UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ
FACULTAD DE INGENIERÍA

Análisis de relevancia de descriptores de forma para el reconocimiento de gestos faciales en imágenes 3D

Julián Severiano Rodríguez Acevedo

Universidad Nacional de Colombia
Facultad de Ingeniería
Bogotá, Colombia
2012

Análisis de relevancia de descriptores de forma para el reconocimiento de gestos faciales en imágenes 3D

Julián Severiano Rodríguez Acevedo

Tesis presentada como requisito parcial para optar al título de:
Magister en Ingeniería Automatización Industrial

Director:
Flavio Prieto, PhD.

Línea de Investigación:
Automatización de procesos

Universidad Nacional de Colombia
Facultad de Ingeniería
Bogotá, Colombia
2012

Dedicatoria

A mis padres Severiano y Nidia, quienes siempre han dado todo lo que está a su alcance para apoyarme y guiarme a cumplir mis metas.
A mi esposa Paola por tener la paciencia de esperar durante el tiempo que he dejado de dedicarle.
A mis Hermanas Bibiana y Laura a quienes nunca les falta una palabra de motivación que me anima a seguir adelante.
A mi sobrinito Esteban a quien ví dar sus primeros pasos y escuché sus primeras palabras durante la realización de este trabajo.

Agradecimientos

El autor agradece en primer lugar a Dios quien me ha conducido a cumplir mis metas, a todos aquellos quienes de una u otra forma han colaborado en la realización de este proyecto, en especial al profesor Flavio Prieto, quien con sus conocimientos y acertadas recomendaciones condujo la realización de este trabajo.

El autor agradece también en general a la Universidad Nacional de Colombia sede Bogotá por brindar los recursos bibliográficos necesarios en la investigación que requirio el desarrollo de este proyecto.

Resumen

Actualmente existe un creciente interés por estudiar técnicas basadas en visión artificial y computación gráfica para caracterizar el rostro humano, con el fin de realizar antropometría facial, reconocimiento e identificación de personas y sistemas de interacción hombre - máquina.

Gracias a los métodos actuales es posible tener modelos 3D del rostro que permiten sacar provecho de la gran cantidad de información geométrica que tiene el rostro. Esta información es obtenida frecuentemente mediante el cálculo de descriptores de forma, los cuales permiten obtener de forma numérica, información relevante acerca de la geometría del rostro.

El objetivo de este trabajo es realizar un análisis de relevancia de tres descriptores de forma (DESIRE, Spherical Spin Image y Cone Curvature), para determinar la viabilidad de realizar reconocimiento de gestos faciales con alguno de ellos o con alguna combinación de los mismos. Se presentan análisis de la capacidad de recuperación de cada uno de ellos mediante las curvas de precision-recall, análisis de la relevancia de sus características mediante el indice de Fisher; así como el entrenamiento de dos clasificadores diferentes, tomando las características más relevantes de cada descriptor; se analiza además el costo computacional de cada uno de ellos y la viabilidad para implementarlos en un sistema de reconocimiento en tiempo real.

Palabras clave: descriptores de forma 3D, extracción de características, reconocimiento de gestos.

Abstract

Currently there is a growing interest in studying techniques based on computer vision and computer graphics to characterize the human face, in order to perform facial anthropometry, recognition and identification of people and systems of human - machine interaction. With current methods it is possible to have 3D facial models that allow to take advantage of the large amount of geometric information with the face. This information is often obtained by calculating shape descriptors, which allow us to obtain numerically relevant information about the geometry of the face.
The objective of this work is to analyze the relevance of three shape descriptors (DESIRE, Spherical Spin Image and Cone Curvature) to determine the feasibility of facial recognition to any of them or any combination thereof. Are presented analysis of the retrieval of each by the precision-recall curves, analysis of the relevance of their characteristics by Fisher index, as well as the training of two different classifiers, taking the most important characteristics of each descriptor, also discusses the computational cost of each and feasibility for deployment

in a recognition system in real time.

Keywords:shape descriptors 3D, feature extraction, gesture recognition

Contenido

Lista de Tablas

Lista de Figuras

Introducción

En los últimos años ha existido un creciente interés en mejorar todos los aspectos referentes a la interacción entre humanos y máquinas, con el objetivo de que dicha interacción sea tan natural como se realiza entre humanos. Una manera de mejorar dicha interacción puede ser mediante la identificación de las emociones; la forma más común en que las personas expresan sus emociones es a través de las expresiones faciales, las cuales identificamos con relativa facilidad; sin embargo, la implementación de un sistema automático de reconocimiento de expresiones faciales tiene un alto grado de complejidad.

Ekman [8], fue uno de los pioneros del reconocimieto de emociones mediante el análisis de expresiones faciales y definió 6 grandes expresiones: alegría (HA), disgusto (DI), miedo (FE), enojo (AN), sorpresa (SU) y tristeza (SA). Se han desarrollado diferentes técnicas para reconocer expresiones faciales, la mayoría de ellas basadas en imágenes 2D estáticas [9] y en el análisis de secuencias de imágenes, estas últimas divididas en las aproximaciones por flujo óptico [1], rastreo de características [2] y las basadas en el alineamiento del modelo ([7]).

Una de las principales motivaciones del reconocimiento de características faciales 3D, es solucionar los problemas existentes en los métodos de reconocimiento 2D, tales como la sensibilidad a las variaciones de iluminación, escala y orientación. Existen diferentes maneras de representar la información en tres dimensiones. Por un lado la información puede ser representada como una nube de puntos en el espacio. Estos puntos pueden representarse también en forma de mallado, donde cada punto es un vértice y a su vez los vértices están unidos por artistas. El polígono o geometría a utilizar para realizar esta unión es habitualmente un triángulo, ya que es la manera más simple de aproximar una nube de puntos a una superficie.

Los modelos tridimensionales del rostro, tienen gran cantidad de información acerca de su morfología, que puede ser extraída empleando diferentes métodos que involucran algún tipo de descriptor de forma. Los descriptores de forma son herramientas matemáticas que permiten extraer en forma de datos númericos, información acerca de la geometría de un objeto 3D. Aunque se han evaluado varios descriptores con el propósito general de reconocer objetos en una escena [10, 11] y se han hecho análisis de algunos otros sobre puntos de interés en el rostro como en [12], donde se análizan varios descriptores basados en curvatura, o en [13] donde usando las imágenes Spin se extraen puntos característicos del rostro y [14, 15] en donde se evalúan características del rostro basándose en análisis de firma y plantillas. A

nuestro conocimiento no se ha realizado un análisis de este tipo con descriptores de forma como DESIRE, Spherical Spin Image y Cone Curvature, por este motivo, en este trabajo se realiza un estudio de estos descriptores, con el propósito de determinar cual de ellos tiene un mejor comportamiento en el reconocimiento de gestos faciales.

1 FUNDAMENTACIÓN TEÓRICA

En este capítulo se presentan los conceptos fundamentales acerca de la representación de los modelos 3D y la estructura de las mallas triangulares, dado que son las empleadas en los modelos de la base de datos utilizada en este trabajo. Se muestra también la fundamentación teórica de los descriptores de forma y su clasificación, así como una breve presentación del estado del arte de la aplicación de los descriptores en reconocimiento de gestos, y finalmente, las métricas empleadas usualmente para verificar la similitud entre características a partir de la información suministrada por los descriptores.

1.1. Representación de modelos 3D

La reconstrucción o representación 3D es el proceso mediante el cual, objetos del mundo real, son reproducidos en la memoria de un computador representando sus características físicas, (dimensiones, forma, volumen). Se han desarrollado variedad de métodos de representación cuyo objetivo principal es realizar la conexión de puntos que conforman el objeto en forma de elementos de superficie (cualquier figura geométrica). La representación de objetos 3D, tiene un gran interés, ya que se han encontrado aplicaciones en diversos campos como son: diseño, automatización de procesos, conducción automática de vehículos, mapeo de terrenos, guiado de robots, entre otros.
Un proceso de representación tridimensional, en general involucra las siguientes etapas: adquisición, registro, integración y ajuste. La etapa de adquisición consiste en sensar la superficie del objeto desde un número determinado de vistas o imágenes de rango. El registro consiste en llevar las múltiples imágenes adquiridas, a un sistema de coordenadas común. La etapa de integración, consiste en eliminar datos redundantes y generar datos en las regiones que tengan ausencia de información. La etapa de ajuste, es la encargada de estimar un modelo matemático, estos modelos pueden tener gran variedad de representaciones, una de las más empleadas es la representación por mallas triangulares. La Figura **1-1** muestra el proceso de construcción del modelo 3D a partir de un escena real, y la Figura **1-2** ilustra un ejemplo del mismo proceso.

1.1.1. Métodos de digitalización 3D

Los métodos de digitalización 3D, hacen referencia a las técnicas existentes para representar digitalmente, en un computador, las características físicas de un objeto o escena real. Estos

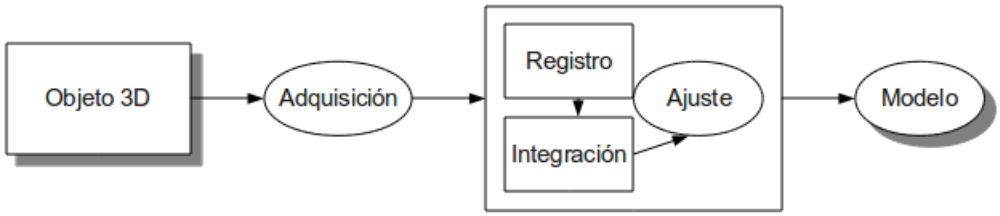
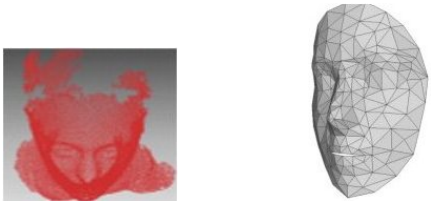


Figura 1-1: Proceso de reconstrucción 3D

(a) Adquisición (b) Registro



(c) Integración (d) Ajuste

Figura 1-2: Ejemplo reconstrucción

métodos se pueden clasificar en pasivos y activos [4].

1.1.1.1. Métodos pasivos

En forma general, los métodos pasivos se limitan al estudio de la información visual de una o más vistas, tomadas en distintos instantes de tiempo y observando la escena, sin actuar sobre ella, dentro de estos métodos se encuentran:

Geometría epipolar. Conociendo la posición de la imagen de un punto en dos fotografías, y la posición y orientación de las cámaras, se puede reconstruir la posición del punto actual por triangulación. Un punto de la escena con los dos centros de proyección forman un triángulo, cuya intersección con los planos de proyección son las líneas epipolares. El resultado es una imagen parcial con profundidades (una nube de puntos en el espacio). En [5] se puede encontrar la explicación detallada de este tipo de reconstrucción, conocida también como reconstrucción stereo.

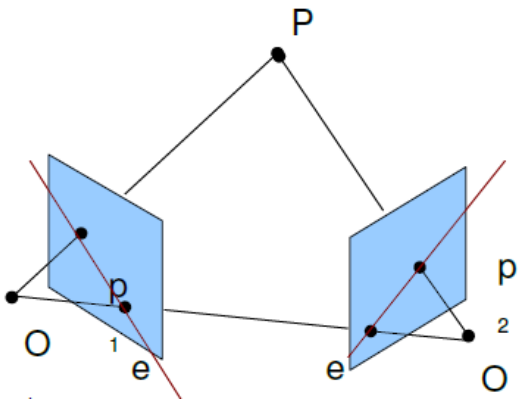


Figura 1-3: Proceso de Reconstrucción, mediante geometría epipolar. [5]

Shape from Shading. Este método [6], permite realizar la reconstrucción, a partir de una sóla imagen del modelo, siempre y cuando se cumplan algunas condiciones:

- La superficie del objeto, tiene un comportamiento Lambertiano, es decir, que la radiancia no varía en función del ángulo de observación.
- La Reflectividad es constante.
- Todos los puntos de la imagen reciben iluminación directa.

A partir de la perspectiva. Dibujo en perspectiva, usando puntos de fuga, para cada punto de la imagen del que sepamos su proyección en el plano, se pueden calcular sus coordenadas.

1.1.1.2. Métodos Activos

Se conocen como técnicas de reconstrucción activa aquellas en las que se actúa con algún tipo de energía externa sobre el objeto a representar, con el fin de capturar de forma más sencilla sus dimensiones. En los métodos activos, existe retroalimentación de la información visual sobre:

- El movimiento, localización y orientación del sensor.
- Los lentes y el sistema de adquisición.
- La métrica y el estado interno del sistema (calibración visual).

Una imagen de rango es un conjunto de puntos del espacio, obtenidos por un escáner de barrido, que representa la superficie de un objeto. La principal característica de las imágenes de rango es que cada píxel es una medida de la distancia entre un punto visible del objeto y un marco de referencia conocido. Por lo tanto, una imagen de rango reproduce la estructura 3D de una escena. (Figura **1-4**)

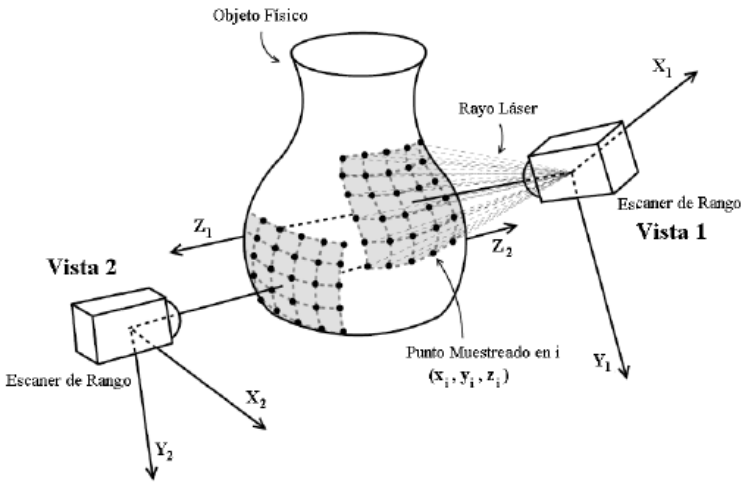


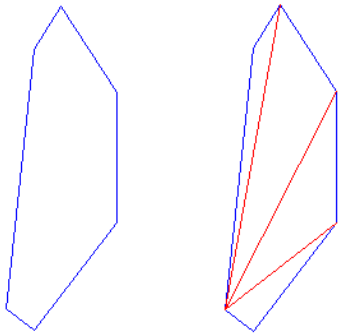
Figura 1-4: Proceso de Adquisición de imágenes de rango para dos vistas diferentes [3]

Los sensores de rango activo, utilizan diferentes técnicas, dentro de las más comunes se encuentran la triangulación, radar/sonar, interferometría de Moiré.

1.1.2. Mallas triangulares

Existen diferentes maneras en que un modelo 3D puede ser representado luego de ser digitalizado, la forma más común de representar modelos 3D es mediante mallas poligonales, estas están conformadas por la unión de poligonos cuyos vértices son coplanares; si todos los poligonos son triángulos, el mallado se conoce como malla triangular; por simplicidad en muchos procesos las mallas poligonales generalmente, son convertidas en triangulares como se ilustra en la Figura **1-5**, donde una malla poligonal con k vértices es transformada en la unión de $k-2$ triángulos.

Una malla triangular está conformada por un conjunto de triángulos $T = \{T_1, T_2, ...T_m\}$, formados por un conjunto de vértices $V = \{v_i | v_i = (x_i, y_i, z_i) \in R^3, 1 \leq i \leq n\}$ y un conjunto de índices $P = \{p_1, p_2, ...p_m\}$ que relacionan a los vértices que conforman cada triángulo, es decir proporcionan información acerca de la topología. La Figura **1-6** ilustra un ejemplo de los listados de vértices e índices que conforman el modelo 3D de un icosaedro



(a) (b)

Figura 1-5: Triangulación de un polígono

$(n = 12, m = 20)$. En este ejemplo, la primera línea del listado de triángulos T_1, indica que el primer triángulo del modelo lo conforman las vértices p_1, p_2 y p_3.
En algunos formatos de representación de los modelos 3D, se cuenta con información adicional como un listado de colores o texturas, sin embargo, la información de la geometría y la topología, son suficientes para representar el modelo.
Algunos programas de computador pueden generar modelos sintéticos de objetos reales, pero como se explicó anteriormente, en la actualidad se cuenta con técnicas de reconstrucción 3D, cuyo propósito general es generar una nube de puntos a partir de muestras geométricas en la superficie de un objeto real. Estos puntos son sometidos luego a un proceso de triangulación para obtener la información toplogica del objeto. En algunos casos, un sólo escaneo no produce el modelo completo del objeto, es necesarío, entonces realizar múltiples tomas desde diferentes direcciones para tener información de todos los puntos de vista del objeto. Estas tomas o escaneos tienen que ser luego integrados en un sistema común de referencia, que transforma las coordenadas locales de cada toma, en coordenadas globales del modelo.

Existen diferentes formatos de modelos en 3D, que dependen de la forma en que fueron construidos, y la información que proporcionan, los más populares son: VRML (Virtual Reality Modeling Language), DXF (AutoDesk Drawing eXchange Format), 3DS (3D Studio file format), OFF(Objet File Format) y OBJ (Wavefront Object files).

1.2. Descriptores de forma

Los descriptores de forma extraen información geométrica que describe un objeto 3D. Generalmente son usados para buscar similitudes entre diferentes modelos (Figura **1-7**).
Existen gran cantidad de descriptores de forma 3D, y en la lieratura se encuentran clasificados

T_1	1	2	3
T_2	1	3	4
T_3	1	4	5
T_4	1	5	6
T_5	1	6	2
T_6	2	6	10
T_7	3	2	1
T_8	4	3	12
T_9	5	4	8
T_{10}	6	5	9
T_{11}	7	9	8
T_{12}	7	10	9
T_{13}	7	11	10
T_{14}	7	12	11
T_{15}	7	8	12
T_{16}	8	4	12
T_{17}	9	5	8
T_{18}	10	6	9
T_{19}	11	2	10
T_{20}	12	3	11

p_1	1	0	0
p_2	0.44	0.88	0
p_3	0.44	1.4	0.85
p_4	0.44	-0.72	0.52
p_5	0.44	-0.72	-0.52
p_6	0.44	1,4	.0,85
p_7	-1	0	0
p_8	-0.44	-0.88	0
p_9	-0.44	-1.4	-0.85
p_{10}	-0.44	0.72	-0.52
p_{11}	-0.44	0.72	0.52
p_{12}	-0.44	-1.4	0.85

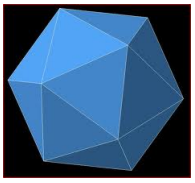


Figura 1-6: Conjunto de vértices e índices que conforman un icosaedro

de diferentes maneras; sin embargo, podemos agruparlos en tres grandes categorías:

1. Basados en características.
2. Basados en gráficos.
3. Basados en vistas.

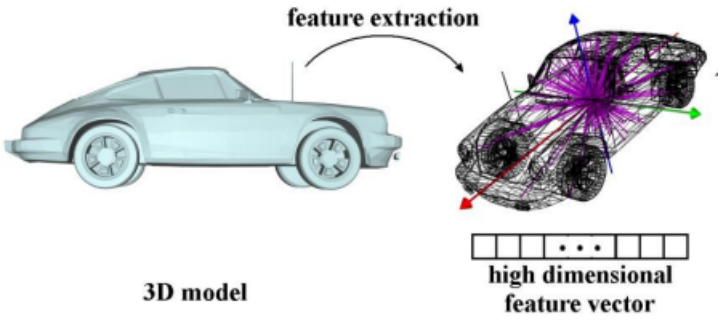


Figura 1-7: Extracción de Características [32]

1.2.1. Características de los descriptores

Un buen descriptor de forma debe tener al menos las siguientes características:

1. Invarianza con respecto a traslación, rotación, escala y refelexión del modelo 3-D.
2. Robustez con respecto al ruido de la superficie y valores atípicos.
3. Robustez con respecto a niveles de detalle.

Si se tiene un modelo 3-D I, y τ es una combinación de traslaciones, rotaciones y escalados de I, y f y f' son los vectores característicos de I, y τ respectivamente, debe cumplirse entonces que $f = f'$.
Actualmente, la mayoría de modelos 3D, se obtienen mediante scanner 3-D, la adquisición de los modelos mediante estos dispositivos, tiene la desventaja de que en algunas ocaciones quedan huecos en la generación del modelo, de manera que si un mismo objeto se escanea dos veces, es posible que la segunda vez el modelo tenga una teselación diferente; sin embargo la apariencia de los 2 modelos 3-D generados será prácticamnete idéntica. El descriptor de forma debe estar en la capacidad de conservar la alta similitud entre los dos modelos.
Generalmente los modelos 3-D tienen gran cantidad de triángulos, pero para algunas aplicaciones no se requiere gran nivel de detalle, en la Figura **1-8** se ilustra un modelo en cuatro niveles de detalle diferentes, el descriptor de forma extraido a cada uno de ellos evidentemente no arrojará el mismo resultado, pero deberá ser capaz de distinguir diferencias entre estos y un modelo diferente.

Figura 1-8: Modelo 3-D en cuatro niveles de detalle. [32]

1.2.2. Descriptores basados en características

Este tipo de descriptores son los más populares, y además pueden ser subclasificados en: 1. características globales, 2. características locales y 3. mapas espaciales.

El cálculo de características globales para representar objetos incluyen area, volumen y momentos. En [16] se calcula un número de distribuciones de forma globales para representar objetos 3D. Las funciones de forma medidas incluyen el ángulo entre tres puntos aleatorios, la distancia entre dos puntos aleatorios, el área formada por tres puntos aleatorios y el volumen entre cuatro puntos aleatorios sobre la superficie. En [17] Ohbuchi mejora la función de forma midiendo no sólo la distancia entre dos puntos, sino también la orientación mutua de la superficie sobre la que están localizados los puntos. Zaharia [21] introdujo un descriptor de forma que calcula la distribución del índice de forma sobre toda la malla. Los descriptores basados en características globales son en general eficientes computacionalmente, sin embargo no discriminan muy bien cuando los objetos o modelos 3D tiene pequeñas diferencias.

Las características locales, por lo general se concentran en puntos que pueden proporcionar cierta información de interés, es el caso de las imágenes spin [37], dónde se seleccionan puntos que son tomados como origen para generar un histograma apartir de los puntos vecinos al origen, representando la geometría local de la superficie entorno al punto origen. La mayoria de descriptores basados en las características locales, usan propiedades geométricas del objeto tales como la curvatura o las normales, para describir un punto sobre la superficie del mismo. En [19] se emplean las propiedades de la curvatura principal para identificar los extremos finales como puntos salientes. Castellani [20] propone otro método para detectar puntos salientes del objeto 3D basado en la medida de qué tanto se desplaza un vértice después de filtrar. Los puntos salientes son descritos usando descriptores locales basados en los modelos ocultos de Markov.

Los descriptores basados en mapas espaciales describen el objeto mediante la captura de lugares físicos sobre ellos. En [22], se emplea la idea de que los coeficientes armónicos esféricos,

reconstruyen una aproximación del objeto en diferentes resoluciones, para mostrar que los armónicos esféricos se pueden usar para transformar una rotación dependiente de la forma, en una independiente, sin necesidad de que exista una normalización previa. Laga en [23] muestrea puntos uniformemente sobre una esfera unitaria y usa la transformada wavelet esférica para representar objetos 3D. Wavelets, son funciones bases que representan una función dada en múltiples resoluciones. Laga propone tres descriptores basados en wavelets esféricas: usando los coeficientes como vectores de características, usando la energía L_1 de los coeficientes, y usando la energía L_2 de los coeficientes. En [18] se captura la forma del objeto 3D usando el mapa de curvatura de la superficie del objeto.
Otro de los descriptores basados en mapas espaciales es el Ray Based, que hace parte del descriptor DESIRE que se implementó en el presente trabajo, y el cual se explicará en el capítulo siguiente.

1.2.3. Descriptores basados en Gráficos

Los descriptores básados en gráficos, en lugar de usar las propiedades geométricas del modelo para definir la forma, emplean la información topológica para describir su forma. Los gráficos se construyen mostrando cómo los diferentes componentes del modelo están unidos entre sí. Las representaciones gáficas incluyen model graphs, Reeb graphs y Skeleton graphs. Estos métodos son costosos computacionalmente y sensibles a pequeños cambios en la topología. En [24] es usado el skeletal graph como un descriptor para codificar tanto las propiedades geométricas como topológicas de un objeto 3D. Hiliaga [25] introduce Reeb graphs para realizar búsqueda de modelos similares.

1.2.4. Descriptores basados en vistas

El más popular de estos descriptores es el Ligth Field (LFD) desarrollado por Chen [26], un ligth field alrededor de un objeto 3D, es una función la proyeccion del objeto 3D desde un punto dado. Cada ligth field de un objeto 3D, es representada como una colección de imágenes 2D obtenidas a partir de puntos de vista diferentes, distribuidos uniformemente sobre una esfera. El método ubica los puntos de vista sobre 20 vértices de un dodecaedro regular y usa proyecciones ortogonales para capturas las siluetas del modelo. Las imágenes obtenidas son descritas usando momentos Zernike y Fourier para definir la región y el contorno del modelo 3D, respectivamente.
Las imágenes de profundidad, y las siluetas empleadas para la construcción del descriptor DESIRE, pueden considerarse dentro de esta categoría de descriptores de forma, puesto que, como se detallará en el capítulo siguiente, se construyen a partir de vistas diferentes de un paralelepípedo que encierra el objeto 3D.

(a) (b) (c)

Figura 1-9: Tres tipos diferentes de descriptores. 1.9(a) muestra las curvaturas principales como ejemplo de características. En 1.9(b) un ejemplo del esquema de construcción del descriptor skeletal graph([24]). En 1.9(c) proceso de construcción del descriptor LFD ([26])

En la Figura **1-9** se ilustran ejemplos de tres tipos diferentes de descriptores.

1.3. Reconocimiento de gestos

En general, los descriptores de forma se han empleado con el objetivo de realizar reconocimiento de objetos 3D dándole utilidad práctica en tareas como inspección visual, guiado de robots y control de calidad. Algunos trabajos se han centrado en extraer ciertas características del rostro humano como en [13], donde mediante el uso de las imágenes spin, se localizan tres puntos característicos del rostro: la punta de la nariz y los puntos que definen el ángulo interno de cada uno de los ojos. En otros trabajos como en [27] se han hecho revisiones acerca de los métodos empleados para realizar reconocimiento facial; sin embargo no son muchos los trabajos que se han centrado en el reconocimiento de gestos faciales a partir de modelos 3D. En [45] se realiza este reconocimiento basándose en las propiedades de los segmentos de línea que unen algunos puntos característicos del rostro obteniendo un desempeño promedio de reconocimiento del 99 % para la expresión de sorpresa; Wang en [46] realiza también un estudio del reconocimiento de expresiones en mallados 3D, basándose en las curvaturas principales. En el último capítulo de este trabajo se presenta un cuadro comparativo entre los resultados obtenidos y los resultados de los trabajos citados anteriormente.

Algunos descriptores de forma han mostrado buenos resultados en la representación de modelos 3D, como es el caso del DESIRE, y Cono Curvatura, y se han desarrollado algunas variaciones de otros conocidos como es el caso de las imágenes Spin esféricas desarrollado a partir de las imágenes Spin, por este motivo se realizó el estudio con estos descriptores para verificar su desempeño frente al reconocimiento de expresiones faciales en modelos 3D.

1.4. Métricas de similitud

En esta sección se presenta una breve descripción, acerca de los fundamentos matemáticos de las métricas o indicadores de similitud empleados en el desarrollo de este trabajo. La similitud, es una cantidad que refleja la relación entre dos objetos.
Con la información proporcionada por el descriptor, generalmente un vector de características del modelo, se requiere, con el objetivo de realizar una búsqueda en una base de datos de modelos, realizar algún tipo de comparación entre vectores característicos para identificar aquellos que puedan tener ciertas similitudes. Una de las formas más comunes, consiste en el cálculo de los vecinos más cercanos, es decir, úbicar aquellos modelos cuyos vectores característicos están más cercanos y hacer un ranking de distancias. Las métricas empleadas generalmente son: distancia l_p y coeficiente de correlación líneal, las cuales se describen a continuación:

1.4.1. Distancia l_p

Es una forma de medir la distancia entre dos vectores y está definida como:

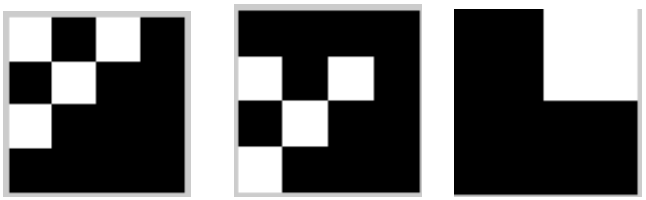
$$l_p(f', f'') = \|f' - f''\|_p = \left(\sum_{i=1}^{N} |f'_i - f''_i|^p\right)^{1/p}, p = 1, 2, ...$$

siendo $f' = (f'_1,\ f'_2, ...f'_N), f'' = (f''_1, f''_2, ...f''_N)$ los vectores característicos de dos modelos diferentes.
Cuando $p = 2$, se tiene la distancia euclidiana, para $p = \infty$, la métrica es llamada distancia máxima, calculada como:

$$l_\infty(f', f'') = \|f' - f''\|_\infty = \left(\begin{smallmatrix} max \\ 1 \leq i \leq N \end{smallmatrix}\right) |f'_i - f''_i|.$$

Estas normas no tienen un buen comportamiento en el dominio espacial. Los valores f'_i y f''_i, corresponden a características relacionadas con una misma región en el modelo 3D, sin embargo, en lugar de calcular las diferencias sólo entre valores característicos de la misma región, la distribución de los valores a través de las regiones vecinas debe tenerse en cuenta también. Como ejemplo consideremos que f', f'' y f''' son los vectores característicos correspondientes a las imágenes de la Figura **1-10**.
Si calculamos la norma l_2, tendríamos $l_2(f', f'') = \sqrt{8} > l_2(f', f''') = \sqrt{6}$ y con l_∞, $l_\infty(f', f'') = l_2(f', f''') = 1$.

En ambos casos tenemos una contradicción con respecto a la percepción humana, ya que es evidente que es mayor el parecido entre f' y f'' que entre f' y f''', es decir, debería cumplirse que $l_p(f', f'') < l_p(f', f''')$.
Una forma de corregir el inconveniente anterior consiste en transformar las imágenes obtenidas al dominio frecuencial, con el objetivo de correlacionar los componentes de cada vector de



(a) (b) (c)

Figura 1-10: Visualización de tres vectores de características (f', f'', f'''), los cuadros blancos corresponden a un valor de 1, los cuadros negros corresponden a 0

características.

Aplicando la transformada discreta de Fourier a cada imagen obtenemos unos coeficientes complejos de Fourier, cuyas magnitudes pueden ser usadas para representar el vector de características en el dominio frecuencial. Por ejemplo, para las imágenes de la Figura **1-10** la transformación al dominio frecuencial está dada en las ecuaciones 1-1, 1-2 y 1-3.

$$\begin{bmatrix} 1\,0\,1\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,0\,0\,0 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & \sqrt{2} & 2 & \sqrt{2} \\ \sqrt{2} & 2 & \sqrt{2} & 0 \\ 2 & \sqrt{2} & 4 & \sqrt{2} \\ \sqrt{2} & 0 & \sqrt{2} & 2 \end{bmatrix} \tag{1-1}$$

$$\begin{bmatrix} 0\,0\,0\,0 \\ 1\,0\,1\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & \sqrt{2} & 2 & \sqrt{2} \\ \sqrt{2} & 2 & \sqrt{2} & 0 \\ 2 & \sqrt{2} & 4 & \sqrt{2} \\ \sqrt{2} & 0 & \sqrt{2} & 2 \end{bmatrix} \tag{1-2}$$

$$\begin{bmatrix} 0\,0\,1\,1 \\ 0\,0\,1\,1 \\ 0\,0\,0\,0 \\ 0\,0\,0\,0 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 2\sqrt{2} & 0 & 2\sqrt{2} \\ 2\sqrt{2} & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 2\sqrt{2} & 2 & 0 & 2 \end{bmatrix} \tag{1-3}$$

con estos vectores de características las nuevas distancias son:

$l_2(f', f'') = 0 < l_2(f', f''') = 48$ y con l_∞, $l_\infty(f', f'') = 0 < l_2(f', f''') = 4$. Lo cual es evidentemente más razonable con respecto a la similitud que se puede apreciar de las imágenes consideradas en el ejemplo.

1.4.2. Coeficiente de correlación líneal

En el presente trabajo, se implementó el descriptor SSI, el cual como se detallará en el siguiente capítulo, proporciona una serie de imágenes que representan el modelo 3D. Con el objetivo de encontrar similitudes entre modelos, es posible emplear el coeficiente de correlación líneal, el cual mide la relación entre las variables a considerar, en este caso entre las imágenes que describen el modelo 3D. El coeficiente se obtiene según la Ecuación 1-4.

$$r = \frac{\sum_m \sum_n (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \overline{A})^2)(\sum_m \sum_n (B_{mn} - \overline{B})^2)}}. \tag{1-4}$$

Donde A_{mn} y B_{mn} son las dos imágenes de tamaño $m \times n$ a comparar.
Los valores que puede tomar el coeficiente de correlación r, son $-1 < r < 1$. La correlación es más fuerte en tanto r se apróxime más a 1, lo cual quiere decir, que al realizar la comparación entre un par de imágenes spin si r es un valor alto, las imágenes son similares.
El coeficiente de correlación líneal, proporciona una manera simple y estable para comparar dos imágenes e impone un orden de correspondencia de puntos, de forma que se pueda diferenciar que tan buenas o malas son las correspondencias entre un par de imágenes que describen un modelo tridimensional.

Existen otros tipos de métricas para cuantificar la semejanza entre vectores de características, como por ejemplo, el coeficiente de Jaccard [28], distancia de Spearman [29], distancia de Cayley, separación angular, distancia de Hamming [30], entre otras. Sin embargo, las métricas explicadas en esta sección son las empleadas en el presente trabajo por ser las sugeridas en la literatura relacionada con cada uno de los descriptores.

2 EXTRACCIÓN DE DESCRIPTORES DE FORMA

Este capítulo presenta de manera detallada la forma en que se calculan los tres descriptores empleados, (DESIRE, Spherical Spin Image y Cone Curvature), se presentan imágenes de los resultados obtenidos extrayéndolos a los modelos de la base de datos utilizada; se realiza además un análisis del costo computacional de cada descriptor teniendo en cuenta las operaciones matemáticas básicas realizadas y su dependencia de parámetros variables.

El descriptor DESIRE es desarrollado en [31], allí, Vranic muestra que este descriptor tiene un alto desempeño en la búsqueda de objetos 3D no sólo en su efectividad, sino en el tiempo promedio de cálculo. En este trabajo se quizo verificar el comportamiento de este descriptor aplicándolo a rostros humanos y más específicamente su desempeño en el reconocimiento de expresiones faciales. El descriptor Spherical Spin Image, del cual se han realizado estudios frente al reconocimiento facial [36], es una variante del descriptor Spin Image, de este último se han realizado estudios en cuanto a su capacidad de recuperación en la búsqueda de modelos 3D. El Spherical Spin Image se implementó en este trabajo con miras a establecer su comportamiento respecto al reconocimiento de gestos faciales. Dados los resultados presentados en [38], del descriptor Cone curvature, frente al reconocimiento de objetos 3D, se quizo en este trabajo realizar su implementación con el fin de verificar su comportamiento respecto al reconocimiento de gestos faciales, y realizar la comparación con los otros descriptores seleccionados.

La implementación de los descriptores empleados, se realizó en lenguaje C ++ empleando la librería de visión por computador OpenCV la cual puede ser usada libremente para propósitos comerciales y de investigación.

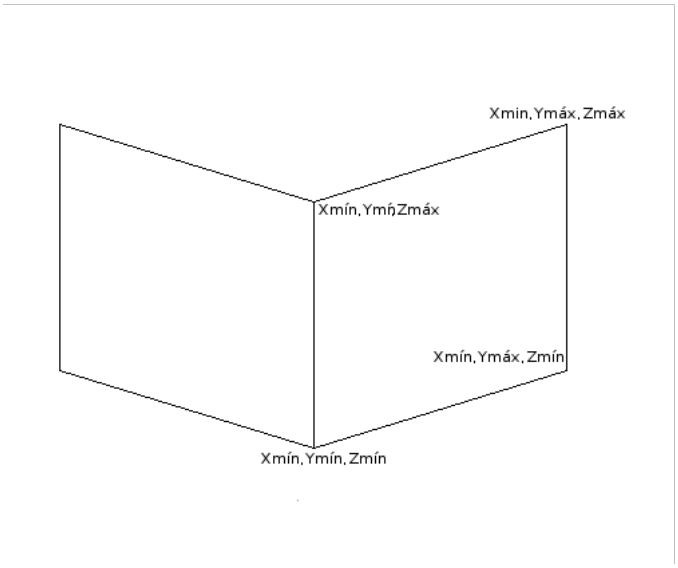
2.1. DESIRE (DEpth SIlhouett Ray-Extent)

El DESIRE [31] , es un descriptor compuesto, el cual está formado por imágenes de profundidad (**de**pth buffer), **s**iluetas y extensiones de rayos (**r**ay-**e**xtents) de una malla poligonal. El vector de características (DE) describe la distancia del objeto desde una cara de un cubo canónico, midiendo la distancia entre las direcciones que son perpendiculares a la cara del cubo. El vector de siluetas caracteriza los puntos de contorno de proyecciones ortogonales

del modelo sobre un cubo de acotamiento. El vector (RE), proporciona información acerca de la extensión del objeto desde el centro de gravedad a lo largo de las direcciones radiales.

2.1.1. Depth Buffer-Based Feature Vector

El vector Depth Buffer [32] se obtiene a partir de imágenes 2D, creadas de objetos 3D para obtener algunas características de forma. Son seis imágenes obtenidas como una medida de la profundidad de los puntos del mallado, a cada una de las caras de un paralelepípedo que encierra el objeto. En principio, todos los valores de cada una de las imágenes son puestos a cero; el mayor valor que puede tomar un píxel de la imagen es 1, indicando la máxima cercanía del punto con la cara del paralelepípedo.
Cada cara está determinada por los correspondientes vértices de la región del paralelepípedo, (Figura **2-1**).



(a)

Figura 2-1: Definición de cada cara del paralelepípedo

De tal forma que la región del paralelepípedo ρ, está definida por:

$$\rho = \{(x, y, z) | x_{min} \leq x \leq x_{max}, y_{min} \leq y \leq y_{max}, z_{min} \leq z \leq z_{max}\} \tag{2-1}$$

La imagen de profundidad correspondiente por ejemplo a la primera cara de la región se forma de la siguiente manera: cada cara de la región, es subdividida en $N \times N$ rectángulos

(o cuadrados, si ρ es un cubo). De esta forma cada rectángulo o cuadrado será un píxel de una imagen en escala de grises de $N \times N$.
Inicialmente, cada píxel (posición (a, b)) es puesto a cero ($v_{ab} = 0$). Luego un punto $P = (x_P, y_P, z_P)$, perteneciente al modelo 3D, es proyectado sobre la cara perpendicularmente, la proyección P' está determinada por las coordenadas (x_{min}, y_P, z_P) si $P' \in v_{ab}$, entonces se actualiza el valor de v_{ab}.

$$v_{ab} \leftarrow max \left\{ v_{ab}, \frac{x_{max} - xp}{x_{max} - x_{min}} \right\} \tag{2-2}$$

Es de notar, que $v_{ab} \in [0, 1]$.
Existen varias posibilidades para definir la región del paralelepípedo ρ que encierra al objeto 3D, en el presente trabajo se consideró el *canonical bounding cube* (CBC) el cual consiste en un cubo cuyos vértices están dados por $\{(x, y, z)|x, y, z \in \{-a_{max}, a_{max}\}\}$, donde

$$a_{max} = max\{max|x_i|, max|y_i|, max|z_i|\}. \tag{2-3}$$

De esta forma la región que encierra el objeto 3D, es un cubo perfecto, y cada cara es subdividida en $N \times N$ cuadrados.

El Algoritmo 1 resume el proceso de construción de las imágenes de profundidad, y en la Figura **2-2** se ilustra el proceso de construcción de las imágenes depth, mientras que en la Figura **2-3** se visualizan los resultados para las 6 caras del modelo 3D mostrado en la Figura 2.2(a).

De acuerdo a la explicado en la Sección 1.4 es necesario aplicar una transformación al dominio frecuencial, a las imágenes obtenidas. La Figura **2-4** muestra un ejemplo de imagen de profundidad y su transformada, la cual se obtiene aplicando la transformada rápida de Fourier, y realizando un desplazamiento para que toda la concentración de la energía, y por ende la mayor información de la imagen quede concentrada en el centro, de esta manera debido a la simetría presente en la transformada, en [32] se propone la expresión 2-4, para tomar los coeficientes de Fourier que conformarán el vector de características final:

$$|p - \frac{N}{2}| + |q - \frac{N}{2}| \leq k \leq \frac{N}{2}, \tag{2-4}$$

en la que $0 \leq p, q \leq N - 1$, son los índices de la imagen transformada, N es el tamaño de la misma, es decir que se tomarán los coeficientes úbicados en (p, q) cuyos valores de p y q satisfagan la expresión 2-4.

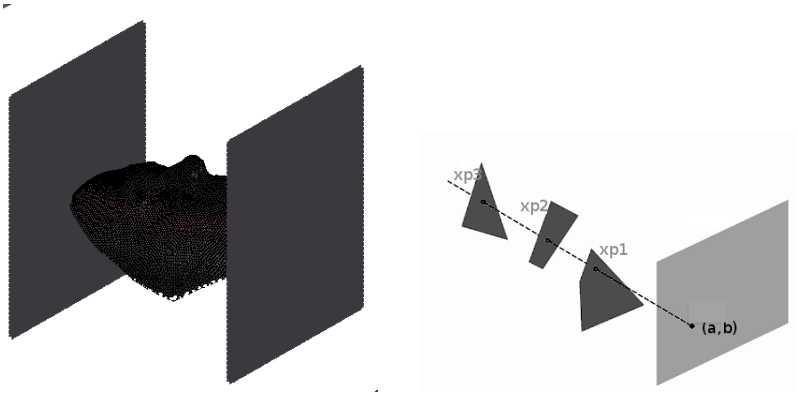
El número de características final está dado por el valor de k de acuerdo a: $k^2 + k + 1$, es decir que para las 6 imágenes de profundidad, la dimensión del vector final es $6(k^2 + k + 1)$.

Algoritmo 1 Depth buffer image

Entrada: Coordenadas de la cara del cubo, $C = \{X_{min}, X_{max}, Y_{min}...\}$, modelo 3D $S = \{V, F\}, V = \{v_i\} \in R^3$

Salida: Imagen de profundidad

1: $DistanciaX \leftarrow X_{max} - X_{min}$
2: $DistanciaY \leftarrow Y_{max} - Y_{min}$
3: $DistanciaZ \leftarrow Z_{max} - Z_{min}$
4: $N \leftarrow$ tamaño deseado de la imagen
5: $dy \leftarrow Y_{max} - Y_{min}/N$
6: $dz \leftarrow Z_{max} - Z_{min}/N$
7: $ya \leftarrow Y_{min} + dy$
8: $zb \leftarrow Z_{min} + dz$
9: Se asignan puntos de la imagen en pasos de dy y dz respectivamente a los vectores pasosY y pasosZ
10: Inicialice la imagen zbuffer de $N \times N$ con ceros
11: Cálculo de las distancias$\leftarrow x_i - x_{min}$
12: **mientras** $k < longituddeV$ **hacer**
13: **para** $i < Tamimagen$ **hacer**
14: **para** $j < Tamimagen$ **hacer**
15: **si** y_i está entre un par de puntos consecutivos de pasosY **entonces**
16: **si** z_i está entre un par de puntos consecutivos de pasosZ **entonces**
17: zbuffer$\leftarrow distancias(k)$
18: **fin si**
19: **fin si**
20: **fin para**
21: **fin para**
22: **fin mientras**
23: **devolver** zbuffer



(a) (b)

Figura 2-2: En la Figura (a) aparecen dos caras opuestas del cubo que encierra el mallado 3D. En (b) distancias de la cara a los puntos proyectados sobre ella.

De acuerdo a los resultados obtenidos y las recomendaciones en [32], en este trabajo se seleccionó $N = 256$ y $k = 5$ con el que se obtiene un vector **DE** de dimensión 186.

2.1.2. Silhouette-Based Feature Vectors

La silueta puede ser definida como los puntos de contorno de un objeto sólido. Para obtener un vector de características basado en siluetas de un modelo 3D, este es proyectado perpendicularmente sobre un cubo que lo encierra, cada cara del cubo es subdividida en $N \times N$ cuadrados, igual que en el caso de Depth Buffer, pero en esta ocasión no interesa la profundidad del objeto con respecto a la cara, de manera que si la proyección sobre la cara pertenece a un píxel, su valor es puesto a cero, los píxeles donde no hay proyección del objeto 3D sobre la cara tienen valor 1; de esta forma se obtienen tres imágenes monocromáticas como se muestra en la Figura **2-5**, el cubo se realiza usando nuevamente CBC.
Una vez tenemos las siluetas, debemos extraer los puntos de contorno. Considerando la imagen como una matriz C de $N \times N$, en la que $C = [c_{ij}]$, donde los elementos c_{ij} son los valores de cada píxel de la imagen, un valor en c_{ij} de 1 (blanco) indica el fondo de la imagen, mientras que un valor de 0 (negro) indica un punto perteneciente a la silueta, un punto de contorno es aquel en el que su valor es 0 y almenos 1 de sus vecinos pertenece al fondo de la imagen.

Con el objetivo de construir el vector de características final, es necesario seleccionar algunos puntos del contorno para definir una secuencia S_K definida por:

$$S_K = \{S_0, S_1, ..., S_{K-1}\}, \tag{2-5}$$

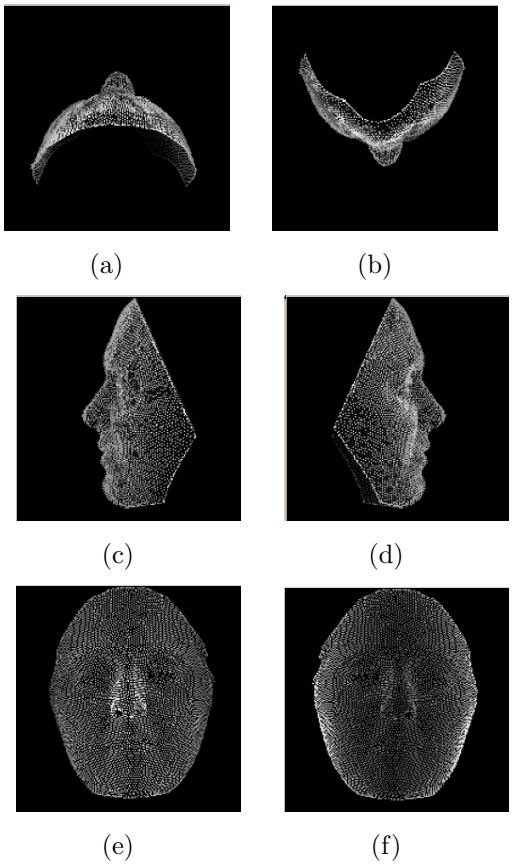


Figura 2-3: Resultados de las imágenes de depth buffer para un modelo de rostro 3D.

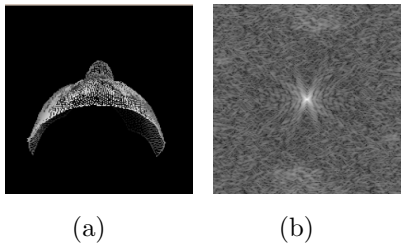


Figura 2-4: Imagen de profundidad y su transformada de Fourier.

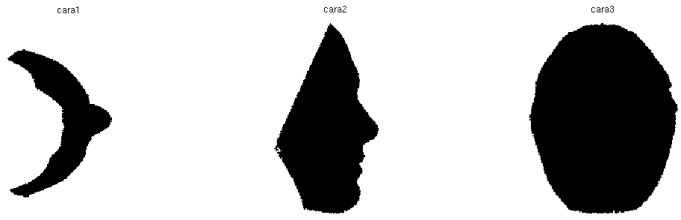


Figura 2-5: Proyección del modelo sobre 3 caras de un cubo.

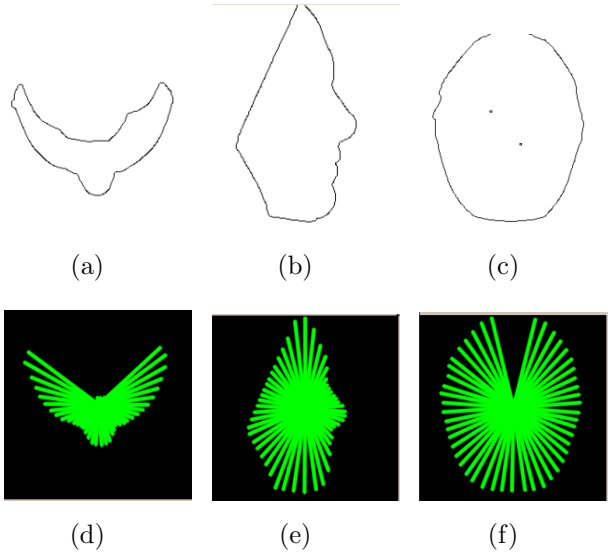


Figura 2-6: Puntos de contorno de la silueta y puntos seleccionados para el vector de características.

el método para seleccionar la secuencia S_K, se decribe a continuación:

- Si $\rho(c_p)$, $\varphi(c_p) \geqslant 0$, $-\pi < \varphi(c_p) \leq \pi$, son las coordenadas del punto de contorno $c_p = (i_p, j_p)$ en el sistema de coordenadas polares con origen O, es decir:
  $c_p = (i_p, j_p) = 0 + \rho(c_p)(cos\varphi(c_p), sen\varphi(c_p))$.

- Entonces, los puntos S_i definidos en 2-5 están dados por:

$$S_i = \left\{ \frac{a_{max}}{N}(c_p - O), \rho(c_p) = max\psi_i, \psi_i = \{\rho(c_p) | \varphi(c_q) \approx \frac{2_{\pi i}}{K}\} \right\}, \tag{2-6}$$

  donde a_{max} está definido en 2-3.

El objetivo es intersectar los puntos del contorno con los rayos que parten desde el origen O (centro de la imagen) y viajan en dirección $(cos(\frac{2i\pi}{K}), sen(\frac{2i\pi}{K}))$, si la interesección existe, entonces el punto más alejado del contorno se toma para hacer parte de S_i, de lo contrario $S_i = 0$.
Del valor de K, depende la dimensión del vector **SI**, de nuevo con base en los resultados de investigación previos, se seleccionó en este trabajo $K = 50$. que será la dimension para cada una de las siluetas.
En la Figura **2-6** se ilustra en la primera fila los puntos de contorno del modelo 3D de un rostro; y en la segunda fila los rayos que parten del centro de la imagen para intersectar los puntos de contorno seleccionados.

Los puntos S_i representan las características basadas en las siluetas en el dominio espacial, sin embargo igual que en la sección anterior para las imágenes Depth Buffer, es conveniente transformar esta información al dominio frecuencial, puesto que una medida de similitud con distancias entre vectores en el dominio espacial no es muy efectiva [32]. El Algoritmo 2, resume el proceso de construcción de las siluetas.

Algoritmo 2 Silhouette-Based Feature Vectors

Entrada: Coordenadas de la cara del cubo, $C = \{X_{min}, X_{max}, Y_{min}...\}$, modelo 3D $S = \{V, F\}, V = \{v_i\} \in R^3$

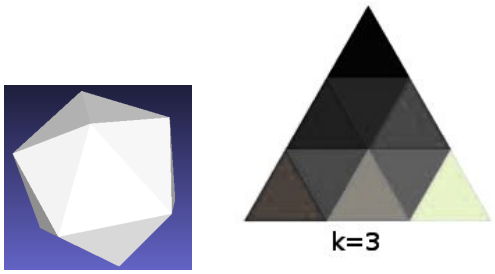
Salida: Silueta de la cara seleccionada

$DistanciaX \leftarrow X_{max} - X_{min}$
$N \leftarrow$ tamaño deseado de la imagen
$dy \leftarrow Y_{max} - Y_{min}/N$
$dz \leftarrow Z_{max} - Z_{min}/N$
$ya \leftarrow Y_{min} + dy$
$zb \leftarrow Z_{min} + dz$
Se asignan puntos de la imagen en pasos de dy y dz respectivamente a los vectores pasosY y pasosZ
mientras $k < longituddeV$ **hacer**
  para $i < tamImagen$ **hacer**
    para $j < tamImagen$ **hacer**
      si y_i está entre un par de puntos consecutivos de pasosY **entonces**
        si z_i está entre un par de puntos consecutivos de pasosZ **entonces**
          silueta$\leftarrow 0$
        fin si
      fin si
    fin para
  fin para
fin mientras
Aplicar detección de bordes a la silueta
devolver silueta

2.1.3. Ray-Based Feature Vector

Este vector de características [33] se obtiene a partir de la medida de la extensión del objeto 3D en direcciones dadas a lo largo de rayos definidos.

Teniendo un conjunto de N vectores unitarios $\mathbf{u}_i$, intersectamos la malla triangular (modelo 3D) con rayos que parten desde el centro de masa del objeto y viajan en dirección de $\mathbf{u}_i$;



(a) (b)

Figura 2-7: Icosaedro y cara subdividida en k triángulos

el vector de características estará conformado por las distancias r_i al punto de intersección más alejado; en caso de no haber intersección, $r_i = 0$. La dimensión del vector será N.

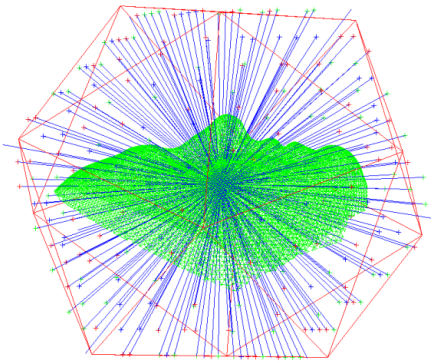
Como la idea es enviar rayos en todas las direcciones del modelo a partir de su centro de masa, pensamos en una esfera unitaria S^2, con centro en el centro de masa del objeto; de manera que los vectores $\mathbf{u}_i$ sean considerados puntos sobre la esfera. De esta forma la medida de los r_i, estará dada por la función:

$$r_{\mathbf{u}} = max\{r \geqslant 0 | r_{\mathbf{u}} \in I \cup \{O\}\}, \tag{2-7}$$

donde I representa el modelo 3D, es decir, que el vector de características $\mathbf{f}$ está dado por $\mathbf{f} = (r_1, ..., r_N)$
Es evidente que no podemos tomar los infinitos puntos sobre la esfera para construir el vector $\mathbf{u}_i$, por tanto hacemos uso del procedimiento descrito en [34], en el cual tomamos un icosaedro cuyas caras subdividimos en k^2 triángulos ($k = 1, 2, ...$), la Figura **2-7** ilustra esto. Luego de la subdivisión tenemos $20k^2$ triángulos con $10k^2 + 2$ vértices $\mathbf{v}_i$, los cuales conforman el vector $\mathbf{u}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$; de aquí se deduce que $N = 10k^2 + 2$.
Un método para encontrar las intersecciones entre los triángulos del mallado que conforman el modelo y los rayos en dirección de $\mathbf{u}_i$, se describe en [35]; sin embargo antes de aplicar este proceso conviene aplicar los pasos considerados en [32], cuyo objetivo es encontrar el conjunto de rayos candidatos a intersectar algún triángulo del mallado. En primer lugar los vectores unitarios direccionales:

$$\mathbf{u}_i = \mathbf{u}_i(\varphi_i, \theta_i) = (cos\varphi_i sen\theta_i, sen\varphi_i cos\theta_i), i = 1, ..., N$$
$$-\pi < \varphi_i \leq \pi, 0 < \theta_i \leq \pi,$$

son almacenados en orden creciente de φ_i y θ_i, es decir,



(a)

Figura 2-8: Rayos intersectando los puntos del mallado

$\mathbf{u}_i(\varphi_i, \theta_i) < \mathbf{u}_i(\varphi_j, \theta_j) \Longleftrightarrow \varphi_i < \varphi_j \vee (\varphi_i = \varphi_j \wedge \theta_i < \theta_j)$.
Luego, para cada triángulo T, perteneciente al mallado determinamos $\varphi_{min}, \phi_{max}, \theta_{min}$ y θ_{max}, siendo estos los ángulos en coordenadas polares de un triángulo del mallado.
Entonces la condición necesaria para que el rayo $\mathbf{u}_i(\varphi_i, \theta_i)$ pueda interseactar a T es:

$$\varphi_{min} \leq \varphi_i \leq \varphi_{max} \wedge \theta_{min} \leq \theta_i \leq \theta_{max}. \tag{2-8}$$

La Figura **2-8**, muestra los rayos que parten del centro de masa del modelo 3D, en dirección de los vértices seleccionados.
El vector de características final del descriptor **DESIRE** se conforma concatenando los vectores de características resultantes de **De**pth, **Si**luetas y **R**ay-**E**xtended $\mathbf{c} = (\mathbf{d}|\mathbf{s}|\mathbf{r})$
de tal forma que la dimension final del vector es $C = D + S + R$, donde, D=Dimensión del vector de características Depth, S= Dimensión del vector de características Si, y R= dimensión del vector RE.

2.2. Spherical Spin Image

El descriptor Spherical Spin Image (SSI) [36], comprende una serie de imágenes descriptivas asociadas con los puntos de orientación en la superficie de un objeto. Estas imágenes se crean mediante la construcción de un sistema de coordenadas 2D sobre un punto orientado (punto 3D con vector normal n). Una Imagen Spin de un punto p es un histograma 2D en el que cada píxel es un bin que almacena el número de vecinos que están a una distancia α a partir de n y una profundidad β de su plano tangente p.
El descriptor SSI se diferencia del Spin Image (SI) [37] en dos aspectos, el SSI utiliza los puntos que están dentro de una esfera de radio dado para construir el histograma, mientras que el SI utiliza todos los puntos sobre la superficie. Por otro lado, el SSI usa la distancia desde otros puntos al punto p, en lugar de la distancia de otros puntos a la normal del punto

p, como el Spin Image.

Teniendo un punto $\mathbf{p}$, perteneciente a I (modelo 3D), con normal $\mathbf{n}$ y plano tangente TP_p, cada uno de los demás puntos $\mathbf{p_k}$ del modelo se relaciona con el punto $\mathbf{p}$ a través de dos parámetros:

1. Distancia entre los puntos $\alpha = \|\mathbf{p} - \mathbf{p_k}\|$.

2. Distancia desde $\mathbf{p_k}$ al plano tangente $\beta = \mathbf{n} \cdot (\mathbf{p_k} - \mathbf{p})$

La Figura 2.9(a) ilustra estas distancias.
Cada Spherical Spin Image (**SSI**) se obtiene de la siguiente manera:

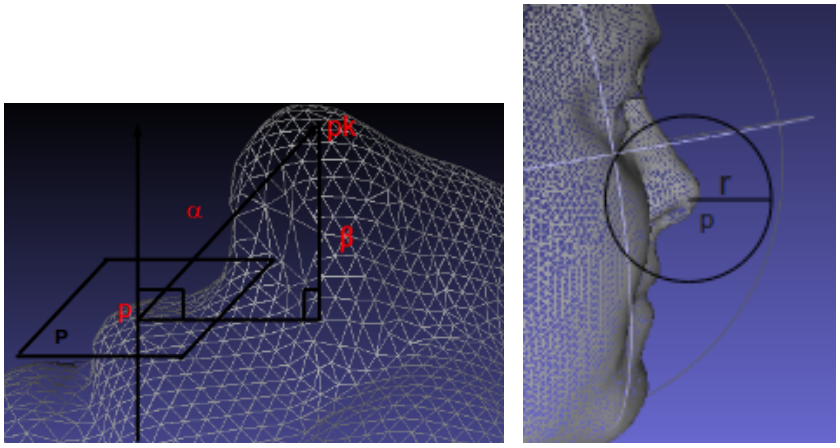
- Para cada punto orientado $\mathbf{p}$ perteneciente a I, una esfera con radio r, se centra en $\mathbf{p}$.
- El **SSI** de $\mathbf{p}$ es un histograma de α y β formado por los puntos que pertenecen al modelo y además están dentro de la esfera.
- Los valores de α y β, son mapeados en un sistema de coordenadas de dos dimensiones donde el eje horizontal corresponde a α, y el vertical a β, el sistema de coordenadas representa el SSI del punto seleccionado, cuando todos los α y β han sido calculados y mapeados.
- En este punto el bin más cercano se incrementará en uno.
- Para encontrar el bin a incrementar se emplea la Ecuación 2-9.

$$i = \lfloor \frac{\frac{W}{2} - \beta}{b} \rfloor, j = \lfloor \frac{\alpha}{b} \rfloor \tag{2-9}$$

2.2.1. Cálculo de las normales

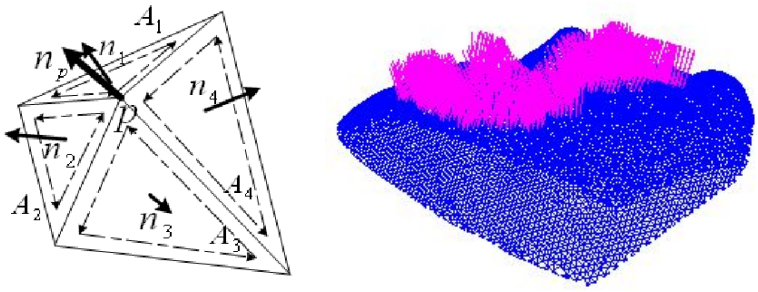
Dado que los modelos empleados para el desarrollo del presente trabajo están conformados por mallas triangulares, se presenta el procedimiento para el cálculo de las normales en este tipo de mallado.
La normal de un vértice en una malla triangular, puede ser calculada realizando un promedio ponderado dependiente de las caras adyacentes al vértice, según la Ecuación 2-10

$$n_p = \frac{n_1 A_1 + n_2 A_2 + n_3 A_3 + n_4 A_4}{A_1 + A_2 + A_3 + A_4} \tag{2-10}$$



(a) (b)

Figura 2-9: En (a) la posición de cada punto $\mathbf{p_k}$, al punto $\mathbf{p}$. (b) Puntos pertenecientes a la esfera de radio r.



(a) (b)

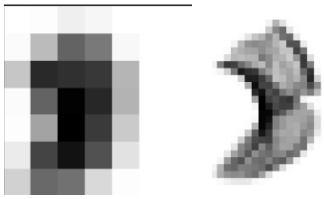
Figura 2-10: En (a) cálculo de la normal del vértice $\mathbf{p}$ [36]. En (b) algunas normales calculadas sobre un modelo de rostro 3D.

Finalmente, según el método empleado en [37], la orientación de todas las normales se determina calculando el producto escalar de la normal de cada vértice y el vector que parte del centroide del objeto al vértice. En la Figura **2-10** se muestra el proceso del cálculo de normales en un vértice, así como el resultado de calcularlas sobre un modelo de la base de datos empleada.

2.2.2. Parámetros del algoritmo

2.2.2.1. Tamaño del bin b

Este parámetro es importante porque determina el tamaño o capacidad de almacenamiento de las SSI generadas. La magnitud de b también afecta la capacidad descriptiva; si el tamaño del bin es muy grande, muchos puntos serán mapeados en el mismo bin, y habrá riesgo de que



(a) (b)

Figura 2-11: En (a) SSI para un $b = 4$ veces la resolución de la malla. Figura (b) SSI con b=resolución de la malla. [10].

exista una correspondencia erronea entre SSI, por otra parte si b es muy pequeño, el proceso de correlacionar dos imágenes no será satisfactorio incluso si las dos imágenes realmente coinciden. De acuerdo a [37], el tamaño del bin debe ser cercano a la resolución de la malla, la cual es definida como la media de las longitudes de todas las aristas del mallado. La Figura **2-11** ilustra las SSI, para dos tamaños de bin.

2.2.2.2. Radio r

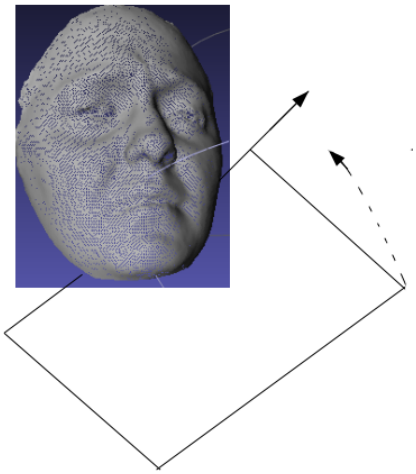
La selección del radio r de la esfera, determina cuantos vértices vecinos del punto orientado, harán parte de la SSI que describirá finalmente al modelo 3D, por tanto no puede ser tan pequeño de forma que no seleccione puntos suficientes, ni demasiado grande, como para seleccionar puntos demasiado alejados que no proporcionarán discriminancia a la imagen final. Según experimentos realizados en [36], se determinó $r = 50\times$ resolución de la malla, como un valor adecuado, y el cual se adoptó en el desarrollo de este trabajo.

2.2.2.3. Ancho de la imagen W

Un SSI, no tiene necesariamente igual número de filas y columnas; sin embargo por simplicidad se acostumbra a que las imágenes sean cuadradas, en la Figura **2-12**, se aprecia cómo se visualiza el ancho de la imagen. La hoja gira alrededor de la normal del punto orientado, y todos los vértices en el barrido de la hoja son asignados al SSI, por esto el nombre de Spherical Spin Image.

Este tamaño se calcula como:

$W = \lfloor \frac{r}{b} \rfloor$.
Como se seleccionó, $r = 50 \times resolucion$, y $b = resolucion$, el valor de W empleado es 50, es decir que cada Spherical Spin Image tiene un tamaño de 50×50 píxeles.



(a)

Figura 2-12: Visualización del ancho de la imagen

2.2.3. Selección de puntos

Calcular las SSI para cada uno de los puntos del mallado, tiene un alto costo computacional; así que para el presente trabajo se seleccionaron algunos puntos sobre los cuales se calculan las SSI; el procedimiento empleado está basado en dos partes:

1. Se extraen las curvaturas mínimas principales k_{min} del modelo; y se seleccionan aquellos puntos que están por debajo de cierto umbral $k_{min} \leq Umbral_c$.

2. Tras el paso anterior se conservan algunos puntos marginales, los cuales son filtrados seleccionando los vecinos de cada punto que están dentro del área definida por un radio determinado $umbral_s$.

En la Figura **2-13**, se muestran los resultados descritos para el cálculo de la SSI de un punto sobre el modelo de un rostro completo, así como para un punto de la región segmentada de la boca para una expresión de sorpresa. El Algoritmo 3, resume el proceso de construcción de las Spherical Spin Image.

2.3. Cone Curvature

2.3.1. Conjunto de Ondas de Modelado (MWS)

Los conjuntos de onda de modelado (MWS) [39], relacionan ampliamente subconjuntos de nodos en una malla. Siendo T, una malla triangular de un objeto, y N un vértice perteneciente a T. Las ondas de modelado (MW) organizan los nodos de la malla en puntos

Algoritmo 3 Spherical Spin Image

Entrada: modelo 3D $S = \{V, F\}, V = \{v_i\} \in R^3$, parámetros (tamaño del bin, radio(r), tamaño imagen), vector normales

Salida: SSIs del modelo

para $indice < Nvertices$ **hacer**
  $ssi \leftarrow 0$
  $punto \leftarrow v(indice)$
  $normal \leftarrow normales(indice)$
  para $indicador < Nvertices$ **hacer**
    $puntoactual \leftarrow v(indicador)$
    $normalactual \leftarrow normales(indice)$
    si $(puntoactual - punto) > r$ **entonces**
      sale del ciclo y toma otro punto
    fin si
    calculo de α
    calculo de β
    calculo de los parámetros i y j
    $ssi(i, j) \leftarrow ssi(i, j) + 1$
  fin para
  $SphericalSpinImage(indice) \leftarrow ssi$
fin para
devolver SphericalSpinImage

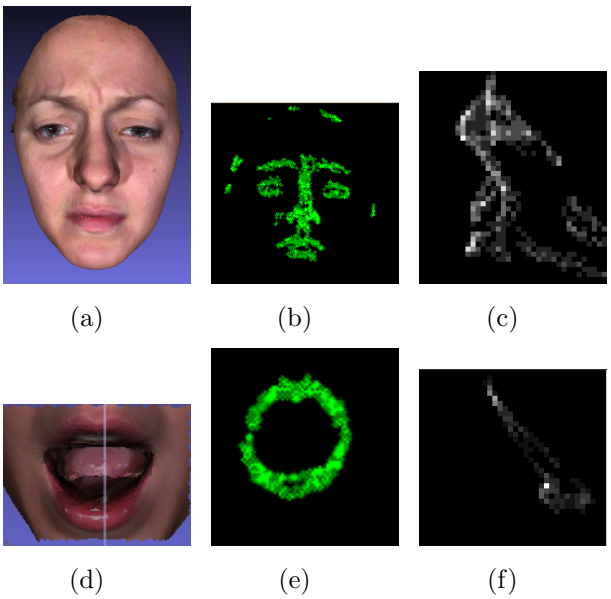


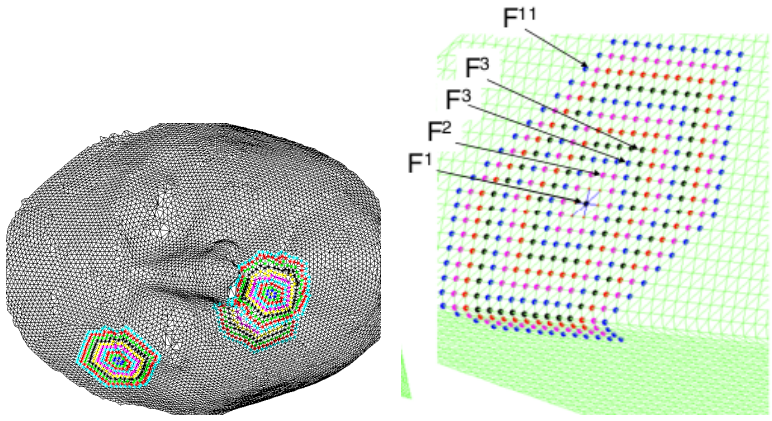
Figura 2-13: En la fila de arriba, aparecen un rostro de la base de datos, los puntos de mayor curvatura, y la SSI de un punto. En la fila de abajo los resultados correspondientes para la región de la boca en una expresión de sorpresa.

concéntricos, espacialmente dispuestos alrededor de N. Cada uno de los grupos concéntricos (Figura 2.14(b)), se llama un *wave front* (F), y el nodo inicial N de cada uno de ellos se conoce como Foco. Entonces todos los posibles MWs que se pueden generar sobre T, se conocen como conjunto de ondas de modelado. Un conjunto de ondas de modelado, se define entonces como: $MWS = \{MW^1, MW^2, ..., MW^q\}$, donde MW^i, es la onda de modelado generada a partir de un foco N, que corresponde con el i-esimo vértice del modelo I.
El cono de curvatura es una representación del modelo 3D, que se calcula a partir del MWs para cada nodo de la malla y cuya definición desarrollada en [38] es la siguiente: se llama Cono Curvatura (CC) j-esima de N (α^j), al ángulo del cono con vértice N cuya superficie está aproximada al j-ésimo frente de ondas, F^j, de la onda de modelado asociada a N. Formalmente:

$$\alpha^j = sign(F^j) \cdot \left| \frac{\pi}{2} - \frac{1}{t_j} \sum_{i=1}^{t_j} \gamma_i^j \right|. \tag{2-11}$$

Donde, $\gamma_i^j = \sphericalangle C^j N N_i \in F^j, t_j$ es el número de nodos de F^j y C^j es el baricentro de F^j.

El rango de valores de la CC es $[-\pi/2, \pi/2]$, donde el signo toma en cuenta la localización relativa de O, C^j y N, siendo O el origen de coordenadas del sistema de referencias fijo a T. Un signo negativo, indica zonas cóncavas, valores cercanos a cero se corresponden a superficies



(a) (b)

Figura 2-14: WF en diferentes focos sobre el rostro.

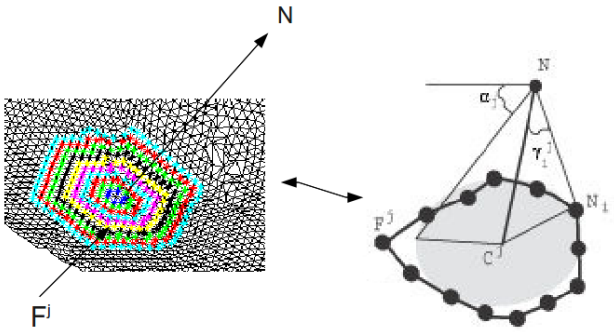


Figura 2-15: Definición del cono de curvatura. [38]

planas, y valores positivos son zonas convexas. La Figura **2-15**, ilustra la definición de la CC.

Dado un foco N, existe un conjunto q de frentes de onda que determinan las CC's para N, $\{\alpha^1, \alpha^2, ..., \alpha^q\}$, las cuales proporcionan una completa información acerca de la curvatura del objeto desde el punto de vista de N.

El Algoritmo 4 ilustra los pasos para obtener las CC, previo cálculo del WM, y los baricentros de cada uno de los frentes de onda.

El resultado luego de calcular el CC a un modelo 3D, es una matriz de $h \times q$, donde h, corresponde al número de vértices pertenecientes al modelo, y q el número de frentes de onda considerado para cada foco; en [38], se demuestra que un valor de q mayor, tiene mejores resultados, allí se experimento con niveles de q =2, 6, 8, 14 y 16. Considerando el

tiempo de cálculo del descriptor y teniendo en cuenta que en promedio, el número máximo podría ser de 15 para la región de la boca (Figura **2-16**), en el presente trabajo se consideró $q = 10$.

Una característica de las CC's es que los valores de estas se encuentran lo suficientemente alejadas para focos diferentes, esta característica se ilustra en la Figura **2-17**. Además las CC son invariantes a la traslación, rotacion y escala. El Algoritmo 4, resume el proceso de construcción de las Cone curvature.

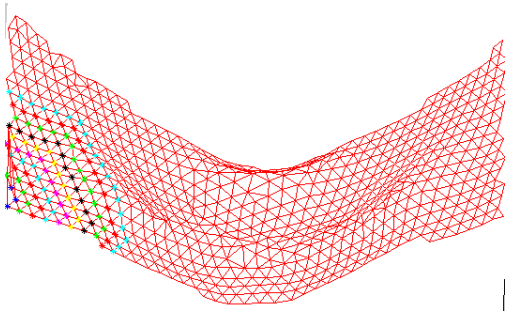


Figura 2-16: 10 frentes de onda en la región de la boca. Se podrían tener un máximo promedio de 15 frentes de onda

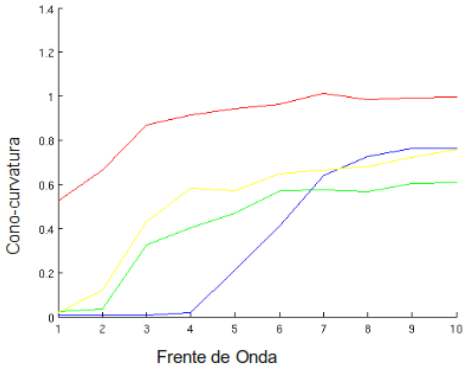


Figura 2-17: Representación de los CC, de 10 frentes de onda, para 4 focos distintos de un mismo modelo

2.4. Costo computacional de descriptores de forma

La estimación del costo computacional es un aspecto importante a tener en cuenta si se quieren implementar los descriptores de forma en un sistema de reconocimiemto en tiempo real; realizando un análisis similar al presentado en [44], se hace una estimación según el número aproximado de cálculos requeridos para obtener los descriptores implementados.

Algoritmo 4 Cone Curvature

Entrada: foco N, arreglo de baricentros, wf (coordenadas de los puntos de cada frente de onda)

para $indice < cantidaddewf$ **hacer**
  $baricentro \leftarrow arreglobaricentros(indice)$
  Cálculo de parámetro γ
  distanciabaricentro← cálculo de la distancia entre baricentro y origen γ
  distanciaFoco← cálculo de la distancia entre N y el origen γ
  si $distanciabaricentro > distanciafoco$ **entonces**
    $signo \leftarrow -1$
  si no
    $signo \leftarrow 1$
  fin si
  cálculo de $\alpha(indice)$
fin para

El costo computacional se presentará en términos del tiempo necesario para efectuar las siguientes operaciones básicas: suma (ts), resta (tr), multiplicacion (tm) y división (td) en cada uno de los descriptores implementados.

2.4.1. Costo computacional de DESIRE

Para el descriptor DESIRE, es necesario evaluar los tiempos de cálculo de cada uno de sus componentes; para el caso de las imágenes de profundidad (**Depth**), de acuerdo al Algoritmo 1, y teniendo encuenta las operaciones básicas, podemos obtener una expresión para calcular el número total de operaciones de la siguiente manera:

$$T_1 = 6(5tr + td + 5ts + 2tm).$$

Esta expresión es independiente de variables como el número de vértices, o el tamaño de la imagen de profundidad.
En cuanto a la asignación de las distancias para conformar las imágenes de profundidad, se puede aproximar las siguiente expresión:

$$T_2 = N * C * (\sum_{i=1}^{Tam} i).$$

La cual se puede expresar como:

$$T_2 = N * C(\frac{Tam^2 + Tam}{2}),$$

donde Tam, corresponde al tamaño de la imagen de profundidad, N es el número de vértices del mallado y C, la constante de tiempo de la operación de asignación de variables.
De esta forma el algoritmo para (**Depth**) tiene complejidad $O(n^2)$ dependiente del tamaño de la imagen, pero líneal con respecto al número de vértices.

De acuerdo al Algoritmo 2, para el proceso del cálculo de las Siluetas, y teniendo encuenta sólo las líneas que tienen dependencia de alguna variable, se aproxima la siguiente expresión:

$T = N * C(\frac{Tam^2+Tam}{2})$,

que tiene la misma complejidad análizada para **DE**.
En cuanto al algoritmo de RE (Ray Extendet based), es necesario realizar varios cálculos para subdividir las aristas del icosaedro con objetivo de encontrar los puntos de intersección con las M caras del mallado, es la componente de mayor costo computacional del descriptor DESIRE; realizando el análisis de las operaciones necesarias, se obtuvo:

$costoc_{RE} = 126 \times ts + 20(14 \times tr + 15 \times ts + 13 \times td) + 12(3 \times tm + traiz) + 240 \times tm + M(9 \times tm + 3 \times traiz + 60 \times tm) + M \times ptosIcosaedro \times (4 \times p_{cruz} + 4 \times p_{punto})$.
donde, $ptosIcosaedro$, corresponde a la cantidad de vértices en los que puede ser subdividido el icosaedro.
Para el cálculo del número de operaciones realizadas por la raíz cuadrada, se estimó de acuerdo al procedimiento presentado en [42], en donde se tiene que:
$Raiz = Raiz^{-1} + td$.
$Raiz^{-1} = 4 \times tm + 3 \times tr$.

En cuanto al cálculo del producto punto y producto cruz de vectores en tres dimensiones el número de cálculos en las operaciones fundamentales es:

$p_{punto} = 3 \times tm + 2 \times ts$.
$p_{cruz} = 6 \times tm + 3 \times tr$.
De lo anterior notamos que para el caso de DE y SI, la complejidad del algoritmo depende en forma cuadrática del tamaño de la imagen, mientras que el RE, tiene una dependencia lineal del producto de la cantidad de caras del mallado y el número de vértices considerados sobre el icosaedro base.
Para tener una idea más clara acerca de lo que esto significa, calculando un ejemplo con un tamaño de imagen=256, M=1800 y puntos del icosaedro= 92, que corresponden a valores promedios considerados en este trabajo, obtenemos las expresiones siguientes:

$T_1 = C(\frac{Tam^2+Tam}{2}) = 32896 \times C$, $T_2 = M \times ptosIcosaedro = 165600 \times C$,
Siendo C, una constante que representa el tiempo en realizar las operaciones matemáticas

básicas.
De aquí observamos que el mayor costo computacional en cuanto a tiempo de proceso del descriptor DESIRE, es debido al componente RE.

2.4.2. Costo computacional de Spherical Spin Image

De acuerdo al tamaño de la imagen a generar, y teniendo en cuenta el procedimiento descrito en la Sección 2.2, se deduce una expresión para el tiempo total de cálculo del descriptor SSI, considerando los tiempos de proceso de cada una de las operaciones básicas, de la siguiente manera:

$cc_{SSI} = (\sum_{i=1}^{N} i) \times (3 \times Tam \times tm + 6 \times ts + 5 \times tr + 3 \times td + tm) = (\frac{N^2+N}{2}) \times (3 \times Tam \times tm + 6 \times ts + 5 \times tr + 3 \times td + tm)$
De esta expresión vemos que el algoritmo tiene una complejidad $O(n^2)$, dependiente del número de vértices del mallado.

2.4.3. Costo computacional de Cone Curvature

El Cone curvature, además del proceso de normalización del modelo, requiere el cálculo de los MW set para cada uno de los puntos del mallado, lo cual incrementa el costo computacional. Un análisis aproximado de las operaciones necesarias para calcular este descriptor se resume en la siguiente expresión:

$costoc_{CC} = 3 \times (\frac{N^2+N}{2}) \times wf(ts + td) + wf(wf(3 \times tr + 3 \times tm + raiz) + 60 \times tm + ts + tr) + tr + td + 2tr + 4 \times tm + raiz$
Considerando sólo las operaciones básicas, vemos que en este algoritmo, existe una dependencia cuadrática del número de vértices, por lo que su complejidad es $O(n^2)$, además depende también del número de wf consideradas. Teniendo en cuenta el tiempo para calcular las posiciones de las MW set, este es el descriptor con mayor tiempo de ejecución.

Como se mencionó al comienzo de este capítulo, los descriptores fueron implementados en lenguaje C++, utilizando la librería OpenCV; se aprovechó además el software Meshlab [41], a través del cual fue posible "cargar" los modelos de la base de datos y obtener la información de los vértices y triángulos del mallado, para su posterior procesamiento y visualización. Gracias a este software, no es necesario preocuparse por el formato del mallado, ya que este "acepta" varios formatos. Sin embargo, es importante conocer la arquitectura del software para conocer con exactitud como queda almacenada la información, en que tipo de estructura y como acceder a ella, lo cual requiere un tiempo de estudio que debe sumarse al tiempo de implementación de los descriptores.

3 ANALISIS DE SIMILITUD

En este capítulo, se aplicarán las métricas de similaridad descritas en la Sección 1.4 para los descriptores implementados, como herramienta de búsqueda de una imagen que actua como query; lo que se interpreta finalmente como una clasificación de los modelos utilizando la estrategia de los vecinos más cercanos. Como herramienta de evaluación, se emplean las curvas de precision-recall, las cuales proporcionan información acerca de la capacidad de encontrar objetos relevantes de cada uno de los descriptores.

3.1. Base de Datos

La base de datos empleada para el desarrollo de este trabajo es la BU-3DFE (Binghamton University 3D Facial Expression) [40] la cual está disponible con fines de investigación y contiene 100 imágenes de personas (56 mujeres, 44 hombres), que van desde los 18 hasta 70 años de edad, con una variedad de grupos étnicos/raciales, incluyendo Blanco, Negro, Asiáticos, Indio y Latinos hispanos.

Cada persona realizó 6 expresiones; alegría (HA), disgusto (DI), miedo (FE), enojo (AN), sorpresa (SU) y tristeza (SA); frente a un escáner 3D, estas imágenes tienen formato vrml, y están constituidas por mallas triangulares. La Figura **3-1**, muestra algunos modelos de la base de datos.

3.2. Análisis DESIRE

Con el objetivo de encontrar las distancias entre vectores, y de esta manera verificar similitudes entre gestos, se extrajo el descriptor a un conjunto de imágenes de la base de datos para cada gesto, se calculó la distancia entre una imagen de cada gesto y 360 más, de las cuales 60 corresponde a la misma clase de búsqueda, de esta manera, en el caso ideal se espera que para cada una de las imágenes buscadas, las distancias a las 60 imágenes del gesto correspondiente sean las menores.

Con el objeto de probar el descriptor DESIRE, se utilizó como métrica de similaridad la norma l_∞, que se define en la Ecuación 3-1, para determinar la distancia entre dos vectores.

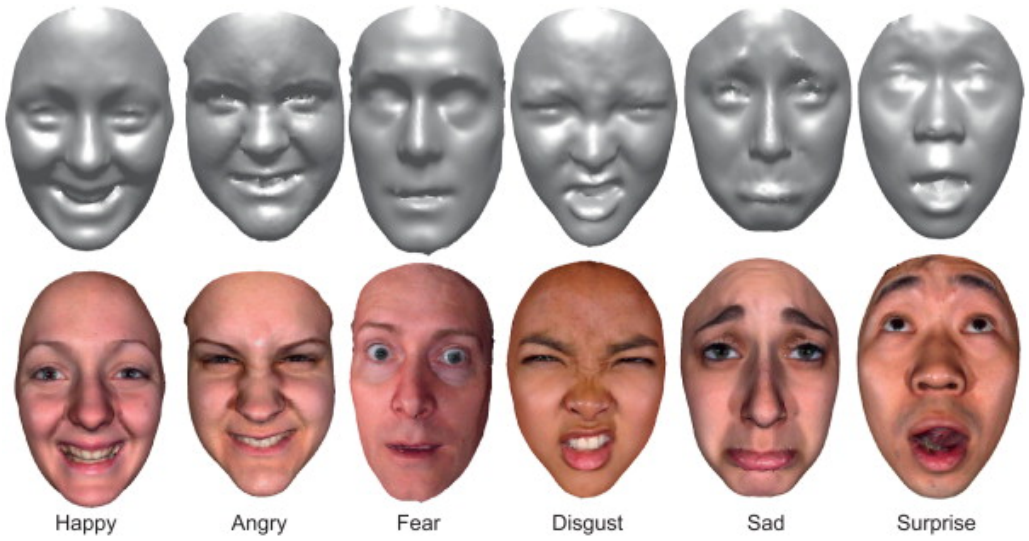


Figura 3-1: Ejemplos de modelos pertenecientes a la base de datos empleada.[40]

$$l_\infty = (\mathbf{f}', \mathbf{f}'') = \|\mathbf{f}' - \mathbf{f}''\|_\infty = \left(\begin{smallmatrix} max \\ 1 \leq i \leq N \end{smallmatrix}\right) |f_i' - f_i''|. \tag{3-1}$$

Recordemos que $\mathbf{f}', \mathbf{f}''$ son los vectores de características de dos modelos diferentes y N, es la magnitud de los mismos.
En la Tabla **3-1**, se muestran las tasas de aciertos en el reconocimiento luego de calcular la l_∞ para la imagen query de cada gesto comparándola con las 60 referencias de la misma clase. Estas se crearon realizando una lista ordenada de los objetos con menor distancia entre vectores. De esta forma, por ejemplo, en la primera fila de la Tabla **3-1**, se indica que de las 60 pruebas realizadas, en 31 casos, las distancias se encontraron en las primeras posiciones, para este primer ejemplo se encontraría una tasa de aciertos del 51,67 %, extrayendo el descriptor sobre todo el rostro.

	Rostro completo		Boca	
Expresión	casos	Porcentaje	casos	Porcentaje
AN	31	51,67 %	24	40 %
DI	6	10 %	17	28,33 %
FE	10	16,67 %	16	26,67 %
HA	18	30 %	23	38,33 %
SA	16	26,67 %	26	43,33 %
SU	36	60 %	39	65 %

Tabla 3-1: Aciertos del descriptor DESIRE.

En la misma Tabla **3-1**, se presentan los resultados de realizar el mismo procedimiento, luego de segmentar la región de la boca (Figura. **3-2**) con ayuda del software Meshlab [41], y extraer los descriptores sobre la misma cantidad de imágenes, en la tabla, se aprecia que a excepción del gesto de enojo (AN), el desempeño del descriptor mejoró levemente, aplicándolo sobre la región de la boca, en cuanto a la cantidad de expresiones reconocidas.

De esta manera, se observa que a pesar de que cada expresión facial se manifiesta en varias regiones del rostro, es posible distinguirlas teniendo en cuenta solamente los cambios presentes en la región de la boca. La Figura **3-3**, muestra algunos ejemplos de la base de datos en los que se aprecia los cambios sobre la boca para cada expresión.

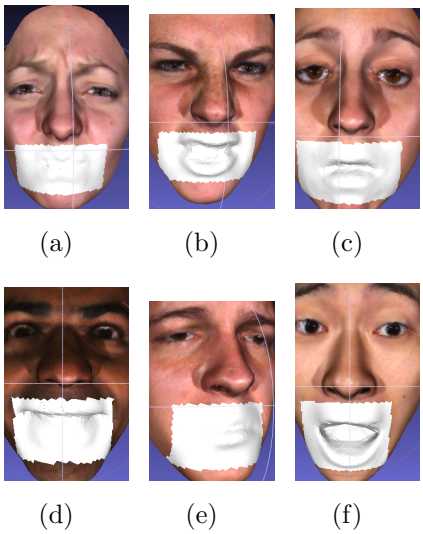


Figura 3-2: Región de la boca segmentada manualmente con meshlab para todos los gestos

Figura 3-3: Diferentes Expresiones de la base de datos. [40]

3.3. Análisis Spherical Spin Image

Para analizar la similitud de las imágenes spin de dos rostros diferentes realizando la misma expresión se utilizó, según se emplea en [37], el cálculo del coeficiente de correlación lineal (Ecuación 1-4); si el valor de r es alto (cercano a 1), las imágenes son similares.

Esta vez se realizó el cálculo del coeficiente de correlación líneal entre todas las Spherical Spin Image de un modelo de cada gesto, y todas las Spherical Spin Image de 60 modelos más. En la Tabla **3-2** se ilustra la tasa de aciertos o coincidencia entre Spherical Spin Image tanto en todo el rostro, como sobre la región de la boca; para esto se analizó, de 60 búsquedas, cuantas tuvieron el valor de r más alto con el gesto correspondiente.

	Rostro completo		Boca	
Expresión	casos	Porcentaje	casos	Porcentaje
AN	12	20 %	20	33,33 %
DI	12	20 %	26	43,33 %
FE	12	20 %	3	5 %
HA	20	33,33 %	10	16,67 %
SA	33	55 %	46	76,67 %
SU	8	20 %	12	33,33 %

Tabla 3-2: Aciertos del descriptor SSI.

Podemos apreciar en la Tabla **3-2** que únicamente para la expresión de miedo (FE) y alegría (HA), hubo mayor cantidad de imágenes coincidentes para el descriptor calculado en el rostro completo, para las demás expresiones se encontraron mejores resultados sobre la región de la boca.

3.4. Análisis Cone Curvature

Para calcular la métrica de similitud se uso nuevamente la distancia l_∞, que determina la separación y por ende la coincidencia entre dos vectores. Sin embargo, luego de calcular el descriptor Cone Curvature, lo que tenemos es una matriz de $h \times q$, donde h es el número de focos considerados; mientras que q corresponde a la cantidad de WF. De esta forma la distancia entre dos modelos T_1 y T_2, según se explica en [38], se define según la Ecuación 3-2:

$$d^j = (T_1, T_2) = \sqrt{\sum_{k=1}^{h} (\theta_{1(k,j)} - \theta_{2(k,j)})^2}. \tag{3-2}$$

Donde, θ_1, θ_2, son las matrices de CC para cada modelo. Finalmente la distancia se calcularía como:

$d(T_1, T_2) = max|d^j(T_1, T_2)|, j = 1, 2, ..., q.$

La Tabla **3-3**, muestra la tasa de modelos con la menor distancia, luego de calcular la métrica de distancia para la región de la boca al descriptor CC sobre 50 modelos.

Debido a la gran cantidad de cálculos que deben ser realizados para obtener este descriptor, el tiempo de cálculo en el rostro completo, teniendo en cuenta la cantidad de vértices (10000 en promedio) es demasiado alto, lo cual lo hace inviable para un sistema de reconocimiento; por este motivo este descriptor se extrajo únicamente sobre la región de la Boca.

	Boca	
Expresión	casos	Porcentaje
AN	18	36 %
DI	4	8 %
FE	8	16 %
HA	10	20 %
SA	10	20 %
SU	30	60 %

Tabla 3-3: Aciertos del descriptor ConeCurvature.

Promediando los porcentajes obtenidos de modelos coincidentes en las tablas, obtenemos un porcentaje del 26.67 %, que ubicaría al descriptor CC en el tercer lugar comparado con el 40.28 % del descriptor DESIRE, y el 34.72 % del SSI realizando la búsqueda de gestos coincidentes haciendo cálculos de distancias entre vectores característicos.

El cálculo de la distancia entre vectores es una forma de clasificar los modelos teniendo en cuenta la cercanía de las características de los modelos de referencia con los de búsqueda. Se realizó un experimento adicional, en el que se calcularon de nuevo las distancias, pero esta vez fueron tomadas las distancias entre el promedio de las características de un conjunto de 60 imágenes y un conjunto de otras 50 imágenes diferentes, utilizando las mismas métricas

	AN	DI	FE	HA	SA	SU
AN	$1{,}11e^{-12}$	$\mathbf{2{,}76e^{-13}}$	$6{,}74e^{-13}$	$5{,}85e^{-13}$	$6{,}43e^{-13}$	$5{,}74e^{-13}$
DI	$\mathbf{5{,}28e^{-13}}$	$7{,}23e^{-13}$	$1{,}25e^{-12}$	$1{,}09e^{-12}$	$7{,}26e^{-13}$	$7{,}984e^{-13}$
FE	$\mathbf{5{,}05e^{-13}}$	$7{,}47e^{-13}$	$1{,}28e^{-12}$	$1{,}11e^{-12}$	$8{,}87e^{-13}$	$8{,}21e^{-13}$
HA	$1{,}21e^{-12}$	$6{,}10e^{-13}$	$\mathbf{5{,}77e^{-13}}$	$6{,}84e^{-13}$	$7{,}41e^{-13}$	$1{,}07e^{-12}$
SA	$9{,}93e^{-13}$	$2{,}99e^{-13}$	$7{,}95e^{-13}$	$6{,}31e^{-13}$	$\mathbf{2{,}62e^{-13}}$	$7{,}36e^{-13}$
SU	$9{,}95e^{-13}$	$9{,}97e^{-13}$	$1{,}02e^{-12}$	$9{,}83e^{-13}$	$\mathbf{8{,}49e^{-13}}$	$1{,}11e^{-12}$

Tabla 3-4: Distancias entre promedios para el descriptor DESIRE sobre el rostro

	AN	DI	FE	HA	SA	SU
AN	68,5	52,06	45,14	33,63	**32,41**	272,29
DI	119,16	**73,07**	90,13	120,57	120,60	262,49
FE	72,28	47,38	41,23	**29,64**	38,91	276,35
HA	129,97	**97,87**	109,16	130,89	130,75	247,68
SA	48,28	62,12	42,58	44,06	**24,54**	274,75
SU	687,40	647,74	654,98	689,34	689,60	**494,96**

Tabla 3-5: Distancias entre promedios para el descriptor DESIRE sobre la región de la boca

anteriores. En las tablas **3-4** y **3-5** se resumen los resultados de las distancias encontradas para cada uno de los descriptores, señalando en negrilla las distancias más cortas.
De las tablas **3-4** y **3-5**, se aprecia que para el primer caso correspondiente a las distancias entre cada uno de los gestos, ninguno coincide con su correspondiente, mientras que en el segundo caso, es decir, para la región de la boca, coinciden teniendo las menores distancias entre sí, las expresiones de DI, SA y SU.

3.5. Evaluación de la efectividad

A continuación se presentan los análisis efectuados con el fin de medir el desempeño de los descriptores respecto a la recuperación de información.

3.5.1. Curvas Precision-Recall

La precisión se define como la fracción de imágenes recuperadas que son relevantes entre el total de imágenes. Recall es la fracción del número de imágenes relevantes que han sido correctamente recuperadas, mide la capacidad de recuperar las imágenes que son relevantes

dentro de todo el conjunto [31].
En general las curvas Precision-Recall miden la efectividad del sistema.

$$precision = \frac{M}{R},\ recall = \frac{M}{q}.$$

R es la cantidad de modelos recuperados, q corresponde al conjunto de objetos relevantes y M son los objetos relevantes recuperados.
Como ejemplo para un query q del descriptor DESIRE, usando la métrica de similaridad empleada en el capítulo anterior, se realiza la búsqueda entre un total de 540 objetos (90 de cada gesto), las distancias calculadas se organizan en un vector, encontrando que los 90 objetos están en las posiciones $1, 2, 3, 4, 5, 9, 16, 22, ..., 489, 504, 518$, con esta información se construye la Tabla **3-6**.

Pos	1	2	3	4	5	9	16	22	...	504	518
Recall	1/90	2/90	3/90	4/90	5/90	6/90	7/90	8/90	...	89/90	90/90
Precision	1/1	2/2	3/3	4/4	5/5	6/9	7/16	8/22	...	89/504	90/518

Tabla 3-6: Ejemplo del calculo de precision y recall

El procedimiento anterior es empleado para un solo query q, para evaluar el descriptor en un conjunto de elementos de búsqueda, se realizan los mismos cálculos para cada elemento y al final se promedian los resultados. Las figuras **3-4** y **3-5**, ilustran los resultados gráficos del cálculo de precision- recall para los descriptores empleados, y su desempeño para cada gesto sobre el rostro completo y la región de la boca, respectivamente.

De acuerdo a la Tabla **3-6**, es evidente que los mejores resultados se obtienen cuando las posiciones encontradas ocupan los primeros lugares, esto quiere decir que el valor de precision es cercano a uno, los resultados mostrados en la Figura **3-4**, demuestran que en general sobre el rostro completo, el descriptor DESIRE tiene un mejor comportamiento recuperando gestos faciales. Salvo las expresiones HA y SA (Figuras 3.5(d) 3.5(e)), en la que se aprecia un mejor efectividad del descriptor SSI, lo cual corresponde con los resultados mostrados en la Tabla **3-2**, donde los mejores porcentajes de aciertos en la búsqueda por coeficiente de correlacion líneal, se presentaron precisamente para estas dos expresiones con el 30 % y el 26.67 %, respectivamente.

De los resultados de la Figura **3-5**, notamos que para los descriptores calculados sobre la región de la boca, excepto en la expresión de tristeza (SA), dónde SSI tiene una evidente mayor efectividad, hay un comportamiento similar entre el CC y el DESIRE, aunque este último presenta un comportamiento levemente superior para todas las expresiones.

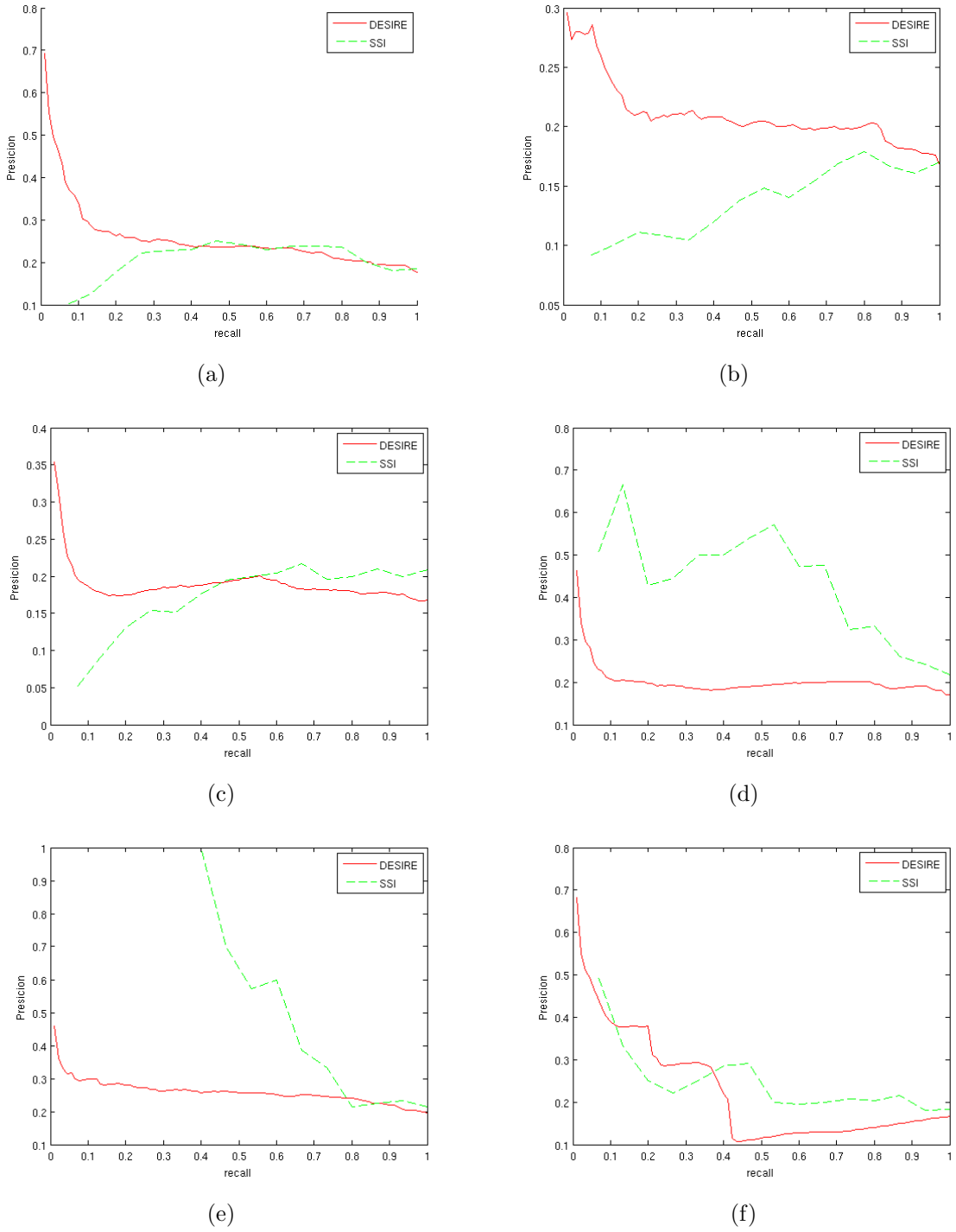


Figura 3-4: Curvas Precision - recall para cada uno de los descriptores en los 6 gestos faciales sobre todo el rostro

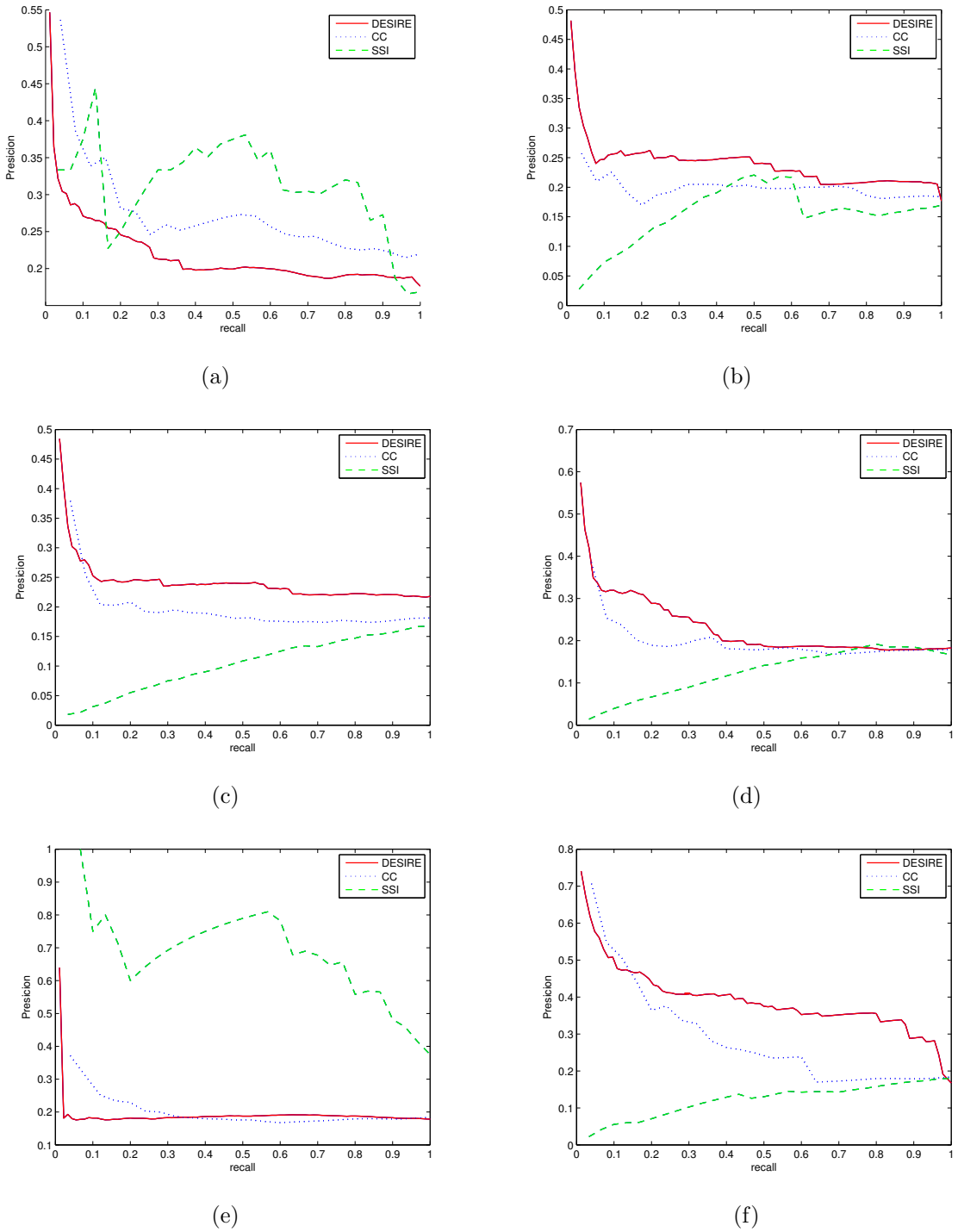


Figura 3-5: Curvas Precision - recall para cada uno de los descriptores en los 6 gestos faciales sobre la región de la boca

3.6. Análisis de similitud con reducción de dimensionalidad

Los tres descriptores implementados tienen en general una alta dimensionalidad, esto dificulta el entrenamiento de algunos clasificadores, y consume gran cantidad de recursos computacionales, en esta sección se emplea el análisis de componentes principales (PCA) para realizar una reducción del número de características producto de la extracción de los tres descriptores, y se practica un análisis similar al efectuado en las secciones anteriores, con el fin de comparar los resultados y evaluar el efecto de aplicar la reducción.

3.6.1. Reducción de dimensionalidad

Una vez son calculados los descriptores, estos generalmente tienen una dimensión bastante grande, para el caso del DESIRE, por ejemplo, considerando la región de la boca, se tiene que la dimensión de **DE** según se comentó en la Sección 2.1 es de $6 \times 31 = 186$; para el SI, de acuerdo a lo descrito en la misma sección se tomó un valor de K=50. Considerando sólo la región de la boca (Figura **3-2**), se analizó que solo son necesarias dos siluetas, debido a que la segmentación manual produce un borde tipo rectangular que es similar para todas las imágenes, y que por lo tanto no producirá información relevante; teniendo en cuenta esto se tiene que la dimensión del vector **SI** es $2 \times 50 = 100$ para la región de la boca, pero de $3 \times 50 = 150$, para todo el rostro.
Respecto a la componente **RE**, se tomó un valor de 92 vértices, para esto se tuvo en cuenta un número suficiente de puntos sobre el icosaedro base, pero que no podía ser demasiado grande, dado el tiempo requerido para el cálculo; de esta manera la dimensión total del vector DESIRE es $186 + 150 + 92 = 428$ para el rostro completo, y $186 + 100 + 92 = 378$ para la región de la boca.

Con el objetivo de realizar más adelante el análisis de discriminancia del descriptor, y entrenar un clasificador, se realiza la reducción de dimensionalidad usando la técnica de componentes principales (PCA), la cual consiste en realizar una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos, en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado el Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente, de esta manera obtenemos un vector de características con un mínimo de redundancia en la información.

Para el caso del descriptor DESIRE se tomaron 10 componentes principales, las cuales explican más del 90 % de la varianza. En la Tabla **3-7**, se muestra el porcentaje de explicación de la varianza para las componentes seleccionadas por cada uno de los gestos.

	AN	DI	FE	HA	SA	SU
componente1	64.64 %	86.62 %	63.17 %	59.28 %	63.26 %	57.80 %
componente2	10.00 %	7.97 %	18.51 %	10.03 %	8.99 %	8.97 %
componente3	6.86 %	1.04 %	4.26 %	5.83 %	5.55 %	6.73 %
componente4	3.24 %	0.64 %	2.37 %	4.55 %	4.48 %	5.16 %
componente5	2.53 %	0.54 %	2.07 %	3.69 %	3.31 %	4.22 %
componente6	2.16 %	0.41 %	1.69 %	2.80 %	2.95 %	3.23 %
componente7	1.98 %	0.36 %	1.53 %	1.60 %	1.95 %	2.55 %
componente8	1.41 %	0.32 %	1.22 %	1.39 %	1.68 %	2.07 %
componente9	1.07 %	0.30 %	0.89 %	1.06 %	0.86 %	1.44 %
componente10	0.79 %	0.21 %	0.61 %	0.96 %	0.64 %	0.99 %
Total	94.69 %	98.41 %	96.30 %	91.18 %	93.68 %	93.16 %

Tabla 3-7: Explicación varianza PCA para el descriptor DESIRE

Respecto al descriptor SSI se construyó un vector con las medias y las varianzas de cada imagen, obteniendo vectores en promedio de dimensión 600, que corresponde al número promedio de vértices sobre el rostro completo, luego de seleccionar los puntos dados por las curvaturas principales, y de 150 correspondiente a la región de la boca, luego de realizar el mismo procedimiento. Aplicando PCA a estos vectores característicos, se tomaron 5 componentes principales, la drástica reducción, se explica con el hecho de que el descriptor SSI es ampliamente redundante en su información, es decir, existe una diferencia prácticamente nula entre la Shperical Spin Image de un punto y su vecino con respecto a los demás vértices del mallado. La Tabla **3-8**, muestra el porcentaje de explicación de la varianza para los componentes seleccionados y se aprecia que con 5 componentes se tiene alrededor del 90 % de la varianza.

	AN	DI	FE	HA	SA	SU
componente1	78.26 %	76.63 %	73.47 %	71.57 %	69.11 %	83.62 %
componente2	14.43 %	10.66 %	12.53 %	13.05 %	19.65 %	5.53 %
componente3	1.96 %	2.22 %	1.89 %	1.64 %	1.61 %	1.13 %
componente4	0.84 %	1.68 %	1.35 %	1.42 %	1.00 %	0.97 %
componente5	0.77 %	1.22 %	1.23 %	1.14 %	0.83 %	0.90 %
Total	96.26 %	92.41 %	90.47 %	88.83 %	92.19 %	92.14 %

Tabla 3-8: Explicación varianza PCA para el descriptor SSI

De forma similar debido a que el resultado del descriptor CC está conformado por una

matriz que contiene los valores de CC de cada punto, calculado por cada uno de los wf según se explico en la sección 2.3. La reducción de dimensionalidad se realiza aplicando el análisis de componentes principales PCA y de acuerdo a lo detallado en [38], se consideraron 2 características, las cuales explican más del 90 % de la varianza para todos los gestos .

En resumen, para el caso del descriptor DESIRE, se redujo su dimensión de 428 y 378 correspondientes al rostro completo, y la región de la boca respectivamente a 10 características. En el descriptor SSI, se redujo de un promedio de 600 para el rostro completo y 150 para la boca, a 5 características y para Cone curvature, se redujo de de una matriz de en promedio 600×10 a un vector de 2 características. En todos los casos la varianza explicada del conjunto de características seleccionadas supera el 90 % para todos los gestos.

3.6.2. Análisis de Similitud y efectividad

De la misma forma que se realizó anteriormente, se quiere revisar la similitud entre modelos calculando la distancia entre los vectores característicos, una vez se ha realizado la reducción de dimensionalidad; además se realizan nuevamente las curvas de presicion-recall, para evaluar nuevamente la efectividad en la recuperación de información relevante de cada uno de los descriptores.
Tomando los datos obtenidos en la sección anterior, producto de aplicar el análisis de componentes principales a cada uno de los descriptores, se construyen las tablas en las que se muestra la cantidad de modelos que ocuparon el primer lugar, luego de calcular las distancias en un conjunto de modelos de todos los gestos.

	Boca	
Expresión	casos	Porcentaje
AN	53	88,33 %
DI	58	96,67 %
FE	53	88,33 %
HA	50	83,33 %
SA	55	91,67 %
SU	57	95 %

Tabla 3-9: Aciertos del descriptor DESIRE luego de aplicar PCA.

Con el objetivo de hacer un tipo de verificación con respecto a las reducciones realizadas, se tomó nuevamente el promedio de las características reducidas de cada descriptor para dos grupos y se evaluaron sus distancias, lo cual permite tener una medida de la similitud entre modelos. Esta verificación, se realizó solamente para los descriptores calculados sobre

	Boca	
Expresión	casos	Porcentaje
AN	2	16,67 %
DI	12	60 %
FE	3	16 %
HA	0	%
SA	4	33,33 %
SU	2	16,67 %

Tabla 3-10: Aciertos del descriptor SSI luego de aplicar PCA.

	Boca	
Expresión	casos	Porcentaje
AN	12	80 %
DI	13	86,67 %
FE	8	53,33 %
HA	10	66,67 %
SA	9	60 %
SU	15	100 %

Tabla 3-11: Aciertos del descriptor ConeCurvature luego de aplicar PCA.

la región de la boca.

	AN	DI	FE	HA	SA	SU
AN	**160,26**	$3{,}08e^3$	$2{,}83e^3$	342,23	336,5	$4{,}214e^3$
DI	$2{,}8580e^3$	**123,46**	250,83	$2{,}97e^3$	$3{,}04e^3$	$2{,}38e^3$
FE	$2{,}59e^3$	472,47	**222,87**	$2{,}88e^3$	$2{,}88e^3$	$8{,}38e^3$
HA	362,44	$3{,}07e^3$	$3{,}03e^3$	272,94	**233,44**	$4{,}07e^3$
SA	272,49	$3{,}06e^3$	$2{,}93e^3$	159,2	**53,27**	$4{,}19e^3$
SU	$6{,}48e^3$	$3{,}55e^3$	$3{,}80e^3$	$6{,}60e^3$	$6{,}67e^3$	$\mathbf{2{,}42e^3}$

Tabla 3-12: Distancias entre promedios para el descriptor DESIRE con reducción de dimensionalidad sobre la región de la boca.

De las tablas **3-12** a **3-14** se aprecia claramente la efectividad de los descriptores DESIRE y CC para los cuales sólamente la expresión de alegría (HA) y miedo (FE) respectivamente, no

	AN	DI	FE	HA	SA	SU
AN	0,21	**0,009**	0,2714	0,2847	0,2312	0,2674
DI	0,2899	**0,0781**	0,3512	0,3645	0,3110	0,3473
FE	0,2618	**0,0840**	0,3232	0,3365	0,2830	0,3192
HA	0,0989	0,3106	0,0375	**0,0372**	0,0777	0,0415
SA	0,2279	**0,0236**	0,2892	0,3025	0,2491	0,2853
SU	0,2786	**0,0669**	0,3399	0,3532	0,2998	0,3360

Tabla 3-13: Distancias entre promedios para el descriptor SSI con reducción de dimensionalidad sobre la región de la boca

	AN	DI	FE	HA	SA	SU
AN	**0,1037**	0,1097	0,1937	0,2549	0,1995	2,7498
DI	0,0835	**0,0602**	0,1061	0,1673	0,1119	2,7994
FE	0,1806	0,0895	0,0573	0,0722	**0,0147**	2,7989
HA	0,2664	0,1753	0,0865	**0,0703**	0,0711	2,7697
SA	0,1787	0,1089	0,1056	0,0894	**0,0434**	2,7506
SU	2,8417	2,8478	2,8445	2,8283	2,7823	**0,0214**

Tabla 3-14: Distancias entre promedios para el descriptor CC con reducción de dimensionalidad sobre la región de la boca

coincidieron en menor distancia con su correspondiente; sin embargo haciendo un ranking, estas ocuparon el segundo lugar para cada caso.

De otro lado para el descriptor SSI no se aprecia un buen comportamiento, y los promedios de cada expresión, encuentran la menor distancia con la expresión de disgusto (DI), a excepción de la expresión de alegría (HA), que si obtuvo la menor distancia con su correspondiente.

La Figura **3-6**, muestra el comportamiento de las curvas de precision-recall, de cada uno de los descriptores para cada gesto luego de realizar reducción de dimensionalidad, allí se aprecia que el descriptor DESIRE, sigue teniendo la mejor efectividad en general respecto a la recuperación de datos relevantes, aunque para el caso de la expresion DI, la mayor efectividad la tuvo el descriptor SSI, lo cual se corrobora con el resultado correspondiente de la Tabla **3-10**.

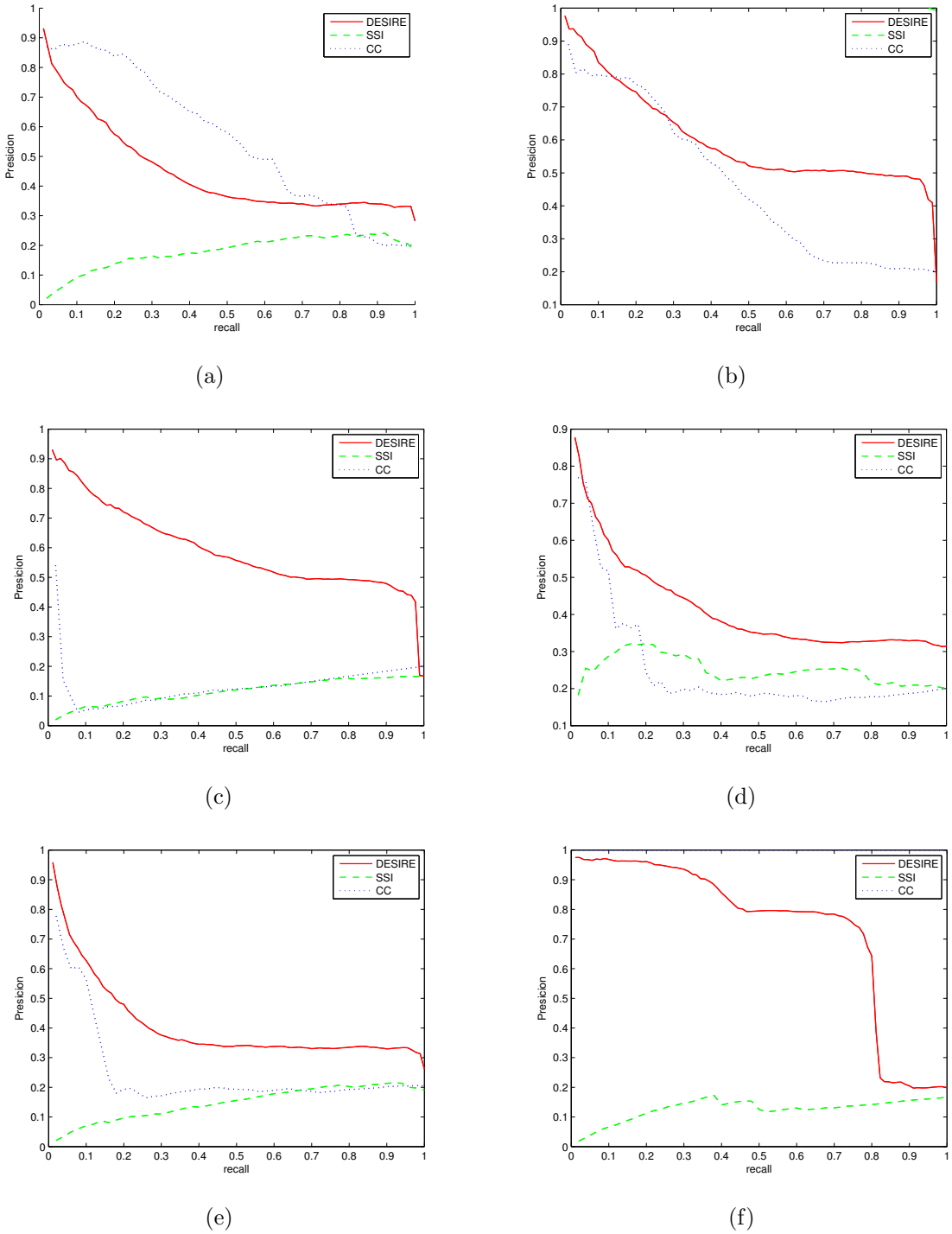


Figura 3-6: Curvas Precision - recall para cada uno de los descriptores en los 6 gestos (a:AN, b:DI, c:FE, d:HA, e:SA, f:SU) sobre la región de la boca luego de aplicar PCA.

4 CLASIFICACIÓN DE GESTOS: RESULTADOS

En este capítulo se realiza el análisis de discriminancia de las características de los tres descriptores para verificar cuales proporcionan la información más relevante, para ello se emplea el análisis discriminante líneal de Fisher, además se entrenan 2 clasificadores y se verifica la capacidad de clasificación con diferentes combinaciones de las características de los descriptores empleados. Se realiza además un análisis multiescala con el descriptor de mejor desempeño, para evaluar las diferencias tras realizar cambios en la resolución del mallado de los modelos.

4.1. Extracción de características discriminantes

El análisis discriminante de Fisher encuentra las características que contienen la mayor información relevante, proyectando los datos en un espacio que maximiza la covarianza entre clases, mientras que minimiza la covarianza intra clases.
El criterio de Fisher se optimiza usando como función criterio la Ecuación 4-1:

$$J(W) = \frac{S_B}{S_W} \tag{4-1}$$

donde S_B, es la matriz de covarianza inter clase, y S_W es la matriz de covarianza intra clase, las cuales son calculadas según las ecuaciones 4-2 y 4-3 respectivamente.

$$S_B = \sum_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \tag{4-2}$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \mu c)(x_i - \mu c)^T \tag{4-3}$$

siendo μ_c la media de cada clase. Es evidente que entre mayor sea el índice de Fisher, más relevante es la información del descriptor. Para los resultados obtenidos en la Sección 3.6.1, se calculó el coeficiente de Fisher de cada descriptor, tanto para el rostro completo como

Descriptor	Indice
DESIRE	5.75
SSI	2.82

Tabla 4-1: Indices de Fisher para cada descriptor todo el rostro

Descriptor	Indice
DESIRE	14.00
SSI	4.82
CC	83.71

Tabla 4-2: Indices de Fisher para cada descriptor región de la boca

para la región de la boca, los resultados se muestran en las tablas **4-1** y **4-2**. Estos indican que la información con mayor relevancia la tiene el descriptor CC para la región de la boca.

El índice de fisher, proporciona una medida del poder discriminante de cada descriptor por separado, es decir, que tanto son capaces de discriminar entre clases las características de cada descriptor. El LDA (Análisis discriminante líneal), nos permite tomar un vector conformado por los tres descriptores, y obtener una reducción de la dimensionalidad seleccionando un número inferior de características formadas como combinación líneal de las originales, las cuales proporcionan la mayor información discriminante.

Se tiene en este caso un vector de carácterísticas $[x_1, x_2, ..., x_{17}]$, donde las características x_1 a x_{10} corresponden a las aportadas por el descriptor DESIRE; las características x_{11} a x_{15} son las correspondientes al SSI, mientras que x_{16} y x_{17} provienen del descriptor CC.

Luego de realizar LDA, y aplicar una regresión líneal a las características obtenidas, para analizar el aporte de cada una de las originales, se obtuvieron 5 funciones, es decir, el numero de clases (6 gestos) menos 1, que separan linealmente cada una de las clases. Las funciones obtenidas son las siguientes:

$y = (A)(X) + C$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y4 \\ y5 \end{bmatrix} = \begin{bmatrix} -0{,}04 & 0{,}019 & -0,013 & 0,027 & -0{,}1375 \\ 0{,}03 & -0{,}044 & -0,022 & -0,0011 & -0{,}0020 \\ -9{,}08 \times 10^{-4} & -9{,}19 \times 10^{-4} & -4 \times 10^{-4} & 4,5 \times 10^{-4} & -0 - 0048 \\ -0{,}022 & 0{,}0235 & 10,5 \times 10^{-4} & 0,0149 & -0{,}05 \\ 0{,}0066 & 0{,}0087 & 0,00178 & -0,0127 & 0{,}0017 \\ -0{,}0283 & 0{,}0291 & 0,00396 & 0,0109 & -0{,}07 \\ 3{,}2 \times 10^{-3} & 5{,}9 \times 10^{-3} & 1{,}27 \times 10^{-3} & -1{,}08 \times 10^{-2} & -1{,}02 \times 10^{-2} \\ 1{,}1 \times 10^{-3} & -1{,}6 \times 10^{-3} & -5{,}96 \times 10^{-4} & -3{,}64 \times 10^{-4} & 2{,}88 \times 10^{-5} \\ 5{,}96 \times 10^{-4} & -1{,}1 \times 10^{-3} & 9{,}31 \times 10^{-4} & -3{,}52 \times 10^{-3} & -2{,}6 \times 10^{-3} \\ -0{,}0039 & 0{,}0063 & -6,7 \times 10^{-5} & 0,008933 & 0{,}0043 \\ -1{,}41 \times 10^{-6} & 2,02 \times 10^{-6} & -2 \times 10^{-6} & 7,4 \times 10^{-6} & 5{,}67 \times 10^{-6} \\ -1{,}13 \times 10^{-7} & 6,47 \times 10^{-8} & -2,59 \times 10^{-9} & 6,23 \times 10^{-9} & 4 \times 10^{-7} \\ -6{,}58 \times 10^{-8} & 8{,}94 \times 10^{-8} & -4,89 \times 10^{-8} & 2,24 \times 10^{-7} & 9{,}8 \times 10^{-8} \\ -2{,}55 \times 10^{-8} & 2{,}83 \times 10^{-8} & -2,13 \times 10^{-8} & 7,5 \times 10^{-8} & 3{,}73 \times 10^{-9} \\ 6{,}48 \times 10^{-8} & 8{,}42 \times 10^{-8} & -3,35 \times 10^{-9} & -8,67 \times 10^{-8} & -1{,}65 \times 10^{-8} \\ 3{,}75 \times 10^{-5} & 3{,}9 \times 10^{-5} & -5,35 \times 10^{-6} & -1,53 \times 10^{-5} & 8{,}97 \times 10^{-5} \\ -9 \times 10^{-6} & 8{,}49 \times 10^{-6} & 1{,}35 \times 10^{-6} & 2{,}21 \times 10^{-6} & -2{,}36 \times 10^{-5} \end{bmatrix}' \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \\ x_{16} \\ x_{17} \end{bmatrix} + \begin{bmatrix} 30,46 \\ 31{,}72 \\ 3{,}70 \\ 14{,}47 \\ -68{,}97 \end{bmatrix} \tag{4-4}$$

De las funciones anteriores se deduce que los elementos que menos aportan a la construcción de las características más discriminantes, son los correspondientes al descriptor SSI (características x_{11} a x_{15}). Los resultados se obtuvieron con 50 imágenes de cada gesto, es decir, 300 en total El análisis con LDA se implementó en Matlab.

4.2. Clasificadores

De nuevo con los resultados de determinar los componentes principales de cada descriptor obtenidos en la Sección 3.6.1, se emplearon dos clasificadores, una red neuronal y un clasificador bayesiano para cada uno de los descriptores, esto se realizó únicamente para los descriptores calculados sobre la región de la boca debido a que fueron los que presentaron los mejores resultados en el análisis previo.

4.2.1. Red Neuronal

Se empleó una red neuronal tipo Backpropagation, con una capa oculta, teniendo en cuenta algunos resultados como el de [47], que demuestran que esta arquitectura es suficiente para resolver este tipo de problemas de clasificación. Para determinar el número de neuronas en la capa oculta, se entrenó la red utilizando el método de validación cruzada, en el que se

está interesado en entrenar y validar a la vez, y detener el entrenamiento en el punto óptimo, es decir, cuando el error de validación sea mínimo.

El procedimiento consistió, en tomar valores pequeños de neuronas en la capa oculta, los cuales se fueron incrementando hasta obtener el mínimo error en la validación. En la Figura **4-1**, se muestra la evolución en el error de aprendizaje y de validación, para una de las pruebas realizadas.

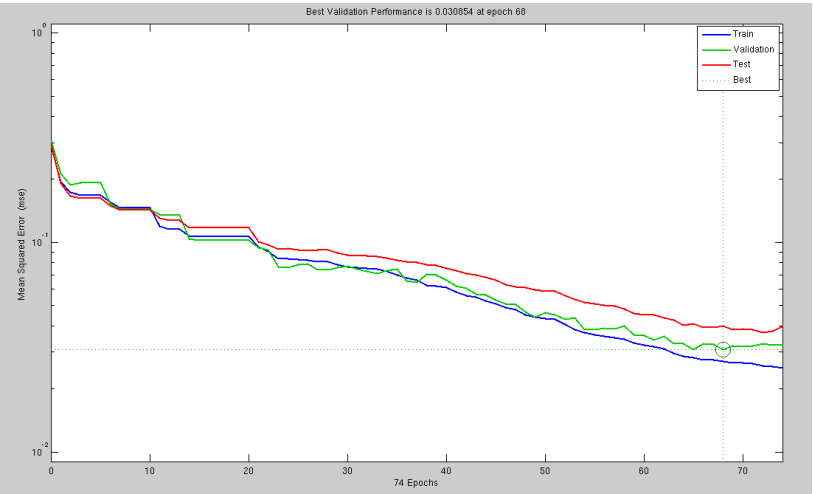


Figura 4-1: Evolucion del error de aprendizaje y validación

Para cada uno de los descriptores se determinó el número adecuado de neuronas en la capa oculta, y se indicó el porcentaje de imágenes clasificadas correctamente. Para el descriptor DESIRE, se encontró que el mejor desempeño se encuentra con 10 neuronas en la capa oculta; para el SSI 8 neuronas; mientras que el menor error en entrenamiento y test para el Cone Curvature, se presentó nuevamente con 10 neuronas en la capa oculta utilizando para todos los casos una red feed forward y el algoritmo backpropagation.

4.2.2. Clasificador Bayesiano

Se utilizó un clasificador Bayesiano simple, definido como:

$$h_{MAP} \equiv argmax_{h \in H} P(h|D) = argmax_{h \in H} P(D|h)P(h) \tag{4-5}$$

El cual es un clasificador probabilístico basado en la teoría de Bayes en el que el aprendizaje se puede ver como el proceso de encontrar la hipótesis más probable, dado un conjunto de ejemplos de entrenamiento D y un conocimiento a priori sobre la probabilidad de cada hipótesis h.

Para este trabajo todas las clases son consideradas equiprobables, puesto que se cuenta con un número igual de muestras para cada una de ellas; además se hace la suposición de que la distribución $P(D|h)$ es Gaussiana.

Las tablas **4-3** a **4-5**, muestran las matrices de confusión resultantes para cada uno de los dos clasificadores aplicado a cada descriptor.

	Red Neuronal						**Bayesiano**					
	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**
AN	78.33	0	0	3.33	18.33	0	90	0	0	8.33	1.67	0
DI	0	95	5	0	0	0	0	95	0	0	0	5
FE	1.67	0	98.33	0	0	0	0	3.33	96.67	0	0	0
HA	1.67	0	0	74	21.67	1.67	5	0	0	78.33	15	1.67
SA	3.33	0	0	13.33	83.33	0	3.33	0	0	10	86.67	0
SU	0	0	0	0	0	100	0	0	0	0	0	100

Tabla 4-3: Matriz de confusion (en porcentajes) del clasificador para el descriptor DESIRE

	Red Neuronal						**Bayesiano**					
	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**
AN	50	8.33	0	0	25	16.67	25	0	0	41.67	16.67	8.33
DI	0	41.67	16.67	0	41.67	0	0	58.33	33.33	8.33	0	0
FE	8.33	0	0	91.67	0	0	0	0	25	75	0	0
HA	0	0	16.67	75	0	8.33	33.33	16.67	0	16.67	16.67	16.67
SA	0	33.33	8.33	41.67	16.67	0	8.33	0	0	91.67	0	0
SU	16.67	16.67	16.67	33.33	8.33	8.33	33.33	8.33	8.33	25	8.33	16.67

Tabla 4-4: Matriz de confusion (en porcentajes) del clasificador para el descriptor SphericalSpinImage

Como era de esperarse, dados los resultados del análisis del capítulo anterior, el mejor desempeño lo tienen los descriptores DESIRE y CC, para este último solo la expresión de miedo (FE) tuvo 0 % de reconocimiento con la red neuronal y 6.67 % con el clasificador bayesiano, las demás expresiones tuvieron una tasa de reconocimiento superior al 80 %.

En cuanto a las tasas de errores de clasificación, tenemos para el caso de la red neuronal que los mayores errores se encontraron para el descriptor DESIRE entre Alegía (HA) y tristeza

	Red Neuronal						Bayesiano					
	AN	**DI**	**FE**	**HA**	**SA**	**SU**	**AN**	**DI**	**FE**	**HA**	**SA**	**SU**
AN	93.33	6.67	0	0	0	0	93.33	6.67	0	0	0	0
DI	13.33	80	6.67	0	10	0	13.33	80	0	0	6.67	0
FE	0	0	0	0	100	0	0	6.67	6.67	0	86.67	0
HA	0	0	0	100	0	0	0	0	0	100	0	0
SA	0	6.67	0	6.67	86.67	0	0	0	6.67	0	93.33	0
SU	0	0	0	0	0	100	0	0	0	0	0	100

Tabla 4-5: Matriz de confusion (en porcentajes) del clasificador para el descriptor Cone Curvature

		AN	**DI**	**FE**	**HA**	**SA**	**SU**
RNA	Sensibilidad	0.78	0.95	0.98	0.74	0.83	1
	Tasa de falsos positivos	0.013	0	0.0099	0.032	0.077	0.0033
	Especificidad	0.98	1	0.99	0.96	0.92	0.99
	Precisión	0.92	1	0.95	0.82	0.68	0.98
	Exactitud	0,644	0,905	0,967	0,587	0,714	1
Bayesiano	Sensibilidad	0.9	0.95	0.97	0.78	0.86	1
	Tasa de falsos positivos	0.0163	0.066	0	0.035	0.032	0.0133
	Especificidad	0,984	0,993	1,000	0,965	0,968	0.987
	Precisión	0,915	0,966	1,000	0,810	0,839	0,937
	Exactitud	0,818	0,905	0,936	0,644	0,765	1,000

Tabla 4-6: Datos estadísticos de las matrices de confusión descriptor DESIRE

(SA) con un 35 %, para Spherical Spin Image, entre miedo (FE) y alegría (HA) con más del 90 % y para Cone Curvature entre miedo (FE) y tristeza(SA) con el 100 %. Las Tablas **4-6** a **4-8** resumen otros datos analizados a partir de las matrices de confusión que nos sirven para evaluar el sistema y que se definen a continuación:

Sensibilidad: Es la probabilidad de que si una instancia pertenece a una categoría, ésta instancia se clasifique con la categoría correcta. Para el caso del descriptor DESIRE, la mejor sensibilidad la tiene la expresión de sorpresa (SU), en los dos clasificadores, para el Spherical Spin Image, es evidente la baja sensibilidad en todas las expresiones para los dos clasificadores, mientras que en el caso del Cone Curvature, la mejor sensibilidad se presenta para las expresiones de Alegría (HA) y sorpresa (SU), las dos con un valor de 1,0.

Tasa de Falsos positivos: Mide la fracción de ejemplos negativos que son clasificados

		AN	DI	FE	HA	SA	SU
RNA	Sensibilidad	0,500	0,417	0,000	0,750	0,167	0,083
	Tasa de falsos positivos	0,045	0,104	0,083	0,317	0,129	0,042
	Especificidad	0,955	0,896	0,917	0,683	0,871	0,958
	Precisión	0,667	0,417	0,000	0,310	0,182	0,250
	Exactitud	0,333	0,263	0,000	0,600	0,091	0,043
Bayesiano	Sensibilidad	0,250	0,583	0,250	0,167	0,000	0,167
	Tasa de falsos positivos	0,130	0,046	0,072	0,414	0,069	0,043
	Especificidad	0,870	0,954	0,928	0,586	0,931	0,957
	Precisión	0,250	0,700	0,375	0,065	0,000	0,400
	Exactitud	0,143	0,412	0,143	0,091	0,000	0,091

Tabla 4-7: Datos estadísticos de las matrices de confusión descriptor Spherical Spin

		AN	DI	FE	HA	SA	SU
RNA	Sensibilidad	0,933	0,800	0,000	1,000	0,867	1,000
	Tasa de falsos positivos	0,026	0,026	0,011	0,013	0,214	0,000
	Especificidad	0,974	0,974	0,989	0,987	0,786	1,000
	Precisión	0,875	0,857	0,000	0,937	0,441	1,000
	Exactitud	0,875	0,667	0,000	1,000	0,765	1,000
Bayesiano	Sensibilidad	0,933	0,800	0,000	1,000	0,933	1,000
	Tasa de falsos positivos	0,026	0,026	0,011	0,013	0,184	0,000
	Especificidad	0,974	0,974	0,989	0,987	0,816	1,000
	Precisión	0,875	0,857	0,000	0,937	0,500	1,000
	Exactitud	0,875	0,667	0,000	1,000	0,875	1,000

Tabla 4-8: Datos estadísticos de las matrices de confusión Descriptor Cone Curvature

incorrectamente como positivos. Lo ideal es que la Tasa de falsos positivos, sea lo menor posible; de acuerdo a las tablas para el caso del DESIRE, la mejor tasa la presenta la expresión de Disgusto (DI) y miedo (FE), para el Spherical Spin Image y el Cone Curvature, la expresión de sorpresa tuvo la mejor tasa de falsos positivos.

Especificidad: Es la proporción de instancias negativas correctamente clasificadas. En general el descriptor DESIRE tiene, de acuedo a la Tabla **4-6**, un excelente nivel de Especificidad, que se refleja con los dos clasificadores empleados, en cuanto al Spherical Spin Image y el Cone curvature, una vez más presentan los mejores resultados para el gesto de sorpresa, aunque en el Cone Curvature se aprecia un mejor comportamiento general para todas las expresiones.

Precisión: Es la probabilidad de que una instancia sea clasificada dentro de una categoría y realmente pertenezca a ella. Las Expresiones de Disgusto (DI), miedo (FE) y sorpresa (SU) tuvieron la mayor precisión para el descriptor DESIRE considerando los dos clasificadores empleados; en cuanto al descriptor Spherical Spin Image, se puede apreciar en la Tabla **4-7**, el pobre nivel de precisión general, teniendo un máximo de 0.7 para la expresión de disgusto. Respecto al Cone Curvatura, aunque tiene expresiones como la de miedo (FE), en la que la precisión es nula, para la expresión de sorpresa se presenta un alto nivel de precisión.

Exactitud: Se refiere a la fracción de las clasificaciones realizadas, en las que se asigna correctamente la categoría. Observando las tablas es evidente que la mayor exactitud se presenta en el reconocimiento del gesto de sorpresa (SU) para los descriptores DESIRE y Cone curvature, mientras que en concordancia con los resultados discutidos anteriormente, el descriptor Spherical Spin Image, presenta un muy mal nivel de exactitud en los reconocimientos para todos los gestos.

Con el objetivo de seguir evaluando el desempeño de los descriptores, frente al reconocimiento de gestos, se planteó un segundo experimento que consiste en construir un vector de características, tomando todos los descriptores, nuevamente se entrenó una red neuronal usando el método de validación cruzada, el cual arrojó que el menor error en la validación, correspondiente al 12 %, se encontró con 10 neuronas en la capa oculta, se empleó también un clasificador Bayesiano y además para este experimento se realizó un clasificador basado en el análisis discriminante líneal, obteniéndose los resultados de la Tabla **4-9**, en la que se resumen las diagonales de las correspondientes matrices de confusión. Para este experimento se aprecia que sólo tuvieron incremento en la tasa de verdaderos positivos, los gestos de FE y SA, con respecto a los mejores resultados de los descriptores individualmente.
Un tercer experimento consiste en construir un vector únicamente con los descriptores de mejor desempeño individual, es decir, DESIRE y CC, que además corresponden a las características más relevantes según el análisis discriminante LDA; para este, el entrenamiento

de la red neuronal arrojó que el menor error, 11 %, se presentaba con 8 neuronas en la capa oculta. Los resultados obtenidos se presentan en la Tabla **4-10**, en la que se observa que únicamente para el reconocimiento de los gestos AN y HA, se presentó un descenso en el desempeño del 19,64 % y el 25 % con la red neuronal, respectivamente. Para los demás, se presenta una tasa de reconocimientos superior al 90 %.

Expresion	Red Neuronal	Bayesiano	LDA
AN	66.67 %	100 %	100 %
DI	91.67 %	91.67 %	100 %
FE	100 %	83.33 %	50 %
HA	75 %	83.33 %	83.33 %
SA	91.67 %	83.33 %	100 %
SU	100 %	100 %	100 %

Tabla 4-9: Expresiones identificadas combinando los tres descriptores

Expresion	Red Neuronal	Bayesiano	LDA
AN	75 %	100 %	100 %
DI	100 %	91.67 %	91.67 %
FE	100 %	100 %	91.67 %
HA	75 %	91.67 %	91.67 %
SA	91.67 %	100 %	100 %
SU	100 %	100 %	100 %

Tabla 4-10: Expresiones identificadas descriptor combinando DESIRE y CC

Finalmente, con el objetivo de realizar una comparación de los resultados obtenidos con otros descriptores basados en las curvaturas principales k_1 y k_2, y como complemento al trabajo realizado en [44], se elaboró el mismo procedimiento anterior para las curvaturas media y Gaussiana, definidas como $H = \frac{k_1+k_2}{2}$ y $K = k_1 \times k_2$, respectivamente.
Los resultados se muestran en las tablas **4-11** y **4-12** que resumen las diagonales de las correspondientes matrices de confusión.
Estos resultados demuestran que según el análisis presentado en [44], los descriptores basados en las curvaturas caracterizan muy bien algunas regiones del rostro, sin embargo su comportamiento frente al reconocimiento de 6 expresiones faciales está por debajo de otros basados en imágenes y volumen, como los presentados en el presente trabajo.
Si promediamos los resultados obtenidos luego de combinar los descriptores de mejor desempeño y considerando únicamente la red neuronal como clasificador, tenemos una tasa

Expresion	Red Neuronal	Bayesiano
AN	58.33 %	16.67 %
DI	66.67 %	50 %
FE	16.67 %	16.67 %
HA	41.67 %	16.67 %
SA	41.67 %	41.67 %
SU	33.33 %	25 %

Tabla 4-11: Expresiones identificadas curvatura media H

Expresion	Red Neuronal	Bayesiano
AN	33.33 %	50 %
DI	58.33 %	25 %
FE	58.33 %	25 %
HA	58.33 %	25 %
SA	25 %	16.67 %
SU	58.33 %	50 %

Tabla 4-12: Expresiones identificadas curvatura Gaussiana K

de reconocimiento promedio del 90.3 % (Tabla **4-10**), lo cual representa un incremento del 3.5 % con respecto a los resultados obtenidos en [45] donde se realiza reconocimiento de gestos faciales empleando un método basado en el análisis de segmentos de línea que conectan puntos característicos del rostro; y del 7.4 % con respecto al mejor de los resultados presentados en [46], donde se realiza el análisis extrayendo algunas características primitivas y evaluando cuatro diferentes tipos de clasificador. La Tabla **4-13** compara los resultados de este trabajo con los obtenidos en [45] y [46].

	Resultados en [45]	Resultados en [46]				**Resultados de este trabajo**		
Clasificador	SVM	QDC	LDA	NBC	SVC	RNA	Bayesiano	LDA
Promedio de reconocimientos	87.1 %	74.5 %	83.6 %	71.7 %	77.8 %	90.3 %	97.2 %	95.8 %

Tabla 4-13: Comparación promedios de reconocimiento

4.3. Análisis multiescala

De los resultados de las secciones anteriores se concluye que el descriptor con el mejor desempeño frente al reconocimiento de gestos faciales considerandolos individualmente, es el

DESIRE, en esta sección se quiere hacer un análisis para verificar el comportamiento de este descriptor con el mismo objetivo, pero considerando múltiples escalas, es decir, realizando una disminución de la resolución al mallado de la base de datos original, con el ánimo de observar el rendimiento del descriptor con una cantidad inferior de puntos.
Se realizan cambios en la resolución de $\frac{N}{2}$, $\frac{N}{4}$ y $\frac{N}{8}$, siendo N, el número de vértices del modelo original. La Figura **4-2** muestra un ejemplo del modelo original de un rostro y las escalas empleadas.

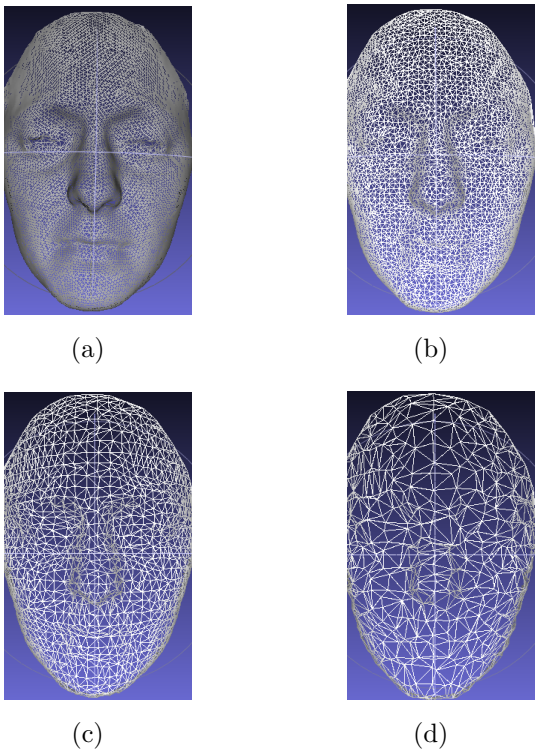


Figura 4-2: Modelo original y cambio de resolución a $N/2$, $N/4$ y $N/8$

Debido a los resultados que demuestran que individualmente el descriptor DESIRE, calculado sobre la región de la boca, presenta un excelente comportamiento, se calculó el descriptor nuevamente para esta región, sobre los modelos remuestreados.
En las tablas **4-14** a **4-16** aparecen los resultados de aplicar dos clasificadores al mallado producto de la reducción en factores de $\frac{N}{2}$, $\frac{N}{4}$ y $\frac{N}{8}$, respectivamente. Se observa que se tiene una gran reducción con respecto a los resultados obtenidos con el modelo original; solamente la expresión de sorpresa, tiene una tasa de verdaderos positivos que podría considerase aceptable con un porcentaje promedio de reconocimientos del 66.57 % para el clasificador con red neuronal; sin embargo para obtener mayor información se realiza la Tabla **4-17** basada en las matrices de confusión para el modelo con reducción de $\frac{N}{2}$.
De acuerdo a los resultados de la Tabla **4-17**, el desempeño del descriptor es en general muy pobre, si comparamos los resultados de, por ejemplo, sensibilidad, precisión y exactitud con

Expresion	Red Neuronal	Bayesiano
AN	3 %	32,65 %
DI	28,5 %	1,024 %
FE	28,5 %	0 %
HA	9,18 %	0 %
SA	2 %	0 %
SU	60,2 %	69,4 %

Tabla 4-14: Expresiones identificadas descriptor DESIRE para resolución 2/N

Expresion	Red Neuronal	Bayesiano
AN	2 %	29,59 %
DI	38,77 %	1,02 %
FE	21,42 %	0 %
HA	5,1 %	1,02 %
SA	0 %	0 %
SU	66,32 %	64,3 %

Tabla 4-15: Expresiones identificadas descriptor DESIRE para resolución 4/N

Expresion	Red Neuronal	Bayesiano
AN	1,02 %	53 %
DI	29,59 %	1,02 %
FE	23,46 %	0 %
HA	1,02 %	0 %
SA	0 %	0 %
SU	72,45 %	34,7 %

Tabla 4-16: Expresiones identificadas descriptor DESIRE para resolución 8/N

respecto a los obtenidos en la Tabla **4-6**. Esto indica que el descriptor DESIRE, aúnque tuvo un muy buen desempeño en el análisis anterior, no presenta suficiente robustez frente a cambios en la escala, lo cual representa un inconveniente en un sistema de reconocimiento, debido a que no se tiene certeza de las posibles diferencias en la resolución del mallado que puedan tener los modelos a reconocer, luego de que el clasificador se ha entrenado con datos a escalas diferentes.

		AN	**DI**	**FE**	**HA**	**SA**	**SU**
RNA	Sensibilidad	0,020	0,286	0,286	0,092	0,020	0,602
	Tasa de falsos positivos	0,000	0,106	0,086	0,086	0,068	0,651
	Especificidad	1,000	0,894	0,914	0,914	0,932	0,349
	Precisión	1,000	0,326	0,374	0,155	0,049	0,149
	Exactitud	0,010	0,167	0,167	0,048	0,010	0,431
Bayesiano	Sensibilidad	0,327	0,010	0,000	0,000	0,000	0,694
	Tasa de falsos positivos	0,329	0,002	0,010	0,009	0,002	0,606
	Especificidad	0,671	0,998	0,990	0,991	0,998	0,394
	Precisión	0,152	0,505	0,000	0,000	0,000	0,181
	Exactitud	0,195	0,005	0,000	0,000	0,000	0,531

Tabla 4-17: Datos estadísticos de las matrices de confusión Descriptor DESIRE para resolución N/2

CONCLUSIONES Y TRABAJO FUTURO

Se han presentado tres descriptores (DESIRE, Cone Curvature y Spherical Spin Image), que pertenecen a las categorías de los basados en características y basados en vistas. Luego de realizar la implementación en lenguaje C++, y calcularlos en rostros de una base de datos de imágenes 3D, sobre el rostro completo y la región de la boca, se evalúa su desempeño frente al reconocimiento de gestos faciales, encontrando que dos de ellos DESIRE y Cone Curvature (CC) tienen un desempeño muy bueno previo entrenamiento de un clasificador basado en redes neuronales artificiales y un clasificador Bayesiano. Mientras que el primero es capaz de reconocer 6 gestos con un desempeño promedio del 88.2 % con la red neuronal, el segundo reconoció 5 gestos con un desempeño promedio del 76.7 %. El descriptor SSI tuvo un pobre desempeño promedio del 36 %.

Realizando una combinación de los tres descriptores se encuentra una mejora en el desempeño respecto a CC, aunque no se supera el resultado del DESIRE. Tomando únicamente un vector de características conformado por los dos descriptores de mejor desempeño (DESIRE y CC), se obtiene un incremento del 2.4 % en la tasa de reconocimientos, respecto al DESIRE, lo cual permite concluir que un vector de características conformado por estos dos descriptores puede hacer parte de un sistema de reconocimiento con un alto grado de presición. Sin embargo, una limitante encontrada tiene que ver con el tiempo de cálculo del descriptor CC, cuyo promedio para la región de la boca (aproximadamente 1000 vértices) es de 25 minutos, lo cual lo haría poco viable para un sistema actuando en tiempo real.

Aunque el análisis se realizó calculando los descriptores sobre todos los puntos del rostro y también únicamente sobre la región de la boca, se verificó que se tiene mejor desempeño si se tiene en cuenta solo esta última, no sólo porque se reduce considerablemente el tiempo de cálculo, sino porque además esta región es la de mayor variabilidad cuando se cambia la expresión, esta segmentación se realizó de forma manual con la ayuda del software Meshlab. Sin embargo, sería interesante explorar en un trabajo futuro, la implementación de un método de segmentación automática de la región de interés, para que el sistema pueda ser empleado en un proceso de reconocimiento en tiempo real.

Luego de analizar el desempeño del descriptor DESIRE, con el que se obtuvieron mejores

resultados, en un análisis multiescala, tomando los modelos a diferentes resoluciones del mallado, se verificó que este descriptor no presenta un buen comportamiento reconociendo gestos de mallados con resoluciones diferentes al del entrenamiento, lo cual afectaría también un sistema de reconocimiento, en el que las resoluciones pueden ser variables.

A pesar de los resultados obtenidos que indican que es posible realizar reconocimiento de expresiones con los descriptores analizados, los tiempos de cálculo hacen poco viable la implementación de un sistema en tiempo real, por este motivo sería interesante explorar mecanismos que permitan decrementar el tiempo bien sea en la implementación o buscando modificaciones a los algoritmos.

Analizando la región de la boca se obtuvieron resultados satisfactorios en cuanto al promedio de expresiones reconocidas, sin embargo podría considerarse el análisis sobre otras regiones que involucren menos puntos, lo cual podría además, ayudar a reducir los tiempos.

Bibliografía

[1] N. Tsapatsoulis, Y. Avrithis, and S. Kollias. "On the use of Radon Transform for Facial Expression Recognition". in *Proceedings of 5th International Conference on Information Systems Analysis and Synthesis (ISAS 1999)*. Orlando, FL, USA, (1999).

[2] Lien, J.J.-J., Kanade, T., Cohn, J.F., Ching-Chung Li. , " Subtly different facial expression recognition and expression intensity estimation", in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference* 853-859, (1998).

[3] L., Silva ; O.R.P., Bellon ; K.L., Boyer: Robust range image registration using genetic algorithms and the surface interpenetration measure. World Scientific, 2005. – 176 p – ISBN 9812561080

[4] L., Chi-Fang: A new approach to high precision 3D measuring system. En: Image Vision Computing 17 (1999)

[5] Hartley R., Zisserman A Multiple View Geometry in Computer Vision 2ed. Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer and Mubarak Shah. " Shape from Shading: A Survey".

[6] E. Prados, O. Faugeras in Handbook of Mathematical Models in Computer Vision, Springer, page 375–388 (2006).

[7] Essa, I.A., Darrell, T., Pentland, A., "Tracking facial motion", *Motion of Non-Rigid and Articulated Objects*, 1994., Proceedings of the 1994 IEEE Workshop, 36-42, (1994).

[8] Ekman, P. Editor. (1982): Emotion In the Human Face. Cambridge University Press, New York, NY, 2nd edition,.

[9] Rosenblum, M.; Yacoob, Y.; Davis, L., "Human emotion recognition from motion using a radial basis function network architecture", *Motion of Non-Rigid and Articulated Objects*, 1994., Proceedings of the 1994 IEEE Workshop, 43-49 (1994).

[10] A. E. Johnson, M. Hebert. Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes. IEEE Transaction on pattern analysis and machine intelligence, vol. 21, NO. 5, (1999).

[11] G. Passalis Â· T. Theoharis Â· I. A. Kakadiaris. "PTK: A Novel Depth Buffer-Based Shape Descriptor for Three-Dimensional Object Retrieval". *International Journal of Computer Graphics archive*, **23**, 5-14, (2006).

[12] Ceron, A.; Salazar, A.; Prieto, F.; , "Relevance analysis of 3D curvature-based shape descriptors on interest points of the face", in *Image Processing Theory Tools and Applications (IPTA) 2010 2nd International Conference* 452-457 (2010).

[13] Cristina Conde, Licesio J. Rodríguez-Aragón, and Enrique Cabello; " Automatic 3D Face Feature Points Extraction with Spin Images".in *Image Analysis and Recognition, Third International Conference, ICIAR 2006*, **4142**, 317-328 (2006).

[14] Chin-Seng Chua; Feng Han; Yeong-Khing Ho; , "3D human face recognition using point signature," Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on , vol., no., pp.233-238, (2000)

[15] G. Gordon, "Face recognition based on depth maps and surface curvature", in SPIE Proc: Geometric Methods in Computer Vision, vol.1570, pp. 234–247, (1991).

[16] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. ACM Transactions on Graphics, 21:807–832, (2002).

[17] R. Ohbuchi, T. Minamitani, and T. Takei. Shape-similarity search of 3D models by using enhanced shape functions. International Journal of Computer Applications in Technology, 23(2):70–85, (2005).

[18] Jurgen Assfalg, A. Del Bimbo, and P. Pala. Retrieval of 3D objects using curvature maps and weighted walktroughs. In ICIAP, (2003).

[19] Chang Ha Lee, Amitabh Varshney, and David W. Jacobs. Mesh saliency. ACM Trans. Graph., 24(3):659–666, (2005).

[20] U. Castellani, M. Cristani, S. Fantoni, and V. Murino. Sparse point matching by combining 3D mesh saliency with statistical descriptors. Computer Graphics Forum,27(2):643–652, (2008).

[21] T. Zaharia and F. Preteux. 3D shape based retrieval within the mpeg-7 framework. SPIE Applications, 4304, (2001).

[22] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3D shape descriptors. In Eurographics, pages 156–164, (2003).

[23] Hamid Laga, Hirko Takahashi, and Masayuki Nakajima. Spherical wavelet descriptors for content-based 3D model retrieval. In Shape Modeling and Applications, pages 15–23,(2006).

[24] H. Sundar, D. Silver, N. Gagvani, and S. Dickenson. Skeleton-based shape matching and retrieval. In Shape Modeling International, pages 130–138, (2004).

[25] M. Hilaga, Y. Shinagawa, and T. Kohmura. Topology matching for fully automatic similarity estimation of 3D shapes. In SIGGRAPH, pages 203–212, (2001).

[26] D. Chen, X. Tian, Y. Shen, and M. Ouhyoung. On visual similarity based 3D model retrieval. Computer Graphics Forum, 22(3):223–232, (2003).

[27] J. Kittler, A. Hilton, M. Hamouz, J. Illingworth.: 3D Assisted Face Recognition: A Survey of 3D imaging, Modelling and Recognition Approaches. IEEE CVPR05 Workshop on Advanced 3D Imaging for Safety and Security. San Diego, CA, (2005)

[28] Jaccard, Paul, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura", Bulletin de la Société Vaudoise des Sciences Naturelles 37: 547–579 (1901)

[29] Spearman. C. "The proof and measurement of association between two things". Amer. J. Psychol. 15: 72–101 (1904).

[30] Benjamin P. Blackburne, Simon Whelan. " Measuring the distance between multiple sequence alignments".Bioinformatics, Vol. 28, No. 4. (15 February 2012), pp. 495-502.

[31] Vranic, D.V.; , "DESIRE: a composite 3D-shape descriptor," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference* 4 (2005).

[32] D.V. Vranic, 3D Model Retrieval, Ph. D. Thesis, University of Leipzig, Germany, (2004).

[33] D. V. Vranic and D. Saupe, "3D Model Retrieval" in *Proc. Spring Conference con Computer Graphics and its Applications (SCCG2000)*, B. Falcidieno, Ed.,Budmerice Manor, Slovakia, 89–93, Comenius University (2000).

[34] D. V. Vranic, "An improvement of Ray-Based Shape Descriptor ", In *Proceedings of the 8. Leipziger Informatik-Tage (LIT2M)*, W. Wittig and S. Paul,Eds., Leipzig, Germany, 55–58, HTWK Leipzig (2000)

[35] T. Moller and B. Trumbore, "Fast, minimum storage ray-triangle intersection," ACM Journal of Graphics Tools, 2(1):21–28, (1997).

[36] Yueming Wang, Gang Pan, Zhaohui Wu, and Shi Han. "Sphere-Spin-Image: A Viewpoint-Invariant Surface Representation for 3D Face Recognition," in *Computer Science, 2004*, **3037**, 427-434 (2004).

[37] A. E. Johnson, M.Hebert. "Surface matching for object recognition in complex three-dimensional scenes," in *Image Vision Computing, 1998*, **16**, 635-651 (1998).

[38] Antonio Adan, Miguel Adan A Flexible Similarity Measure for 3D Shapes Recognition. IEEE Transaction on pattern analysis and machine intelligence, vol. 26, NO. 11, november (2004)

[39] Adan, A., Cerrada, C., Feliu, V.: Modeling Wave Set: Definition and Application of a new Topological Organization for 3D Object Modeling. Computer Vision and Image Understanding 79, 281–307 (2000)

[40] Lijun Yin Xiachov wei,Yi Sun, Jun Wang, Matthew J. Rosato, "A 3D Facial Expression Daabase For Facial Behavior Research," in *IEEE 7th International conference on Automatic Face and Gesture Recognition (FGR06)*, Southampton, 211-216 (2006).

[41] Cignoni P., Corsini M., Ranzuglia G. "Meshlab: an open-source 3d mesh processing system," in *ERCIM News, 2008*, **73**, 45–46, (2008). http://meshlab.sourceforge.net.

[42] Chris Lomont. Fast inverse square root. Technical report, Purdue University, 1–14 (2003)

[43] Fisher, R. A.. "The Use of Multiple Measurements in Taxonomic Problems". Annals of Eugenics 7 (2): 179–188, (1936)

[44] Cerón Correa, Alexander." Análisis comparativo de descriptores de forma 3D para detección de características faciales". Maestria tesis, Universidad Nacional de Colombia. (2011)

[45] Hao Tang; Huang, T.S."3D facial expression recognition based on properties of line segments connecting facial feature points" in *Automatic Face and Gesture Recognition, 2008. FG '08. 8th IEEE International Conference* 1-6 (2008)

[46] Jun Wang; Lijun Yin; Xiaozhou Wei; Yi Sun; , "3D Facial Expression Recognition Based on Primitive Surface Feature Distribution," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference* **2**, 1399- 1406 (2006).

[47] Funahashi,K.I. "On the approximate realization of continuous mappings by neural networks". Neural Networks, 2, 183-192 (1989).