



**Julián S. Rodríguez**  
Universidad de San Buenaventura  
Bogotá  
jrodriguez@usbog.edu.co

**Flavio Prieto**  
Universidad Nacional de Colombia  
fprieto@unal.edu.co



# Análisis y comparación del descriptor *Cone curvature* frente al reconocimiento de expresiones faciales

## *Analysis and comparison of the cone curvature descriptor in facial gesture recognition tasks*

### Resumen

Se presenta el resultado de analizar el comportamiento del descriptor de forma *Cone Curvature* (CC) en la tarea de reconocimiento de expresiones faciales en imágenes 3D. El descriptor CC es una representación del modelo 3D que se calcula a partir de un conjunto de ondas de modelado para cada vértice de una malla poligonal. Se empleó la base de datos de rostros 3D (BU-3DFE), la cual contiene imágenes con seis expresiones faciales. Con el uso del descriptor CC, las expresiones fueron reconocidas en un porcentaje promedio del 76.67 % con una red neuronal y del 78.88 % con un clasificador bayesiano. Al realizar una combinación del descriptor CC con otros descriptores como DESIRE y *Spherical Spin Image*, se logró un porcentaje promedio de reconocimiento de gestos del 90.27 % y del 97.2 %, usando los mismos clasificadores mencionados previamente.

**Palabras claves:** Descriptores de forma, reconocimiento facial, visión artificial.

### Abstract

This article presents the results of analyzing the behavior of the Cone Curvature shape descriptor (CC) in the task of recognition of facial expressions in 3D images. The CC descriptor is a representation of the 3D model computed from a set of waves modeling for each vertex of a polygon mesh. The 3D Facial Expression Database (BU-3DFE) was used, which contains images with six facial expressions. With the use of the CC descriptor, the expressions were recognized in an average percentage of 76.67 % with a neural network, and of 78.88 % with a Bayesian classifier. By combining the CC descriptor with other descriptors such as DESIRE and *Spherical Spin Image*, it was achieved an average percentage of gesture recognition of 90.27 % and 97.2 %, using the mentioned classifiers.

**Key words:** Artificial vision, facial recognized, shape descriptors.

Recibido: 15-05-2015  
Modificado: 07-08-2015  
Aceptado: 24-08-2015

# 1. Introducción

Recientemente se ha despertado un fuerte interés por el reconocimiento automático de expresiones faciales, lo cual posee variadas aplicaciones, que van desde el desarrollo de interfaces hombre-máquina, sistemas de seguridad que detectan comportamientos “sospechosos” en sitios públicos, hasta la posibilidad de desarrollar robots que interactúen con humanos, de acuerdo con expresión manifestada por estos últimos.

El psicólogo Paul Ekman [1], uno de los pioneros del reconocimiento de emociones mediante el análisis de expresiones faciales, definió seis grandes expresiones: alegría (HA), disgusto (DI), miedo (FE), enojo (AN), sorpresa (SU) y tristeza (SA). Se han desarrollado diferentes técnicas para reconocer dichas expresiones faciales, algunas de ellas basadas en imágenes 2D estáticas [2], otras basadas en secuencias de imágenes empleando aproximaciones por flujo óptico [3], [4], por rastreo de características [5], o por alineamiento del modelo [6]. Generalmente, los trabajos de reconocimiento de expresiones a partir de imágenes 2D, presentan inconvenientes, como sensibilidad a la variación en la iluminación, orientación y escala. También se han reportado desarrollos de reconocimiento facial en tres dimensiones [7]–[9]. El objetivo de realizar un análisis con modelos 3D del rostro, es obtener características que sean invariantes a estos parámetros, lo que permite mayor robustez y confiabilidad en los resultados obtenidos.

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, reconocimiento de rostros y control de calidad. En particular, en la tarea de reconocimiento de rostros dichos descriptores son de mucha utilidad, debido a la gran cantidad de información morfológica que pueden extraer. Algunos trabajos se han centrado en obtener ciertas características del rostro humano, como en [10], 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 [11] se analizan varios descriptores basados en curvatura. En otros trabajos como en [12] y [13], se han hecho revisiones acerca de los métodos empleados para realizar reconocimiento facial. En [14] 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 87.1 %. Wang en [15] realiza también un estudio del reconocimiento de expresiones en mallados 3D, basándose en las curvaturas principales. En [16] se realiza un análisis comparativo de dos descriptores de forma, *Spherical Spin Image* (SSI) y DESIRE, para determinar el de mejor desempeño ante el reconocimiento de expresiones faciales en modelos 3D. En el presente trabajo se realiza la implementación del descriptor *Cone Curvature* (CC) sobre modelos 3D del rostro; los resultados obtenidos son comparados con los obtenidos en [16] para los descriptores DESIRE y SSI, además se hace una combinación de todos los descriptores y se evidencia la mejoría en los resultados de reconocimiento de expresiones.

El descriptor *Cone Curvature* (CC) ha mostrado ser eficiente en el reconocimiento de objetos [17], por este motivo se realizó el estudio con este descriptor para verificar su desempeño frente al reconocimiento de expresiones faciales en modelos 3D. Los resultados de este estudio son útiles para comparar la capacidad de reconocimiento de expresiones faciales utilizando este descriptor *versus* los resultados obtenidos con otros descriptores de forma, lo que permitirá seleccionar el método adecuado para reconocer expresiones faciales para posibles aplicaciones como interfaces avanzadas hombre-máquina.

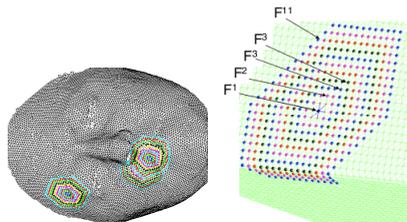
El artículo está organizado de la siguiente manera: en la Sección 2 se hace una breve explicación acerca de los fundamentos del descriptor empleado, así como de los utilizados en [16]. En la Sección 3 se describen los análisis realizados para evaluar la efectividad del descriptor CC en la tarea de clasificación de gestos faciales. En la Sección 4 se muestran los resultados de reconocimiento del descriptor tras entrenar dos tipos de clasificadores, también se realiza la combinación del descriptor CC con otros dos descriptores de forma, de igual forma se presenta el análisis de los resultados obtenidos y se hace la comparación con lo obtenido en [16]. Finalmente, las conclusiones son presentadas en la Sección 5.

## 2. Descriptores de forma

Los descriptores de forma son herramientas matemáticas que permiten extraer, en forma de datos numéricos, información acerca de la geometría de un objeto 3D. Un buen descriptor debe tener invarianza con respecto a traslación, rotación, escala y reflexión del modelo 3D, además debe ser robusto frente al ruido presente en la superficie y a valores atípicos. A continuación se hace una breve descripción del descriptor *Cone Curvature* (CC) implementado en este trabajo, así como de otros dos descriptores: *Spherical Spin Image* y DESIRE, implementados en [18] y [19] respectivamente. Dichos descriptores fueron combinados con el descriptor CC para evaluar o mejorar su desempeño en la tarea de reconocimiento de expresiones faciales.

### 2.1. Cone Curvature (CC)

Antes de definir el descriptor *Cone Curvature* es necesario describir los conjuntos de onda de modelado (MWS) [20], dichos conjuntos relacionan subconjuntos de nodos en una malla triangular  $T$  de un objeto (Modelo 3D), organizando los nodos en puntos concéntricos, dispuestos alrededor de un vértice  $N$  perteneciente a  $T$ . Cada uno de los grupos concéntricos (figura 1) se llama un *wave front* (WF), y el nodo inicial  $N$  de cada uno de ellos se conoce como foco. Todas las posibles ondas de modelado (MW) que se pueden generar sobre  $T$ , se conocen como conjunto de ondas de modelado. Entonces,  $MWS = \{MW^1, MW^2, \dots, MW^q\}$ , donde  $MW^i$  es la onda de modelado generada a partir del foco  $N$ , que corresponde con el  $i$ -ésimo vértice de la malla  $T$ .



**Figura 1.** WF en diferentes focos sobre el modelo 3D del rostro (Izquierda). Diferentes frentes de ondas de un nodo  $N$  (derecha) [21].

El *Cone Curvature* es una representación del modelo 3D, que se calcula a partir del MWS para cada nodo de la malla y cuya definición en [22]: se llama Cone Curvature (CC)  $j$ -ésimo de  $N$  ( $\alpha^j$ ), al ángulo del cono con vértice  $N$  cuya superficie está aproximada al  $j$ -ésimo frente de onda ( $F^j$ ), de la onda de modelado asociada a  $N$ . Formalmente:

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

donde  $\gamma_i^j = \angle C^j N N_i$ ,  $N_i \in F^j$ ,  $t_j$  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 planas y valores positivos son zonas convexas. La figura 2 ilustra la definición del CC.

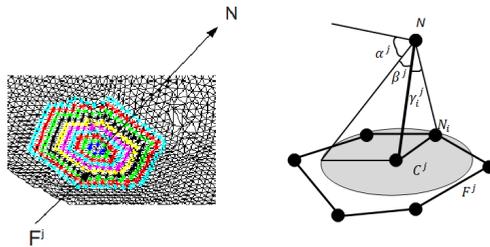


Figura 2. Definición del Cone Curvature [22].

Dado un foco  $N$ , existe un conjunto de frentes de onda que determina todas las CC para  $N$   $\{\alpha^1, \alpha^2, \dots, \alpha^q\}$ , las cuales proporcionan información acerca de la curvatura del objeto desde el punto de vista de  $N$ . La información de la curvatura del objeto es útil en el reconocimiento de expresiones faciales porque en cada expresión las diferentes regiones del rostro, y en particular la de la boca, cambian su forma, lo que implica un cambio en la curvatura de la región desde el punto de vista de cada vértice para el cual se calcula el CC. El algoritmo 1 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 [22] se demuestra que un valor de  $q$  mayor tiene mejores resultados, allí se experimentó 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 quince para la región de la boca (figura 3), en el presente trabajo se consideró  $q = 10$ .

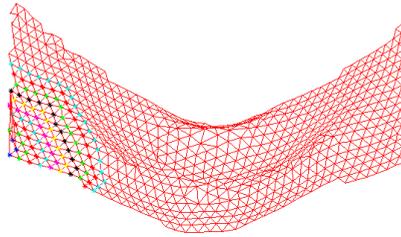


Figura 3. diez frentes de onda en la región de la boca. Se podrían tener un máximo promedio de 15 frentes de onda.

---

**Algorithm 1** Cone Curvature

---

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

**for**  $indice < cantidad\ de\ wf$  **do**

$baricentro \leftarrow arreglobaricentros(indice)$

Cálculo de parámetro  $\gamma$

$distanciabaricentro \leftarrow$  cálculo de la distancia entre baricentro y origen  $\gamma$

$distanciaFoco \leftarrow$  cálculo de la distancia entre  $N$  y el origen  $\gamma$

**if**  $distanciabaricentro > distanciafoco$  **then**

$signo \leftarrow -1$

**else**

$signo \leftarrow 1$

**end if**

cálculo de  $\alpha(indice)$

**end for**

---

## 2.2. Spherical Spin Image (SSI)

El descriptor *Spherical Spin Image* (SSI) [18] 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 donde cada píxel es un bin que almacena el número de vecinos que están: a una distancia  $\alpha$  a partir de  $n$  y a una profundidad  $\beta$  de su plano tangente  $p$  (Figura 4).

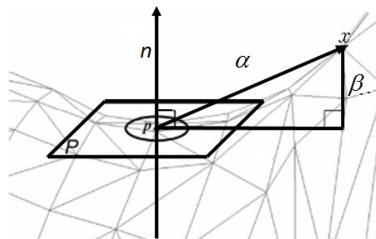


Figura 4. Punto orientado para la extracción de *Spherical Spin Image* [18].

### 2.3. DESIRE

El descriptor DESIRE [19] es un descriptor compuesto conformado por imágenes de profundidad (**DE**pth), siluetas (**SI**) y rayos extendidos de una malla poligonal (**Ray-Ext**ent). 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 (SI) 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. La Figura 5 ilustra cada uno de los tres componentes de este descriptor.

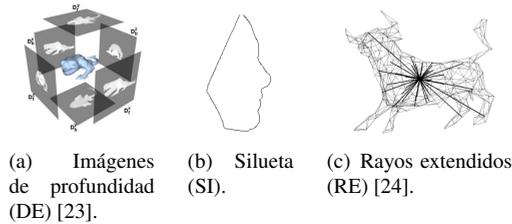


Figura 5. Los tres componentes del descriptor DESIRE.

## 3. Análisis de similitud

### 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) [25], 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 seis expresiones: alegría (HA), disgusto (DI), miedo (FE), enojo (AN), sorpresa (SU) y tristeza (SA); frente a un escáner 3D. Estas imágenes tienen formato vml y están constituidas por mallas triangulares. La figura 6 muestra algunos modelos de la base de datos.

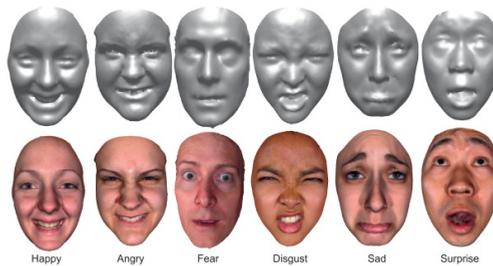


Figura 6. Ejemplos de modelos pertenecientes a la base de datos empleada [25].

### 3.2. Clasificación por distancia

Una forma de encontrar similitudes entre vectores característicos es calcular la distancia entre ellos. Para verificar dicha similitud, se extrajo el descriptor CC a un conjunto de imágenes de la base de datos para cada gesto, se calculó la distancia entre el vector característico de una imagen de búsqueda representativa de cada gesto denominada  $Q$  ( $Q_{AN}$ ,  $Q_{DI}$ ,  $Q_{FE}$ ,  $Q_{HA}$ ,  $Q_{SA}$ ,  $Q_{SU}$ ) y los vectores característicos de otras 300 imágenes, de las cuales 50 corresponden a la misma clase de  $Q$ . Idealmente se espera que las distancias entre el vector de características de la imagen buscada y cada una de las 50 imágenes del gesto correspondiente sean las menores. Luego de calcular el descriptor CC se tiene una matriz de tamaño  $h \times q$ , 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  $T_m$  y  $T_n$  se define de como [22]:

$$d^j(T^m, T^n) = \sqrt{\sum_{k=1}^h (\theta_{(k,j)}^m - \theta_{(k,j)}^n)^2}, \tag{2}$$

donde,  $\theta_1$  y  $\theta_2$  son las matrices de CC para cada modelo. Para cada pareja de modelos ( $m$ ,  $n$ ) se obtiene un vector de distancias  $\{d^1, d^2, \dots, d^q\}$ .

Como medida de similitud se utilizó la métrica de distancia norma-infinito  $l_\infty$ , calculada de acuerdo a la ecuación 3:

$$d(T^m, T^n) = \max |d^j(T^m, T^n)|, j = 1, 2, \dots, q. \tag{3}$$

La tabla I muestra la tasa de modelos con la menor distancia  $d(T^m, T^n)$  aplicada al descriptor CC de 50 modelos y sobre la región de la boca. De esta forma, por ejemplo, en la primera fila de la tabla I se indica que de las 50 imágenes correspondientes a la expresión de enfado (AN), las primeras dieciocho posiciones fueron ocupadas por imágenes de AN por lo que se tiene una tasa de aciertos del 36 %. El descriptor CC fue extraído únicamente sobre la región de la boca debido a la gran cantidad de cálculos que deben ser realizados para obtenerlo; en promedio el rostro tiene 10000 vértices, si se quisiera extraer el descriptor para todo el rostro el sistema de reconocimiento sería inviable.

**Tabla I.** Aciertos del descriptor Cone Curvature sobre la región de la boca

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

En [16] se realizó un análisis similar al presentado en la tabla I pero para el descriptor DESIRE, dicho análisis muestra que para este tipo de clasificación por norma-infinito el descriptor CC presenta un desempeño más bajo.

### 3.3. Reducción de dimensionalidad

El descriptor implementado tiene alta dimensionalidad lo que puede dificultar el entrenamiento de algunos clasificadores y, en general, consume gran cantidad de recursos computacionales. Además, en muchas ocasiones las componentes del descriptor pueden estar correlacionadas o no aportar información importante para el problema de clasificación. Por lo anterior, con el objetivo de reducir la dimensión de los vectores característicos del descriptor, se emplea la técnica del análisis de componentes principales (PCA) y, de acuerdo a [22], el uso de las dos componentes principales asegura más del 90 % de la variabilidad de los gestos. Con el uso de estas dos componentes principales como características se repitió el experimento de cálculo de las distancias, obteniendo los resultados presentados en la tabla II.

**Tabla II.** Aciertos del descriptor Cone Curvature luego de aplicar PCA

Expresión	Casos	Porcentaje
AN	40	80 %
DI	43	86 %
FE	27	54 %
HA	33	66 %
SA	30	60 %
SU	50	100 %

Comparando las tablas I y II se evidencia el mejor desempeño en la clasificación de los gestos cuando se usan las características más relevantes proporcionadas por el PCA.

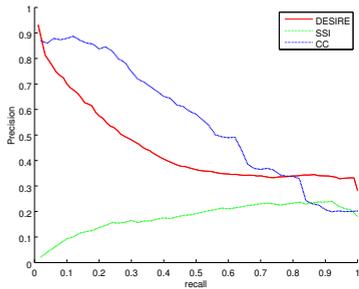
Con el objetivo de verificar la similitud entre imágenes de una misma expresión, se dividió el conjunto de 100 imágenes por gesto en dos subconjuntos de 50 imágenes cada uno, a estos subconjuntos se le calcularon los vectores característicos por cada expresión y se calculó la distancia entre promedios. Los resultados se encuentran en la tabla III, donde se aprecia la efectividad del descriptor CC, solamente la expresión de miedo (FE) no presentó la menor distancia con su correspondiente, sin embargo esta fue la segunda menor distancia.

### 3.4. Análisis del desempeño mediante curvas precision-Recall

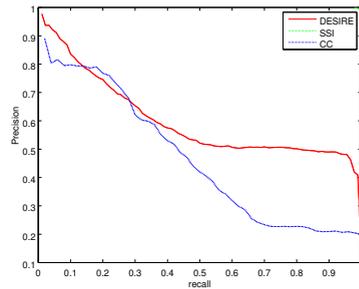
La precisión se define como la fracción de imágenes recuperadas que son relevantes entre el total de imágenes. Recall es la fracción del número de imágenes relevantes que han sido correctamente recuperadas, mide la capacidad de recuperar las imágenes que son relevantes dentro de todo el conjunto [24]. Se obtuvo la curva precision-Recall del descriptor CC y se comparó con las curvas realizadas en [16], para evaluar su desempeño contra los descriptores DESIRE [24] y SSI [18].

**Tabla III.** Distancias entre promedios para el descriptor CC con reducción de dimensionalidad sobre la región de la boca

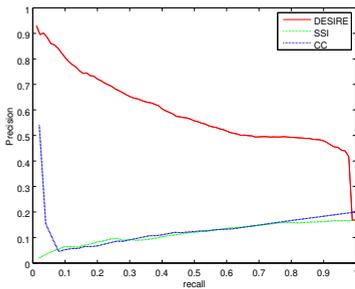
	AN	DI	FE	HA	SA	SU
AN	<b>0,1037</b>	0,1097	0,1937	0,2549	0,1995	2,7498
DI	0,0835	<b>0,0602</b>	0,1061	0,1673	0,1119	2,7994
FE	0,1806	0,0895	0,0573	0,0722	<b>0,0147</b>	2,7989
HA	0,2664	0,1753	0,0865	<b>0,0703</b>	0,0711	2,7697
SA	0,1787	0,1089	0,1056	0,0894	<b>0,0434</b>	2,7506
SU	2,8417	2,8478	2,8445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



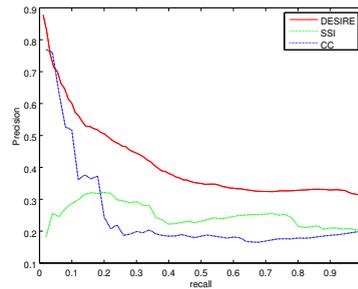
(a) Expresión enojo (AN).



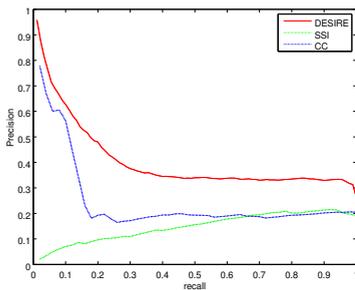
(b) Expresión disgusto (DI).



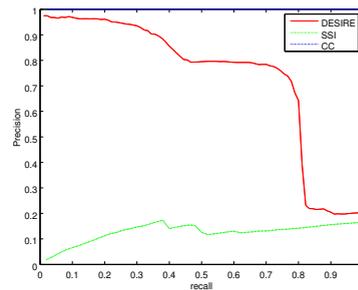
(c) Expresión miedo (FE).



(d) Expresión alegría (HA).



(e) Expresión tristeza (SA).



(f) Expresión sorpresa (SU).

**Figura 7.** Curvas Precision-recall para cada uno de los descriptores en los seis gestos sobre la región de la boca luego de aplicar PCA.

La figura 7 muestra las curvas de precision-recall de los tres descriptores para cada gesto, luego de realizar reducción de dimensionalidad. Se aprecia que el descriptor DESIRE tiene la mejor efectividad, en general, respecto a la recuperación de datos relevantes. Aunque para el caso de la expresión DI la mayor efectividad la tuvo el descriptor SSI y para la expresión SU el mejor comportamiento lo tiene CC.

### 3.5. Clasificación mediante análisis discriminante

Se generó un descriptor de diecisiete características  $D_T = [x_1, x_2, \dots, x_{17}]$ , conformado por las diez características del descriptor DESIRE, cinco características del descriptor SSI y las dos características arrojadas al aplicar PCA al descriptor CC. El Análisis Discriminante Lineal (LDA) permite tomar dicho vector y reducir la dimensionalidad, seleccionando un número inferior de características formadas como combinación lineal de las originales, las cuales proporcionan la mayor información discriminante. Luego de realizar LDA y aplicar una regresión lineal a las características obtenidas, para analizar el aporte de cada una de las características originales, se obtuvieron cinco funciones discriminantes es decir, el número de clases (seis gestos) menos uno, que separan linealmente cada una de las clases. Este análisis mostró que las características menos relevantes son las correspondientes al descriptor SSI.

El análisis LDA permite determinar cuales de las características son las más relevantes para la discriminación de las clases. Sin embargo, las funciones obtenidas también permiten separar linealmente los gestos, resultado que se usó para clasificación de gestos empleando el descriptor compuesto por los tres descriptores ( $D_T$ ). La matriz de confusión para dicha clasificación se presenta en la tabla IV. Se observa el buen desempeño para los gestos AN, DI, SA y SU, el menor desempeño se obtuvo en el gesto FE.

**Tabla IV.** Matriz de confusión (en porcentajes) de la clasificación empleando LDA y el descriptor  $D_T$

	LDA					
	AN	DI	FE	HA	SA	SU
AN	100	0	0	0	0	0
DI	0	100	0	0	0	0
FE	28	0	50	0	22	0
HA	16.67	0	0	83.33	0	0
SA	0	0	0	0	100	0
SU	0	0	0	0	0	100

## 4. Resultados

### 4.1. Clasificación no lineal

El análisis de similitud estudiado en la sección anterior permitió conocer cuál descriptor presenta mejor desempeño en la tarea de reconocimiento de gestos, mediante una análisis lineal de distancias entre vectores característicos. En esta sección se emplean dos clasificadores

no lineales: una red neuronal artificial y un clasificador bayesiano, para evaluar el desempeño del descriptor CC en la tarea de reconocer gestos faciales, se compara su desempeño contra los resultados obtenidos con los descriptores DESIRE y SSI reportados en [16]. Los clasificadores fueron entrenados con un conjunto de 50 imágenes de cada gesto y se evaluó su desempeño con otro grupo de veinte imágenes por gesto en la tabla V se presenta la matriz de confusión obtenida con los resultados de clasificación con los dos clasificadores usando solo el descriptor CC.

**Tabla V.** Matriz de confusión (en porcentajes) de cada clasificador para el descriptor Cone Curvature

	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

Según lo observado en la tabla V, 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%.

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 los tres descriptores, es decir, el CC implementado en este trabajo junto con el DESIRE y el SSI implementados en [16] y realizar de nuevo la clasificación con los clasificadores mencionados. Para este experimento se obtuvieron los resultados presentados en la tabla VI, en la que se resumen las diagonales de las correspondientes matrices de confusión incluyendo las del clasificador con LDA (tabla IV). Se aprecia que solo hubo incremento en la tasa de verdaderos positivos para el gesto FE, con respecto a los resultados logrados con el descriptor CC. Además, las columnas 2 y 3 de la tabla VI muestran la capacidad discriminante no lineal del descriptor conformado por la concatenación de los tres descriptores (CC, DESIRE y SSI).

**Tabla VI.** Porcentaje de correcta clasificación combinando los tres descriptores (diagonal matriz de confusión)

Expresión	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 %

Finalmente, un tercer experimento consistió en construir un vector conformado por los descriptores de mejor desempeño individual, es decir, DESIRE y CC, y con las características

más relevantes según el análisis discriminante lineal. Los resultados obtenidos se presentan en la tabla VII, en la que se observa que para el reconocimiento de los gestos AN y HA con la red neuronal, se presentó un descenso en el desempeño del 19,64 % y el 25 %, respectivamente. El reconocimiento de los demás gestos presentó una tasa de reconocimientos superior al 90 %.

**Tabla VII.** Porcentaje de correcta clasificación usando un descriptor que combina los descriptores DESIRE y CC.

Expresión	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 %

## 4.2. Discusión

Tanto el descriptor *Cone Curvature* (CC) implementado en este trabajo, como los descriptores DESIRE y *Spherical Spin image* han demostrado ser eficientes en el reconocimiento de objetos en 3D [18], [22], [24], dado este antecedente se evaluó su desempeño frente al reconocimiento de expresiones faciales. De acuerdo a los resultados de la tabla V, el descriptor CC tiene un aceptable desempeño en el reconocimiento de expresiones, salvo en la expresión de miedo (FE), lo cual puede atribuirse a la menor variabilidad en la curvatura de la región de la boca para esta expresión, en comparación con las demás. Los resultados de la tabla VII muestran que esta limitante del descriptor CC es corregida cuando se combina con el descriptor DESIRE, el cual tiene mejor desempeño que el descriptor *Spherical Spin Image* [16].

Por último, en la tabla VIII se comparan los resultados obtenidos con los reportados en [16]. Se aprecia que para todos los clasificadores se obtiene una mayor tasa de reconocimiento de los gestos faciales cuando se combinan los descriptores DESIRE y CC, que la tasa de reconocimiento obtenida solo con el descriptor DESIRE. El promedio relacionado en la tabla VIII corresponde al calculado para cada columna de la tabla VII.

**Tabla VIII.** Tasa de reconocimiento promedio

	Clasificador	Promedio de reconocimiento
DESIRE [16]	RNA	88.16 %
	Bayesiano	91.11 %
	LDA	80.5 %
DESIRE + CC	RNA	90.3 %
	Bayesiano	97.2 %
	LDA	95.8 %

## 5. Conclusiones

Se implementó el descriptor Cone Curvature sobre los modelos de rostros de una base de datos de imágenes 3D en la región de la boca, se evaluó su desempeño frente al reconocimiento de gestos faciales encontrando que tiene un buen desempeño previo entrenamiento de un clasificador basado en redes neuronales artificiales y un clasificador Bayesiano. El descriptor fue capaz de reconocer seis gestos con un desempeño promedio del 76.7 % usando el primer clasificador y del 78.88 % usando el segundo.

Al combinar el descriptor CC con otros descriptores implementados en trabajos anteriores se obtuvo un incremento del 13.6 % en la tasa de reconocimientos respecto al CC con la red neuronal, y del 18.3 % con el clasificador Bayesiano, lo cual permite concluir que un vector de características conformado por los descriptores *Cone Curvature* y DESIRE puede hacer parte de un sistema de reconocimiento con un alto grado de precisión. No obstante el buen desempeño de los descriptores, una limitante encontrada tiene que ver con el tiempo de cálculo del descriptor CC, cuyo promedio para la región de la boca (aproximadamente 100 vértices) es de 25 minutos, lo cual lo haría poco viable para un sistema actuando en tiempo real; sin embargo, una implementación de los algoritmos en paralelo podría disminuir este tiempo de cómputo.

Aunque el análisis se realizó calculando los descriptores sobre la región de la boca, no solo porque se reduce considerablemente el tiempo de cálculo, sino porque además esta región es la de mayor variabilidad cuando se cambia la expresión, podría considerarse el análisis sobre otras regiones que involucren menos puntos, lo cual ayudaría a reducir los tiempos. La segmentación de la región de la boca se realizó de forma manual con la ayuda del software Meshlab [26], 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 automático.

## Referencias

- [1] P. Ekman *et al.*, “How do we determine whether judgments of emotion are accurate?,” in *Emotion in the Human Face* (P. E. V. F. Ellsworth, ed.), vol. 11 of *Pergamon General Psychology Series*, pp. 15 – 19, Pergamon, 1972.
- [2] M. Rosenblum, Y. Yacoob, *et al.*, “Human emotion recognition from motion using a radial basis function network architecture,” in *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pp. 43–49, Nov 1994.
- [3] N. Tsapatsoulis, Y. Avrithis, and S. Kollias, “On the use of radon transform for facial expression recognition,” in *Proc Int Conf on Information Systems Analysis and Synthesis (ISAS), Orlando, FL, August, 1999.*
- [4] R. Pericet-Camara, G. Bahi-Vila, J. Lecoeur, and D. Floreano, “Miniature artificial compound eyes for optic-flow-based robotic navigation,” in *Information Optics (WIO), 2014 13th Workshop on*, pp. 1–3, July 2014.
- [5] J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, “Subtly different facial expression recognition and expression intensity estimation,” in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pp. 853–859, IEEE, 1998.
- [6] I. A. Essa, T. Darrell, and A. Pentland, “Tracking facial motion,” in *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pp. 36–42, IEEE, 1994.

- [7] A. Cerón Correa, A. E. Salazar Jiménez, and F. A. Prieto Ortiz, "Reconocimiento de rostros y gestos faciales mediante un análisis de relevancia con imágenes 3d," *Revista de Investigación, desarrollo e innovación*, vol. 4, no. 1, 2014.
- [8] L. Xiaoli, R. Qiuqi, and A. Gaoyun, "Analysis of range images used in 3d facial expression recognition," in *TENCON 2013 - 2013 IEEE Region 10 Conference (31194)*, pp. 1–4, Oct 2013.
- [9] X. Li, Q. Ruan, and G. An, "3d facial expression recognition using delta faces," in *Wireless, Mobile and Multimedia Networks (ICWMMN 2013), 5th IET International Conference on*, pp. 234–239, Nov 2013.
- [10] C. Conde, L. J. Rodríguez-Aragón, and E. Cabello, "Automatic 3d face feature points extraction with spin images," in *Image Analysis and Recognition*, pp. 317–328, Springer, 2006.
- [11] A. Ceron, A. Salazar, and F. Prieto, "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 on*, pp. 452–457, IEEE, 2010.
- [12] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3d assisted face recognition: A survey of 3d imaging, modelling and recognition approaches," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pp. 114–114, IEEE, 2005.
- [13] Z. Guo-Feng *et al.*, "New research advances in facial expression recognition," in *Control and Decision Conference (CCDC), 2013 25th Chinese*, pp. 3403–3409, May 2013.
- [14] H. Tang and T. S. Huang, "3d facial expression recognition based on properties of line segments connecting facial feature points," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pp. 1–6, IEEE, 2008.
- [15] J. Wang *et al.*, "3d facial expression recognition based on primitive surface feature distribution," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 2, pp. 1399–1406, IEEE, 2006.
- [16] J. S. Rodríguez A and F. Prieto, "Analyzing the relevance of shape descriptors in automated recognition of facial gestures in 3d images," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 8650, International Society for Optics and Photonics, 2013.
- [17] C. Cerrada, A. Adan, J. A. Cerrada, M. Adan, and S. Salamanca, *Experiences in Recognizing Free-Shaped Objects from Partial Views by Using Weighted Cone Curvatures*. INTECH Open Access Publisher, 2011.
- [18] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 5, pp. 433–449, 1999.
- [19] D. V. Vranić and D. Saupe, "3d model retrieval," *Proc. SCCG 2000*, pp. 3–6, 2004.
- [20] A. Adán, C. Cerrada, and V. Felíu, "Modeling wave set: Definition and application of a new topological organization for 3d object modeling," *Comput. Vis. Image Underst.*, vol. 79, pp. 281–307, Aug. 2000.
- [21] A. Adán, M. Adán, S. Salamanca, and P. Merchán, "Using non local features for 3d shape grouping," in *Structural, Syntactic, and Statistical Pattern Recognition* (N. da Vitoria Lobo, T. Kasparis, F. Roli, J. Kwok, M. Georgiopoulos, G. Anagnostopoulos, and M. Loog, eds.), vol. 5342 of *Lecture Notes in Computer Science*, pp. 644–653, Springer Berlin Heidelberg, 2008.
- [22] A. Adan and M. Adan, "A flexible similarity measure for 3d shapes recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 11, pp. 1507–1520, 2004.
- [23] G. Passalis, T. Theoharis, and I. A. Kakadiaris, "Ptk: A novel depth buffer-based shape descriptor for three-dimensional object retrieval," *The Visual Computer*, vol. 23, no. 1, pp. 5–14, 2007.
- [24] D. V. Vranic, "Desire: a composite 3d-shape descriptor," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pp. 4–pp, IEEE, 2005.
- [25] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pp. 211–216, IEEE, 2006.
- [26] P. Cignoni, M. Corsini, and G. Ranzuglia, "Meshlab: an open-source 3d mesh processing system," *Ercim news*, vol. 73, pp. 45–46, 2008.

---

### Julián S. Rodríguez A.

Ingeniero Electrónico de la Escuela Colombiana de Ingeniería "Julio Garavito", de Bogotá, Colombia. Obtuvo su título de Maestría en Ingeniería en la Universidad Nacional de Colombia. Actualmente se desempeña como docen-

te investigador en el programa de ingeniería Mecatrónica en la Universidad de San Buenaventura Bogotá. e-mail: jrodriguez@usbbog.edu.co

---

### **Flavio Prieto**

Ingeniero Electrónico de la Universidad Distrital Francisco José de Caldas, Físico de la Universidad Nacional de Colombia Sede Bogotá, Magister en Ingeniería Eléctrica de la Universidad de los Andes, DEA y Doctor del Institut National des Sciences Appliquées (INSA) Lyon, Francia y PhD de l'École de Technologie Supérieure (ÉTS), Université du Québec, Montreal, Canadá. Actualmente se desempeña como docente de la Universidad Nacional de Colombia y sus áreas de interés son: sistemas de visión artificial (2D y 3D), Diseño electrónico y la Filosofía de la tecnología. e-mail: fprieto@unal.edu.co