Surveys.* Vol. 31. 1999. pp. 264-323.

37. S. H. Leung, S. L. Wang, W. H. Lau. "Lip Image Segmentation Using Fuzzy Clustering Incorporating an Elliptic Shape Function". *IEEE Transactions on Image Processing*. Vol. 13. 2004. pp. 51-62.

38. A. W. C. Liew, S. H. Leung, W. H. Lau. "Segmentation of color lip images by spatial fuzzy clustering". *IEEE Transactions on Fuzzy Systems*. Vol. II. 2003. pp. 542-549.

39. S. L. Wang, W. H. Lau, S. H. Leung, A. W. C. Liew. "Lip segmentation with the presence of beards". *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'04*. Vol. 3. 2004. pp. 529-532.

40. Y. Mitsukura, M. Fukumi, N. Akamatsu. "A design of face detection system by using lip detection neuralnetwork and skin distinction neural network". *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics.* Vol. 4. 2000. pp. 2789 - 2793.

41. H. Takimoto, Y. Mitsukura, M. Fukumi, N. Akamatsu. "Face detection and emotional extraction system using double structure neural network". *Proceedings of the International Joint Conference on Neural Networks*. Vol. 2. 2003. pp. 1253 - 1257.

42. Y. Mitsukura, M. Fukumi, N. Akamatsu. "A design of face detection system using evolutionary computation". *Proceedings of TENCON 2000*. Vol. 2. 2000. pp. 398 - 402.

43. W. N. Lie, H. C. Hsieh. "Lips detection by morphological image processing". *Proceedings of the 1998 Fourth International Conference on Signal Processing, ICSP'98*. Vol. 2. 1998. pp. 1084 - 1087.

44. R. A. Rao, R. M. Mersereau. "Lip modeling for visual speech recognition". *1994 Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*. Vol. 1. 1994. pp. 587-590.

45. T. F. Cootes, D. Cooper, C. J. Taylor, J. Graham. "Active shape models - their training and application". *Computer Vision and Image Understanding*. Vol. 61. 1995. pp. 38 - 59.

46. T. F. Cootes, G. J. Edwards, C. J. Taylor. "Active appearance models". *Proceedings of the European Conference on Computer Vision.* Vol. 2. 1998. pp. 484-498.

47. A. Caplier. "Lip detection and tracking". *Proceedings of 11th International Conference on Image Analysis and Processing.* 2001. pp. 8 - 13.

48. A. Caplier, P. Delmas, D. Lam. "Robust initialisation for lips edges detection". *Proceedings of 11th Scandinavian Conference on Image Analysis*. 1999. pp. 523-528.

49. A. Turkmani, A. Hilton. "Appearance-based inner-lip detection". *Proceedings of the 3rd European Conference on Visual Media Production (CVMP 2006).* 2006. pp. 176-176.

50. M. Jiang, Z. H. Gan, G. M. He, W. Y. Gao. "Combining particle filter and active shape models for lip tracking". *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA 2006).* Vol. 2. 2006. pp. 9897- 9901.

51. Y. D. Jian, W. Y. Chang, C. S. Chen. "Attractor-guided particle filtering for lip contour tracking". *Lecture Notes in Computer Science.* Vol. 3851. 2006. pp. 653 - 663.

52. J. E. Hernández, F. Prieto, T. Redarce. "Fast active contours for sampling". *Proceedings of Electronics, Robotics and Automotive Mechanics Conference.* Vol. 2. 2006. pp. 9 - 13.

53. C. Xu, J. L. Prince. "Gradient Vector Flow: A new external force for snakes". *Proceedings of Computer Vision and Pattern Recognition (CVPR'97).* San Juan, Puerto Rico. 1997. pp. 66-71.

54. C. Xu, J. L. Prince. "Snakes, shapes, and gradient vector flow". *IEEE Transactions on Image Processing.* Vol. 7. 1998. pp. 359-369.

55. A. S. M. Sohail, P. Bhattacharya. "Automated lip contour detection using the level set segmentation method". *14th International Conference on Image Analysis and Processing (ICIAP 2007)*. 2007. pp. 425-430.

56. P. Viola, M. J. Jones. "Robust real-time face detection". *International Journal of Computer Vision.* Vol. 57. 2004. pp. 137-154.

57. N. Eveno, A. Caplier, P. Y. Coulon. "Accurate and quasi-automatic lip tracking". *IEEE Trans. on Circuits and Systems for Video Technology.* Vol. 14. 2004. pp. 706-715.

58. Z. Hammal, N. Eveno, A. Caplier, P. Y. Coulon. "Parametric models for facial features segmentation". *Signal Processing.* Vol. 86. 2005. pp. 399-413.

59. H. Seyedarabi, W. S. Lee, A. Aghagolzadeh. "Automatic lip tracking and action units classification using two-step active contours and probabilistic neural networks". *Proc. of the Canadian Conf. on Electrical and Computer Engineering, CCECE'06*. 2006. pp. 2021-2024.

60. S. Werda, W. Mahdi, A. B. Hamadou. "Colour and geometric based model for lip localisation: Application for lip-reading system". *14th International Conference on Image Analysis and Processing (ICIAP 2007).* 2007. pp. 9-14.

61. J. S. Chang, E.Y. Kim, S. H. Park. "Lip contour extraction using level set curve evolution with shape constraint". *Lecture Notes in Computer Science*. Vol. 4552. 2007. pp. 583-588.

62. M. K. Moghaddam, R. Safabakhsh. "TASOM-based lip tracking using the color and geometry of the face". *Proceedings of the Fourth International Conference on Machine Learning and Applications, ICMLA'05*. 2005. pp. 6.

63. L. Xie, X. L. Cai, Z. H. Fu, R. C. Zhao, D. M. Jiang. "A robust hierarchical lip tracking approach for lipreading and audio visual speech recognition". *Proceedings of 2004 International Conference on Machine Learning and Cybernetics.* Vol. 6. 2004. pp. 3620-3624.

64. K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre. "XM2VTSDB: the extended M2VTS database". *Proceedings of the Second International Conference on Audio- and Video-based Biometric Person Authentication, AVBPA'99*. 1999. pp. 72-77.

65. A. M. Martínez, R. Benavente. "The AR face database". *Technical Report 24. Computer Vision Center (CVC)*. Universidad Autónoma de Bacelona, Barcelona, España. 1998.

66. G. Chetty, M. Wagner. "Automated lip feature extraction for liveness verification in audio-video authentication". *Proceedings of Image and Vision Computing*. 2004. pp. 17-22.