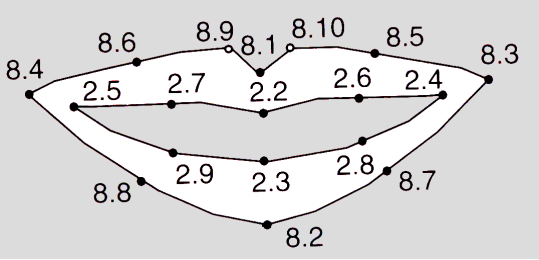
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MANIZALES

Extracción de los puntos característicos MPEG-4 de los contornos labiales sobre secuencias de video

Tesis de Maestría 2009



Silvana Lorena Vallejo Córdoba

slvallejoco@unal.edu.co

Asesor:

Nubia Liliana Montes C.

DIEEC

Universidad Nacional de Colombia Sede Manizales

Grupo de Percepción y Control Inteligente

Extracción de los puntos característicos MPEG-4 de los contornos labiales sobre secuencias de video

por:

Silvana Lorena Vallejo Córdoba

TESIS DE MAESTRÍA

Presentada a:

Departamento de Ingeniería Eléctrica, Electrónica y Computación

Facultad de Ingeniería y Arquitectura

En cumplimiento de los Requerimientos

para el Grado de

MAGISTER EN INGENIERÍA - AUTOMATIZACIÓN INDUSTRIAL

Asesor:

Nubia Liliana Montes Castrillón

Universidad Nacional de Colombia

Sede Manizales

A mis padres...

Abstract

The study of facial images, specifically those focused in mouth region, has grown exponentially during the last decade. This interest is motivated by the wide range of potential applications for systems able to code, interpret and recognize information transmitted by means of lip movements.

This document presents an automatic feature point extraction system for locating points of groups second and eighth of the MPEG 4 video face animation standard in facial images. These groups are designed for inner and outer lip contour definition. It also presents several approaches for the development of execution stages of the system, beginning with a study of different options for the mouth extraction and their correct segmentation in photographic images and video sequences; then, a shape based landmark extraction methodology from segmented image is introduced; finally, two tracking algorithms are implemented allowing the visual characteristics extraction on video sequences, without newly calculate them in each video frame.

Resumen

El estudio de imágenes faciales y específicamente de la región de la boca ha crecido exponencialmente durante la última década. Este interés está motivado por el amplio rango de aplicaciones potenciales para sistemas capaces de codificar, interpretar y reconocer información transmitida con el movimiento de los labios.

En este documento se presenta el desarrollo de un sistema que permite la extracción de los parámetros de animación facial de los grupos 2 y 8 del estándar MPEG 4, diseñados para la definición de los contornos labiales; también se presentan varios enfoques para el desarrollo de las etapas incluidas en su ejecución. Se inicia con un estudio de diferentes opciones para la extracción de la boca y su correcta segmentación en imágenes fotográficas y secuencias de video, luego se plantea una metodología basada en restricciones de forma para la ubicación de los puntos requeridos a partir de la imagen segmentada. Finalmente se implementan dos algoritmos de rastreo que permiten la extracción de las características visuales sobre secuencias de video, sin tener que recalcularlas para cada cuadro de la secuencia.

Porque el hombre cree con más disposición lo que preferiría que fuera cierto. En consecuencia rechaza cosas difíciles por impaciencia en la investigación; silencia cosas, porque reducen las esperanzas; lo más profundo de la naturaleza, por superstición; la luz de la experiencia, por arrogancia y orgullo; cosas no creídas comúnmente, por deferencia a la opinión de la mayoría. Son pues innumerables los caminos, y a veces imperceptibles, en que los afectos colorean e infectan la comprensión.

— FRANCIS BACON

Agradecimientos

Agradezco con mucho cariño a todas las personas que contribuyeron con mi desarrollo como investigadora, especialmente a los integrantes del grupo de investigación PCI, con los que he compartido mucho más que un lugar de trabajo.

A mi directora Nubia Liliana Montes y a mi director honorario Juan Bernardo Gómez, por brindarme la oportunidad de trabajar bajo su constante guía, apoyo e infinita paciencia.

A COLCIENCIAS y al proyecto Identificación de posturas labiales en pacientes con labio y/o paladar hendido corregido, que financiaron mi trabajo en esta tesis.

A mis amigos, los que se encuentran cerca o lejos, quienes a pesar de todo siempre estuvieron para sacarme de mis bloqueos mentales y momentos de incalculable tedio.

A mi familia, por todo su apoyo y especialmente a mi tía Myriam por todos sus cuidados.

De manera muy especial a mis padres, que con todo su amor y respaldo han sido mi inspiración en todas las etapas de mi vida.

Índice

Lista de figuras

Lista de tablas

Acrónimos

PCI	Percepción y Control Inteligente
MPEG	Motion Picture Expert Group
FDPs	Facial Definition Parameters
FAPs	Facial Animation Paratemers
FAPUs	Facial Animation Paratemers Units
AVSR	Audio Visual Speech Recognition
HMM	Hidden Markov Model
ASM	Active Shape Model
AAM	Active Aparence Model
DCT	Discrete Cosine Transform
ICA	Independent Component Analysis
PCA	Pricipal Component Analysis
ROI	Region of Interest
BMA	Block Matching Algorithm
OL	Over Lap
SE	Segmentation error
E	Eficiencia
P	Perdida
ECMN	Error Cuadratico Medio Normalizado

Índice de algoritmos

1 Introducción

Reconocer automáticamente un objeto mediante una computadora es una tarea tradicional dentro de la Inteligencia Artificial. El rostro es uno de los patrones más comunes en nuestro entorno; para cualquier persona, aún un niño, su reconocimiento no constituye mucha dificultad. La identificación de la información expresada en él de forma automática representa, aún en nuestros días, uno de los desafíos más fuertes en el área del reconocimiento de patrones.

El procesamiento de imágenes faciales, específicamente la detección e identificación de gestos bucales en imágenes y secuencias de video, es un campo con gran variedad de aplicaciones como la réplica de los gestos para transmisión de información visual comprimida, la generación de caracteres virtuales, las mediciones antropométricas en medicina e incluso el comando automático de instrumentos y robots en interfaces hombre máquina.

La identificación de posturas y/o de la dinámica labial permite el estudio y seguimiento de la expresión del rostro y de la información que quiere expresar, por esta razón en el área de visualización científica, en las últimas décadas, se percibe un aumento sustancial de investigaciones en lectura de labios, reconocimiento de gestos y análisis de expresiones faciales [63]. Muchos autores se han dedicado a generar algoritmos para la correcta segmentación y caracterización de esta región ideando múltiples metodologías que resultan muy útiles dentro del campo de acción en que fueron planteadas. Esta multiplicidad en las características de definición de la boca hace que sea muy difícil la integración de estos métodos entre si, y más aún con aplicaciones comerciales existentes que normalmente se definen bajo un estándar ya sea en imágenes o video.

MPEG 4 es la primera representación estándar que modela una escena audiovisual como una composición de objetos audiovisuales con características y conducta específicas, eficientemente en espacio y tiempo ([40], [41]); una de las aplicaciones más interesantes es la composición de volumen natural y sintético.

El propósito de este trabajo fue aprovechar las características de animación facial definidas dentro del estándar MPEG 4, para realizar una aplicación lo bastante general como para funcionar con robustez

ante diferentes biotipos faciales y con la intención de que sea fácilmente integrable y escalable con otras aplicaciones que trabajen bajo este estándar.

La animación de modelos faciales requiere de un conjunto de características que pueden generarse sintéticamente o pueden extraerse por el análisis de caras reales. Adoptando los parámetros de definición y animación facial del estándar MPEG 4 se pueden sintetizar expresiones bucales (visemas) detectadas en cuadros de secuencias de video.

Existen dos grupos de parámetros de animación facial (FAPs) especialmente creados para la definición de los contornos labiales, en el desarrollo de esta tesis se usaron estos grupos como base para la definición de los contornos y su deformación en una secuencia dinámica de habla.

Este trabajo se desarrollo dentro del proyecto Identificación de posturas labiales en pacientes con labio y/o paladar hendido corregido, financiado por Colciencias y ejecutado por la Universidad de Caldas, la Universidad Nacional de Colombia sede Manizales y el Hospital Infantil Rafael Henao Toro. En proyectos de investigación desarrollados con anterioridad en la Universidad Nacional de Colombia sede Manizales [31], asociados al tratamiento de pacientes con LPH, se hizo evidente que la evaluación post-operatoria en niños con LPH es más efectiva si se analiza conjuntamente el control de la calidad de emisión acústica y la dinámica de cambio de las posturas labiales; razón por la cual en este proyecto se propuso el uso de secuencias de video para la construcción de un sistema que brinde solución al problema de estimación de la función articulatoria de niños con LPH corregido. Adicionalmente, trabajar con secuencias de video ofrece la posibilidad de implementación de sistemas finales de bajo costo y no invasivos.

En esta tesis se consiguió extraer de una secuencia de video los indicadores característicos relacionados a los dos grupos de características MPEG 4 que describen la boca, desde la segmentación acertada de la región y el seguimiento de estos indicadores en la secuencia.

En el documento se presentan varios enfoques para el desarrollo de las etapas presentes en la ejecución del sistema de extracción de puntos característicos, comenzando por un estudio de diferentes opciones para la extracción de la región de interés y su correcta segmentación. Se propone una nueva transformación de color que permite la segmentación de los labios en secuencias de video y bajo condiciones variables de iluminación.

A partir de la región segmentada se definen de los contornos interno y externo de la boca, para lo cual se desarrollo una metodología que permite la extracción de los grupos 2 y 8 de animación facial como se especifican en el estándar MPEG 4.