

RECONOCIMIENTO DE ROSTROS Y GESTOS FACIALES MEDIANTE UN ANÁLISIS DE RELEVANCIA CON IMÁGENES 3D

FACE AND GESTURE RECOGNITION BY USING A RELEVANCE ANALYSIS WITH 3D IMAGES

Alexander Cerón Correa¹
Augusto Salazar²
Flavio Augusto Prieto Ortiz³

Recibido: 1 de octubre de 2013
Aceptado: 20 de diciembre de 2013

Resumen

El reconocimiento facial tridimensional busca subsanar las falencias que presentan los métodos basados en imágenes bidimensionales. Este tipo de reconocimiento tiene la ventaja de que las representaciones no son afectadas por cambios en la iluminación, dado que viene dada como una nube de puntos o una malla 3D donde la geometría juega un papel crucial. En este trabajo se presenta un sistema de reconocimiento de rostros, que utiliza un conjunto de descriptores de forma 3D, seleccionados a partir de un análisis de relevancia mediante coeficientes de Fisher en diferentes regiones del rostro que hacen parte de un modelo antropométrico del rostro. Se realizó un conjunto de experimentos para reconocer individuos e identificar sus expresiones y género a partir del análisis de relevancia planteado. Los resultados obtenidos muestran que la elección de características utilizando un análisis de relevancia incrementa el rendimiento del sistema de reconocimiento.

Palabras clave: análisis de relevancia, aprendizaje de máquina, descriptores de forma 3D, reconocimiento de rostros 3D, segmentación 3D.

Abstract

The 3D face recognition technique aims to reduce the flaws that present the bi-dimensional based methods. This kind of recognition method has the advantage to be invariant to illumination changes because the faces are represented as a points cloud or a 3D mesh where geometry is the most remarkable feature. In this research work we present a recognition system that uses a set of 3D shape descriptors that were selected from a relevance analysis by using the Fisher coefficients in different regions of the face which are part of an anthropometric face model. A set of experiments for face, expression, and gender recognition were performed by using the relevance analysis proposed. The obtained results show that the relevance analysis increases the performance in face recognition systems.

Key words: 3D face recognition, 3D segmentation, 3D shape descriptor, machine learning, relevance analysis.

¹Ingeniero de Sistemas, MSc., Profesor Asociado, Facultad de Ingeniería, Universidad Militar Nueva Granada, Bogotá, Colombia. Estudiante de Doctorado en Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia.
E-mail: alexander.ceron@unimilitar.edu.co

²Ingeniero Electrónico, MSc. y candidato a PhD, Grupo de Automática y Electrónica, Instituto Tecnológico Metropolitano. Medellín.
E-mail: aesalazarj@unal.edu.co

³Ingeniero Electrónico, PhD en ingeniería, Profesor Titular, Facultad de Ingeniería, Universidad Nacional de Colombia - Sede Bogotá.
E-mail: faprieto@unal.edu.co

1. Introducción

Los modelos tridimensionales del rostro dan gran cantidad de la información sobre la morfología del mismo, la cual puede ser extraída por diversos métodos. Entre ellos se pueden mencionar los basados en modelos de superficies cuádricas que buscan clasificar los puntos de una superficie como planar o no planar, de ahí se estima si puede ser esférica, cilíndrica o cónica (Flynn & Jain, 1988), otros permiten obtener el perfil del rostro en imágenes de rango aprovechando la simetría del mismo (Cartoux et al., 1989).

Por otra parte, existen descriptores basados en las curvaturas que están presentes en cualquier objeto 3D, los cuales han mostrado gran efectividad ya que la curvatura como propiedad de superficie local tiene la capacidad de ser invariante al punto de vista. Los descriptores de curvatura, el cual es un concepto proveniente de la geometría diferencial, permiten identificar ocho formas fundamentales: hiperbólica silla valle, parabólica cóncava hacia arriba, elíptica cóncava hacia abajo, hiperbólica mínima, plana, hiperbólica silla arista, parabólica cóncava hacia arriba y elíptica cóncava hacia arriba (Leal, 2007).

A nivel general se pueden mencionar algunos trabajos relacionados con el análisis y reconocimiento del rostro mediante imágenes 3D. En (Gordon, 1991), se presenta un reconocimiento de rostros basado en mapas de profundidad y la medición de la curvatura principal de superficies sobre un conjunto de datos específico. El cálculo de descriptores de superficies generales se basa en la curvatura y el cálculo de características del rostro basados en conocimiento a priori sobre la estructura del mismo. En (Lee & Milios, 1990), se busca una correspondencia entre imágenes en regiones convexas, haciendo uso de la curvatura gaussiana y de un mapeo para obtener imágenes extendidas Gaussianas. En (Lee, 2005) se presenta una evaluación de algoritmos de reconocimiento de rostros mediante imágenes de rango y se propone el uso de la información de la curvatura, la

cual puede ser adquirida mediante enfoques geométricos en el proceso de proyección lineal.

Pese a que en varios trabajos se realiza detección de características con este tipo de descriptores, no es muy común encontrar aquellos donde se realizan comparaciones cuantitativas entre ellos. Dentro de los trabajos más relacionados con el reconocimiento de rostros en imágenes de rango utilizando descriptores de forma, se puede contar el de (Moreno-Díaz, 2004), donde se hace uso de una segmentación basada en los signos de las curvaturas media y gaussiana, para encontrar las regiones de interés y así caracterizarlas posteriormente, haciendo uso de un conjunto de descriptores. Adicionalmente se realiza una clasificación de 86 descriptores aplicados en las regiones encontradas, mediante el coeficiente de Fisher (Duda et al., 1998), (Bishop, 2006). Los descriptores usados fueron áreas de regiones, relaciones entre áreas, media de áreas, distancias entre centros de masas de regiones, media de distancias simétricas, ángulos entre centros de masas de regiones, media de ángulos de regiones, promedio y varianza de la curvatura media y gaussiana de los puntos de una región y líneas de la cara. Los resultados obtenidos en el trabajo en mención muestran que los mejores desempeños en las técnicas utilizadas ocurren en imágenes de pose frontal. Sin embargo, también es bueno notar que el autor sólo utilizó una segmentación basada en los signos de las curvaturas media y gaussiana, para encontrar las regiones de interés.

En este trabajo se presenta un sistema de reconocimiento de rostros, que utiliza un conjunto de descriptores de forma 3D, seleccionados a partir de un análisis de relevancia mediante coeficientes de Fisher en diferentes regiones del rostro que hacen parte de un modelo antropométrico del rostro.

La organización del artículo es la siguiente, primero se comenzará hablando de los descriptores de forma utilizados, los resultados de un análisis de relevancia haciendo uso de un

modelo antropométrico del rostro. Luego se hablará de los elementos que hacen parte de un sistema de reconocimiento de rostros. Posteriormente se describirá la metodología en la cual se encuentra el marco experimental. Finalmente se mostrarán los resultados en diferentes pruebas de reconocimiento de rostros e identificación de características, las conclusiones y trabajo futuro.

1.1 Descriptores de forma en imágenes 3D

Los descriptores de forma son un conjunto de herramientas matemáticas y procedimentales que permiten obtener diversas medidas sobre los contornos y las estructuras morfológicas de los modelos o imágenes 3D. Dentro de los descriptores de forma existen diferentes tipos: los que están basados en los modelos de superficies cuádricas que buscan clasificar puntos dentro de las diferentes formas geométricas en 3D (planos, esferas, cilindros, conos, entre otros); los que están basados en curvaturas principales, los que están basados en transformaciones de información 3D a 2D como es el caso de las imágenes spin que pueden ser calculadas a diferentes resoluciones, los multi-escala, entre otros.

Descriptores basados en curvaturas: mediante el análisis de los valores de curvatura es posible extraer el plano de simetría del rostro (Cartoux et al., 1989), detectar las regiones de los ojos y la boca (Gordon, 1991; Colombo et al., 2006), segmentar un modelo facial como entrada de un sistema de reconocimiento (Tanaka et al., 1998; Hallinan et al., 1999) o detectar diversas marcas sobre la superficie facial (Irfanoglu et al., 2004; Deo & Sen D, 2005).

Otras aproximaciones combinan la información de curvatura con otras características obtenidas de información 2D (Xue & Ding, 2006), y/o un conocimiento a priori de la geometría del rostro (Sun & Yin, 2008). El proceso de estimación de las curvaturas principales se puede observar en (Taubin, 1995).

Se pueden encontrar una amplia gama de descriptores basados en las curvaturas principales utilizados en análisis y reconocimiento del rostro, los cuales se mencionarían a continuación. Con base en las curvaturas principales k_1 , k_2 , se han propuesto otros descriptores como el índice de forma S_I (Koenderink, 1992).

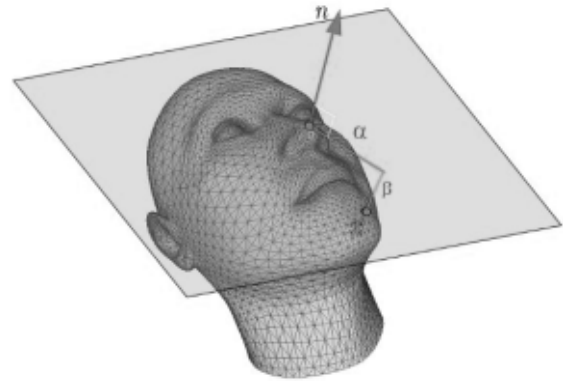


Figura 1. Generación de imágenes spin

Para un punto p sobre una superficie, el S_I se define como:

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \quad (1)$$

El S_I ha sido usado para localizar características faciales en (Lu et al., 2004; Colbry, 2005; Colbry, 2006; Guangpeng & Yunhong, 2007). Otro descriptor es la curvatura (curvedness) definida como:

$$R(p) = \sqrt{(k_1^2(p) + k_2^2(p))} / 2 \quad (2)$$

Representa la cantidad de curvatura en una región, lo cual hace posible la percepción de variación en escala de la forma de los objetos. R ha sido utilizado como criterio para segmentación automática de mallas triangulares (Jagannathan & Miller, 2007).

Imágenes spin: la imagen spin (Johnson, 1997) es un descriptor local, calculado en un punto orientado (p, n) (un punto y su normal respectivamente), que codifica 2 de las 3 coordenadas de un sistema cilíndrico en su

vecindad. La imagen spin X para un punto p en una superficie, es un histograma 2D en el cual cada píxel es un *bin* (unidad de almacenamiento de información) que almacena el número de vecinos que están a una distancia α desde n y a una profundidad β desde su plano tangente P (ver figura 1).

El parámetro α puede ser estimado como la distancia perpendicular desde n hasta el punto x . El segundo parámetro es la distancia relativa paralela entre n y x . La representación matemática de estos parámetros conocidos como coordenadas spin es:

$$\alpha = \sqrt{\|x - p\|^2 - (n \cdot (x - p))^2}, \quad (4)$$

$$\beta = n \cdot (x - p)$$

En (Dinh & Kropac, 2006) se muestra que imágenes spin obtenidas mediante diferentes resoluciones, pueden ser utilizadas para emparejar superficies con densidades de puntos no uniformes, mediante un proceso denominado imágenes spin multi-resolución.

Descriptores multi-escala: Un enfoque de obtención de características multi-escala que resuelve el inconveniente de trabajar con mallas de diferente resolución, fue propuesto en (Ho, 2008), el cual se basa en la obtención del índice de forma a múltiples escalas, ajustando la superficie a vecindades de diferente tamaño, para aplicaciones de registro. Este método busca medir la variación local de superficie calculando la variación del índice de forma de cada punto en la vecindad. Los puntos correspondientes a la variación máxima de la superficie son seleccionados como características adecuadas, de acuerdo a los valores de la desviación estándar. En este trabajo se utilizan los siguientes descriptores: curvaturas principales, curvatura media, curvatura gaussiana, índice de forma (shape index), curvacidad (curvedness), las imágenes *spin*, el índice de forma multi-escala y las imágenes *spin* multiresolución.

2. Análisis de relevancia

Gracias a un análisis de relevancia de los siguientes descriptores: curvaturas principales, curvatura media, curvatura gaussiana, índice de forma (shape index), curvacidad (curvedness), las imágenes spin y el índice de forma multi-escala se determinó cuales ofrecen información más relevante en ubicaciones específicas del rostro mediante la utilización del análisis discriminante de Fisher. Se realizó un análisis de los descriptores planteados en diferentes áreas del rostro (ojos, nariz, mejillas, boca, quijada), el cual es denominando análisis local como el presentado anteriormente en (Cerón et al., 2010), pero con un conjunto mayor de descriptores, obteniendo un ranking de los más relevantes.

2.1 Modelo del rostro

Cada modelo del rostro es segmentado en 28 regiones (ver figura 2), las cuales corresponden a regiones anatómicas del tejido blando del rostro, usadas para describir una lesión en medicina forense y/o para planear una cirugía en muchos otros contextos médicos. El contorno de una región es definido mediante curvas de Bézier, cada una con dos puntos de quiebre y dos puntos de control. La implementación de la curvas de Bézier se realizó en C++, tomando como base la presentada en (Hearn & Baker, 2003) y se adicionó al software MeshLab. La plantilla completa está compuesta por 68 contornos de regiones, 46 puntos de quiebre, 136 puntos medios y 338 triángulos; tanto los puntos de quiebre como los puntos medios son vértices de la plantilla del rostro 3D como se observa en la figura 3 (Salazar & Prieto, 2010).

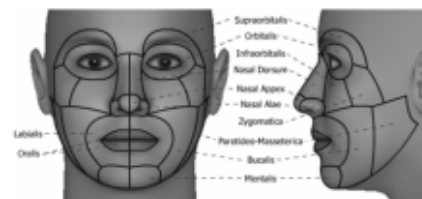


Figura 2. Regiones del rostro humano que se analizan en este trabajo

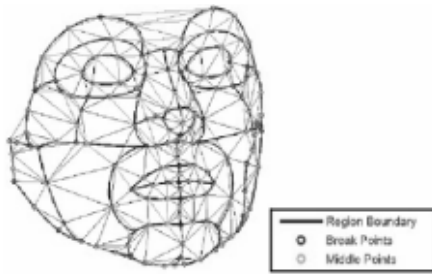


Figura 3. Modelo parametrizado del rostro

2.2 Ranking de descriptores por área del rostro

Gracias a las pruebas anteriores, se definieron cuáles son los descriptores más adecuados en los contornos y regiones del rostro de las 5 áreas propuestas para el análisis local, estos se pueden observar en la tabla 1. Estas características pueden ser evaluadas en un sistema de reconocimiento de rostros, mediante imágenes 3D para cada una de las regiones, así como en mediciones de antropometría facial.

Área	Contornos	Regiones
Ojos	k_1, S_l, R, H, k_2	H, k_1, R, S_l, S_{ims}
Nariz	k_1, S_l, H, R	k_1, S_l, H, R
Mejillas	k_1, S_l, H, R	k_1, S_l, H, R
Boca	S_{ims}, K, R, k_2	k_1, S_{ims}, K, R
Quijada	H, k_2, R, K	R, S_{ims}, k_2, k_1

Tabla 1. Descriptores de acuerdo a las regiones del rostro

3. Sistema de reconocimiento de rostros

Un sistema de reconocimiento de rostros está compuesto de un conjunto de procesos, que permiten alimentar una base de datos con una o más imágenes de un conjunto de individuos etiquetados, y realizar un entrenamiento mediante técnicas de aprendizaje de máquina, que permitan identificar a los individuos con imágenes diferentes a las que fue entrenado. Un sistema más sofisticado podría determinar características propias de los individuos, como género, edad, expresión facial, entre otros.

Un sistema de reconocimiento de rostros consta de las siguientes etapas: adquisición, pre procesamiento, registro, segmentación,

extracción de características y reconocimiento (clasificación o interpretación). Se puede decir que estas etapas permiten generar una base de conocimiento propia del sistema, en la figura 4 se muestran las principales etapas del sistema de reconocimiento.

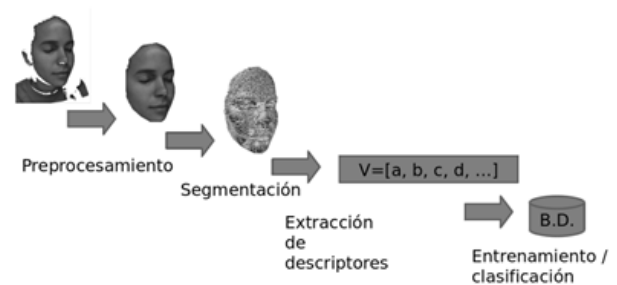


Figura 4. Esquema del sistema de reconocimiento

3.1 Adquisición

El objetivo del sensor es crear una nube de puntos a partir de muestras geométricas tomadas en la superficie del objeto. A partir de estos puntos se puede generar una malla 3D. Un escáner 3D al igual que una cámara tradicional tiene un campo de visión en forma de cono, pero puede obtener información acerca de superficies sin iluminación. La etapa de adquisición de la imagen depende de las características propias del sensor utilizado (tipo de radiación que emite y recibe), distancia de captura, capacidad de apertura, posición. También depende del punto de referencia y distancia al objeto.

3.2 Preprocesamiento

En esta etapa se deben considerar los siguientes aspectos: filtrado, mejora y refinamiento de la malla, los cuales facilitan la manipulación de las imágenes y la obtención de descriptores con mayor confiabilidad. Dado que las imágenes 3D obtenidas mediante sensores de rango contienen ruido, se hace necesario el uso de técnicas de filtrado para las mallas, que mejoran los resultados obtenidos. Se han logrado buenos resultados al filtrar las imágenes con los filtros Laplaciano y Taubin los cuales se encuentran incorporados al software libre MeshLab.

Adicionalmente, existen mallas en las que se presentan discontinuidades o huecos debido al proceso de captura, estas se generan debido a limitaciones del sensor para capturar todas las estructuras morfológicas presentes en el rostro humano, como es el caso de la región bajo la nariz o las cejas, donde por lo general se generan huecos en las mallas 3D y se pierde información real, que puede ser completada mediante diversas técnicas que incluyen interpolación. Por otra parte, es posible que tanto en la captura como en los pasos anteriores se generen vértices repetidos, aislados, caras repetidas, mal conformadas, o con área cero, las cuales deben ser eliminadas para evitar tener mediciones incongruentes en los descriptores.

3.3 Registro

El proceso de registro que implica el ajuste de dos superficies 3D, es un problema que tiene básicamente tres enfoques. (1) Registrar la superficie de búsqueda a la superficie de referencia. (2) Extraer y ajustar representaciones centradas en el observador. (3) Extraer y ajustar representaciones centradas en el objeto de las superficies. En algunos trabajos relacionados como en (Moreno-Díaz, 2005), se realiza un ajuste de cada rostro en lo que se conoce como: normalización del rostro en cuanto a la pose. Lo cual tiene como objeto que todas las caras se encuentren en la misma posición y con la misma orientación en el espacio 3D. Luego del proceso anterior, es necesario ajustar una grilla 3D, o un modelo 3D, que permita la segmentación e identificación de áreas y regiones de interés en las imágenes 3D.

En este trabajo se realizó un ajuste de un modelo a cada rostro, lo cual evita modificar cada rostro. En este caso lo que se modifica es el modelo anatómico del rostro planteado, el cual está compuesto de una menor cantidad de vértices (182 puntos) que un rostro real, que puede estar compuesto de 4000 vértices y llegar a más de 54000 vértices, cuando el modelo es de alta

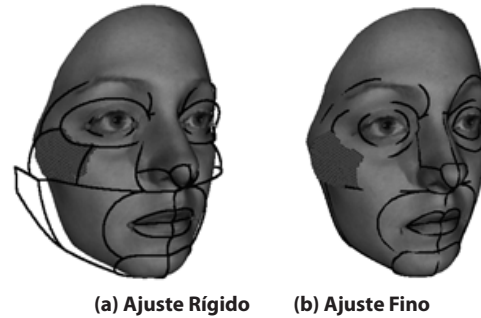


Figura 5. Ajuste de la plantilla y segmentación obtenida

3.4. Segmentación

En esta etapa se realiza una división de las regiones de cada rostro de acuerdo al ajuste del modelo desarrollado en (Salazar & Prieto, 2010). En principio, se realiza la segmentación buscando capturar todos los puntos que se encuentran dentro de las regiones, pero se evalúa que factibilidad tiene el realizar una segmentación basada en un ajuste rígido ver figura 5(a), con el objeto de reducir el tiempo de cómputo que implica el realizar el ajuste fino ver figura 5(b), de cada una de las curvas de Bézier que componen al modelo. Como se puede observar en la figura 5 el resultado de segmentación de la región 11, ver figura 6 (a) tanto con el ajuste rígido como con el fino presenta gran similitud.

3.5 Extracción de características

En este trabajo se desarrollaron procedimientos para extraer características en diferentes partes de rostro. En primer lugar se encontraron los puntos pertenecientes a la malla, más cercanos a los que hacen parte del modelo. Posteriormente se desarrollaron procedimientos que obtenían los puntos que hacen parte de cada contorno del modelo más cercanos a la malla así como procedimientos para encontrar puntos que hacen parte de cada región para lo cual se realizó una segmentación 3D.

Extracción de descriptores en puntos de interés: el proceso de extracción de descriptores de curvaturas k_1 , k_2 , H , K , R y SI en un punto

seleccionado de la malla, se realiza haciendo uso de las librerías que hacen parte del MeshLab, en las cuales se encuentra un algoritmo de obtención de curvaturas principales k_1 y k_2 , basado en (Taubin, 1995).

Las imágenes spin se calcularon para 3 resoluciones, de acuerdo a un esquema multi-resolución, el cual puede ser utilizado como descriptor local al realizar un promedio de la imagen, para lo cual se realizó una implementación propia y se incluyó a un complemento del MeshLab llamado editpickpoints. El índice de forma multi-escala no se utilizó como descriptor local, dado que su algoritmo necesita un conjunto de vértices, por lo tanto se extraerá y se analizará en regiones. En un trabajo anterior enfocado al análisis de descriptores de curvatura, se realizó una obtención manual de descriptores en 24 puntos de interés, en mallas sintéticas y reales, con los cuales se podía obtener otros 29 puntos, gracias a la simetría y las proporciones del rostro, llegando a 53 puntos (Cerón et al., 2010b).

Extracción de descriptores en contornos: se desarrolló un procedimiento para obtener los contornos de cada región, obteniendo los puntos más cercanos a las curvas para todas las 28 regiones, las cuales fueron etiquetadas, como se puede observar en la figura 6(a). Para el cálculo de las curvas de Bézier se define un incremento en su parámetro independiente, inferior a la mitad de la distancia promedio de los puntos de la malla, con el objeto de realizar un muestreo que capture suficiente cantidad de puntos, evitando saltar puntos que pertenezcan a un contorno y realizar muestreos innecesarios. Con estos criterios se obtienen los puntos a evaluar en las mallas, tanto sintéticas como reales, que se consideraron para los experimentos. También se pueden variar los parámetros de los procedimientos, para obtener un muestreo de los puntos en la región. Se realiza una lista de vértices para cada región, denominada contorno, en la cual se almacenan los vértices pertenecientes a la malla que estén más cercanos

a cada una de las curvas que pertenecen al modelo. Este procedimiento se desarrolla siguiendo un orden consecutivo hasta volver al punto inicial, formando una figura cerrada, lo cual puede ser considerado una polilínea en 3D, ver figura 6(b).

Extracción de descriptores en regiones: se desarrolló un procedimiento para obtener los puntos que están dentro de cada una de las 28 regiones acotadas, basada en una búsqueda incremental de puntos al interior de cada región definida en la plantilla, para luego extraer los descriptores que se mencionaron anteriormente. Para obtener los puntos que están dentro de una región, se desarrolló un procedimiento que realiza una búsqueda en dirección, que apunta hacia el punto central de la región. Este punto es definido como el punto perteneciente a la malla más cercano a la distancia media entre dos puntos opuestos de la región en la plantilla. Cuando la región es convexa, el centro de la región cr es cercano al centro geométrico. Cuando la región no es convexa, se puede separar en secciones que sean convexas, y aplicar el mismo procedimiento encontrando puntos centrales en las mismas.

El recorrido se realiza en contornos internos cada vez más pequeños, usando una subdivisión de la distancia entre cada punto del contorno de la región y el centro de la región. La subdivisión depende de la resolución de la malla. El valor de la subdivisión indica la cantidad de incrementos que se deben hacer para llegar al centro de la región. La subdivisión en general se obtiene al dividir la distancia de cada punto del contorno hasta el punto central. Se encontró, empíricamente, que la subdivisión puede ser 15 para imágenes reales y puede reducirse hasta 7 en imágenes sintéticas.

La ventaja de este procedimiento basado en subdivisiones de la distancia, con respecto a uno basado en vecindades, es que no se queda estancado en regiones que presenten discontinuidades o huecos, lo cual es un caso común en imágenes 3D. Otra ventaja es que se

puede disminuir la subdivisión, y realizar un muestreo de puntos en un tiempo mucho menor. Después de tener los puntos pertenecientes a cada región, es posible realizar la extracción de los descriptores planteados, y almacenarlos en un vector de características, de acuerdo al ranking obtenido en el análisis discriminante de Fisher, con el objeto de extraer información relevante para la clasificación a partir de las entidades cuantificables.

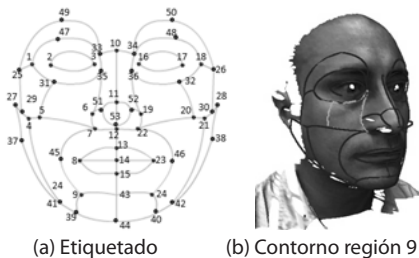


Figura 6. Etiquetado de las regiones y un contorno obtenido

3.6 Entrenamiento

Haciendo uso de técnicas de aprendizaje de máquina como: redes bayesianas, redes neuronales, máquinas de vectores de soporte (SVM), se realiza un entrenamiento del sistema con base en los datos de entrada etiquetados. En una fase inicial se utiliza el vector de características para generar una colección (posiblemente ordenada y estructurada) de descriptores, denominada patrón, el cual define a cada individuo de una base de datos BD. La BD es compuesta de una cantidad de N individuos con M imágenes por individuo.

3.7 Clasificación

En esta etapa se evalúan patrones correspondientes a la información de algún individuo en particular con el objeto de determinar: 1) si el individuo hace parte de la base de datos o no, 2) identificar al individuo y 3) si el individuo posee alguna característica de interés.

4. Diseño del experimento

Para evaluar la utilidad del procedimiento de

extracción de características y el análisis de relevancia se plantea un conjunto de experimentos enfocados al reconocimiento de individuos, género y expresiones.

4.1 Bases de datos

Se utilizaron 10 individuos de la base de datos BU-3DFE (Yin et al., 2006), de los cuales 5 son hombres y 5 son mujeres, y se tienen tres muestras de cada individuo que corresponden a tres expresiones faciales básicas: neutral, enojo y alegría.

4.2 Esquema de reconocimiento

En este caso se realiza el ajuste de la plantilla compuesta de 182 puntos y 26 regiones sobre cada uno de los modelos, se realiza la segmentación para obtener las características de cada región. Es bueno mencionar que aunque anteriormente se definieron 28 regiones, existen 2 regiones laterales que no pueden obtenerse con precisión (estas corresponden a las regiones 19 y 20). Para los individuos en expresión neutral se realizó un ajuste de cada una de las curvas del modelo para que pasen por los contornos de las imágenes 3D. En las imágenes con expresiones de enojo y alegría se utilizó un ajuste rígido para evaluar su funcionamiento en la segmentación, dado que implica un menor tiempo de cómputo que hacer que las curvas pasen exactamente por el rostro. El ajuste rígido planteado consiste en escalar los puntos que componen la plantilla al tamaño de un rostro particular y la segmentación de las regiones se realiza buscando los puntos más cercanos a las regiones definidas dentro de las superficie que hacen parte de cada región. Se evaluaron tres clasificadores: una red bayesiana con un estimador simple y el algoritmo de búsqueda K_2 , una red neuronal y una máquina de vectores de soporte utilizando el algoritmo de entrenamiento SMO (Sequential Minimal Optimization) (Dong et al., 2002).

Para el caso de la red neuronal se utilizó una capa de entrada con un número de neuronas

correspondiente a la cantidad de descriptores multiplicada por la cantidad de puntos o regiones a evaluar, una capa de salida que indica la cantidad de clases (número de individuos (10), género (2), número de expresiones (3)), para la capa intermedia se utilizaron dos alternativas: la primera con la misma cantidad de neuronas de salida y la segunda con el promedio entre el número de neuronas de entrada y salida.

Existen varios métodos para evaluar la capacidad del clasificación, en este caso se realizó una validación cruzada con un pliegue (folding) de 10, lo cual permite tener un indicio del error de generalización.

4.3 Prueba identificación del individuo

Esta prueba se divide en 4 experimentos: 1) identificación a partir de todos puntos característicos y todos los descriptores locales, 2) identificación a partir de puntos y características relevantes, 3) identificación con regiones y 4) identificación con regiones más relevantes y reducción de características.

4.4 Prueba identificación del género

En ésta prueba se etiquetaron todos los individuos como hombre ó mujer, según el caso, con el objeto de analizar la factibilidad de reconocer el género a partir de los descriptores más relevantes extraídos, también, en las regiones más relevantes.

Prueba identificación de la expresión: dado que el data set analizado estaba compuesto de individuos con 3 expresiones: neutral, enojo y alegría. Se etiquetaron las 3 instancias de los 10 individuos según el caso, con el objeto de

encontrar la factibilidad de reconocer una expresión, a partir de los descriptores más relevantes extraídos en las regiones más relevantes.

5. Resultados y análisis

5.1 Identificación de un individuo

Identificación a partir de puntos característicos: en este experimento se evaluaron los puntos que componen a la plantilla, evitando puntos laterales, los cuales no se encuentran presentes en las imágenes procesadas. Los descriptores empleados fueron k_1 , k_2 , H , K , R y S_I y las medias de las imágenes spin obtenidas con 3 diferentes subdivisiones (5, 10,20) μl_5 ; μl_{10} ; μl_{20} . En este caso se cuenta con 168 puntos característicos y 9 descriptores por cada uno llegando a 1512 características por instancia, contando con 10 individuos y 3 instancias de cada uno.

Los resultados de utilizar un clasificador tipo red bayesiana, una red neuronal con 10 neuronas en la capa oculta (igual al de la capa de salida), una red neuronal con 761 neuronas en la capa oculta (el promedio entre neuronas en la capa de entrada y de salida) y una máquina con vectores de soporte usando el algoritmo SMO, se presentan en la tabla 2. En este caso se puede observar que el uso de los 168 puntos para representar el rostro humano, no presenta un rendimiento adecuado para identificar rostros con los clasificadores evaluados, dado que los resultados obtenidos no superan el 60% y pueden tomar un tiempo considerable debido a la cantidad de puntos y de descriptores. Esto también indica que una red neuronal puede tomar un tiempo importante, lo cual la hace inviable para un sistema que tenga que entrenar

Resultados	Red Bayesiana	RN(10)	RN(761)	SMO
Correctamente	53.33% (16)	60% (18)	53.33% (16)	46.66%(14)
Incorrectamente	46.66% (14)	40% (12)	46.66% (14)	53.33% (16)
Error medio absoluto	0.0984	0.1147	0.1175	0.1695
Tiempo en construir en seg.	0.17	6.85	487.99	0.14

Tabla 2. Resultados identificación con 168 puntos y 9 descriptores con diferentes clasificadores

los descriptores en los puntos de una cantidad considerable de imágenes.

Identificación a partir de puntos y características relevantes: en este caso se tomaron solo 62 puntos de los 182 de la plantilla, los cuales corresponden a las regiones de los ojos y la nariz, que son las que tienen coeficientes de Fisher más altos, como se expuso en la tabla 1. Teniendo en cuenta lo obtenido en la tabla 1 se seleccionaron solo 4 descriptores en dichas regiones, correspondientes a k_1, H, R y S_1 . Debido a lo anterior se cuenta con 248 características por individuo, 10 individuos y 3 instancias por individuo. Los resultados de este experimento se pueden observar en la tabla 3, en la columna 2 se presentan los resultados de clasificación haciendo uso de una red bayesiana, en la columna 3 con una red neuronal de 10 neuronas en la capa oculta y en la cuarta columna se presentan los resultados de clasificación haciendo uso de una red neuronal con 133 neuronas en la capa oculta.

Como se puede observar, los tiempos de entrenamiento fueron mucho menores que en la prueba anterior (se disminuyó el tiempo para el entrenamiento de las redes bayesianas en un 94%, un en la RN (10) un 86% en la RN (133) un 97% y se llegó a una mejor clasificación. Adicionalmente, se puede llegar a una mejor clasificación, al aumentar el número de neuronas en la capa oculta, a expensas de gastar más tiempo en el

entrenamiento (12 segundos con 133 neuronas en la capa oculta y 0.92 segundos utilizando solo 10 neuronas en la capa oculta), lo cual hace que se tenga que evaluar más a fondo el uso de redes neuronales para grandes conjuntos de datos.

Identificación a partir de regiones: para realizar este experimento, se utilizaron 26 regiones y 10 descriptores por región. En primer lugar se experimentó con una red bayesiana, posteriormente se utilizó una red neuronal con 10 neuronas en la capa oculta, luego con una red neuronal con 135 neuronas en la capa oculta (la cantidad de neuronas en la capa intermedia se calculó de la misma forma que en las pruebas anteriores) y finalmente se utilizó una máquina con vectores de soporte obteniendo los resultados presentados en la tabla 4.

En este caso no se encuentran resultados superiores de clasificación al método anterior. Sin embargo, se pueden obtener resultados en menor tiempo. Esto indica que al promediar la información de descriptores en una región pueden existir pérdidas de información local, la cual puede ser necesaria para la identificación. En un estudio futuro se espera analizar otras características de las regiones faciales que guarden mayor información sobre el comportamiento de los descriptores en superficie de cada región.

Resultados	Red Bayesiana	RN(10)	RN(761)	SMO
Correctamente	53.33% (16)	60% (18)	53.33% (16)	46.66%(14)
Incorrectamente	46.66% (14)	40% (12)	46.66% (14)	53.33% (16)
Error medio absoluto	0.0984	0.1147	0.1175	0.1695
Tiempo en construir en seg.	0.17	6.85	487.99	0.14

Tabla 3. Resultados identificación con 64 puntos y 4 descriptores con diferentes clasificadores

Resultados	Red Bayesiana	RN(10)	RN(135)	SMO
Correctamente	50% (15)	63.33% (19)	63.33%(19)	60 %
Incorrectamente	50% (15)	36.66% (11)	36.66%(11)	40 %
Error medio absoluto	0.0982	0.1019	0.0933	0.167
Tiempo en construir en seg.	0.01	0.9	11.66	0.32

Tabla 4. Resultados identificación para 26 regiones con diferentes clasificadores

Resultados	Red Bayesiana	RN(10)	RN(57)	SMO
Correctamente	60% (18)	63.33% (19)	76.66% (23)	80% (24)
Incorrectamente	40% (12)	36.66% (11)	23.33% (7)	20% (6)
Error medio absoluto	0.0846	0.0933	0.079	0.162
Tiempo en construir en seg.	0.001	11.66	0.81	0.21

Tabla 5. Resultados identificación para el área de los ojos y nariz con reducción de características con diferentes clasificadores

Identificación con regiones más relevantes y reducción de características

Gracias a los resultados del análisis de relevancia, para realizar el entrenamiento, se tomaron solamente las regiones que tenían los coeficientes de Fisher más altos, las cuales hacen parte de las áreas de los ojos y la nariz. Teniendo en cuenta lo obtenido en la tabla 1 se seleccionaron solo 4 descriptores en dichas regiones, correspondientes a k_1 , H , R y S_1 . En este caso se entrenó una red bayesiana, una red neuronal con 10 neuronas en la capa oculta, una red neuronal con 57 neuronas en la capa oculta, y un SMO obteniendo los resultados de la tabla 5. Aunque en las pruebas anteriores el algoritmo SMO no mostró un buen desempeño, para esta prueba se alcanzó un acierto de 80 %, y un tiempo de construcción del modelo inferior al de las redes neuronales, lo cual indica que puede ser una herramienta a considerar cuando se hace una reducción de características utilizando un análisis de relevancia.

5.2 Identificación de género

En éste caso se desarrolló un experimento

Resultados	Red Bayesiana	RN(2)	RN(53)	SMO
Correctamente	80% (24)	86.66% (26)	86.66% (26)	86.66% (26)
Incorrectamente	20% (6)	13.33% (4)	23.33% (4)	23.33% (4)
Error medio absoluto	0.1962	0.1698	0.1794	0.13%
Tiempo en construir en seg.	0.01	0.27	11.38	0.01

Tabla 6. Resultados identificación de género con diferentes clasificadores

Resultados	Red Bayesiana	RN(2)	RN(53)	SMO
Correctamente	83.33%(25)	86.66% (26)	86.66% (26)	86.66% (26)
Incorrectamente	16.66%(5)	13.33% (4)	23.33% (4)	23.33% (4)
Error medio absoluto	0.1113	0.1494	0.1458	0.2519
Tiempo en construir en seg.	0.06	0.34	11.35	0.07

Tabla 7. Resultados identificación de expresiones faciales con diferentes clasificadores

enfocado a ver la capacidad de discriminancia de los descriptores más relevantes en las regiones más relevantes, para la identificación del género de cada individuo. Teniendo en cuenta lo obtenido en la tabla 1 se seleccionaron solo 4 descriptores en dichas regiones, correspondientes a k_1 , H , R y S_1 . Se utilizaron 4 clasificadores para evaluar la capacidad representativa de los descriptores en términos de permitir la identificación del género. Los clasificadores utilizados fueron: una red bayesiana, una red neuronal con 2 neuronas en la capa oculta, una red neuronal con 53 neuronas en la capa oculta, y una máquina con vectores de soporte utilizando el algoritmo SMO. Los resultados obtenidos se pueden observar en la tabla 6. Al igual que en la prueba anterior el algoritmo SMO mostró un buen rendimiento como clasificador (87%) y en términos de tiempo (disminuyendo un 95% en comparación con caso anterior), lo cual indica que puede ser tenido en cuenta en un sistema de reconocimiento de características faciales en el cual exista un requerimiento de tiempo de ejecución corto, el cual se vuelve difícil de alcanzar cuando se tiene una gran cantidad de individuos.

5.3 Identificación de expresión facial

La última prueba consiste en la identificación de la expresión a partir de los descriptores más relevantes en las regiones más relevantes (ojos y nariz), para lo cual se evaluaron 4 descriptores (k_1 , H , R y S_l). La base de datos BU-3DFE contiene rostros humanos con diferentes expresiones (neutral, enojo, tristeza, alegría, sorpresa, entre otras), en diferentes niveles de cada una de ellas excepto la neutral. Para esta prueba se utilizaron las expresiones: neutral, enojo y alegría (ver figura 7). Las dos últimas en el nivel más bajo de la expresión, es decir ligeramente, enojado o alegre. Los resultados obtenidos con los clasificadores utilizados pueden observarse en la tabla 7, donde se puede observar que existe gran factibilidad de reconocer una expresión facial teniendo en cuenta solamente las regiones consideradas. Se destacan estos resultados dado que los rostros no fueron exactamente ajustados a la plantilla, es decir que se usó un ajuste rígido y se utilizaron individuos con expresiones faciales en el nivel menos acentuado que proporcionaba la base de datos BU-3DFE. Sin embargo, es bueno considerar el desarrollar un estudio enfocado a encontrar cuales son los descriptores y regiones más relevantes para caracterizar expresiones faciales.

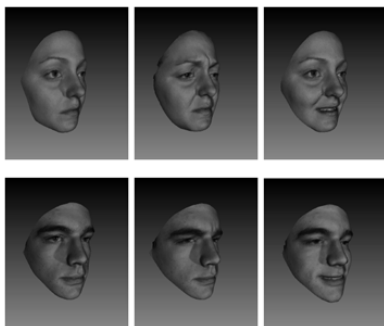


Figura 7. Imágenes de las bases de datos BU-3DFE con expresiones, primera columna expresión neutral, segunda columna expresión de enojo y tercera columna expresión alegría

Conclusiones

Se logró un gran desempeño en el reconocimiento al utilizar el análisis de relevancia para la selección

de descriptores y la utilización de regiones más relevantes con todos los clasificadores en comparación al utilizar todos los descriptores llegando a una mejora de un 12.5% para redes bayesianas, 5% y 43.7% para dos casos de redes neuronales, y un 71% con el algoritmo SMO, en la identificación del individuo. Es de resaltar que este rendimiento fue más alto incluso cuando se utilizó un ajuste rígido entre el modelo del rostro y las imágenes 3D, el cual no es del todo preciso. En los experimentos de reconocimiento realizados se observó que al seleccionar los descriptores más discriminantes se puede lograr aumentar el rendimiento de los clasificadores utilizados y por ende la cantidad de individuos reconocidos correctamente.

Las pruebas de identificación de género y expresión indican que al utilizar un clasificador tipo red neuronal, puede ser suficiente utilizar una cantidad de neuronas en la capa oculta cercana al número de neuronas de salida. Es decir, para identificar la expresión se obtuvo un buen resultado con 3 neuronas en la capa intermedia, las cuales corresponden al número de expresiones a identificar con el número de individuos evaluados.

Estos resultados plantean como trabajo futuro, la obtención de los parámetros óptimos para los clasificadores evaluados en grandes volúmenes de datos. Por ejemplo, la arquitectura de una red neuronal más adecuada para identificar individuos o características del rostro, así como los parámetros que presentan mejor desempeño en la clasificación cuando se usa una máquina de vectores de soporte. Se espera mejorar los resultados obtenidos al usar un proceso de registro más preciso y/o desarrollando plantilla que involucre expresiones faciales. El análisis de relevancia planteado en este trabajo, puede ser extendido al análisis de descriptores en puntos y regiones más relevantes involucradas en la representación de: expresiones faciales, género, raza, edad, entre otros.

Referencias

- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer Science Business + Media, LLC.
- Cartoux J., Lapreste J., & Richetin, M. (1989). Face authentication or recognition by profile extraction from range images. *Proceedings of the Workshop on Interpretation of 3D Scenes*, 194–199.
- Cerón, A., Salazar, A., & Prieto, F. (2010a). Análisis de relevancia de descriptores de forma 3D sobre la superficie del rostro. *Memorias del Encuentro Nacional de Investigación y Desarrollo – Enid*.
- Cerón, A., Salazar, A., & Prieto, F. (2010b). Relevance analysis of 3d curvature-based shape descriptors on interest points of the face. *Image Processing Theory Tools and Applications (IPTA), Proceedings of International Conference on Image Processing Theory, Tools and Applications*. Pp. 452-457. DOI: 10.1109/IPTA.2010.5586721
- Colbry, D., Stockman, G., & Jain, A. (2005). Detection of anchor points for 3D face verification. *Proceedings of IEEE Workshop on Advanced 3D Imaging for Safety and Security*.
- Colombo, A., Cusano, C., & Schettini, R. (2006). 3D face detection using curvature analysis. *Journal of Pattern Recognition*, 39, 444–455.
- Deo, D., & Sen, D. (2005). Automatic Recognition of Facial Features and Landmarking of Digital Human Head. In: *6th International Conference on Computer-Aided Industrial Design and Conceptual Design (CAID&CD)*, Delft, Netherlands.
- Dinh, H., & Kropac, S. (2006). Multi-resolution spin-images. *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 06)*. vol.1, pp. 863 - 870. DOI: 10.1109/CVPR.2006.197
- Dong, J., Krzyzak, A., & Suen, C. (2002). A practical SMO algorithm. *Proceedings International Conference Pattern Recognition*.
- Duda, R., & Hart, P. (1998). *Pattern Classification and Scene Analysis*. Recuperado de <http://www.gbv.de/dms/hebis-darmstadt/toc/32595522.pdf>
- Flynn, P.J., & Jain, A.K. (1988). Hypothesis testing and parameter estimation. *Computer Society Conference on Computer Vision and Pattern Recognition*, 261 – 267.
- Gordon, G. (1991). Face recognition based on depth maps and surface curvature. *Proceedings of SPIE Geometric methods in Computer Vision*, 234–247.
- Guangpeng, Z., & Yunhong, W. (2007). A 3D facial feature point localization method based on statistical shape model. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2007*, vol 2, 249 – 252.
- Hallinan, P., Gordon, G., Yuille, L., Giblin, P., & Mumford, D. (1999). Two- and three-dimensional patterns of the face. A.K. Peters, Ltd.
- Hearn, D. & Baker, P. (2003). *Computer Graphics with OpenGL*. Prentice Hall.
- Ho, H., & Gibbins, D. (2008). Multi-scale feature extraction for 3D surface registration using local shape variation. *Image and Vision Computing New Zealand, IVCNZ .23rd International Conference*. DOI: 10.1109/IVCNZ.2008.4762120
- Irfanoglu M., Gokberk B., & Akarun, L. (2004). 3D shape-based face recognition using automatically registered facial surfaces. *Pattern Recognition. Proceedings of 17th International Conference on Pattern Recognition (ICPR 2004)*, vol 4, 183–186. DOI: 10.1109/ICPR.2004.1333734

Jagannathan, A., & Miller, E. (2007). Three-dimensional surface mesh segmentation using curvedness-based region growing approach. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 29(12):2195–2204.

Johnson, A. (1997). *Spin-Images: A Representation for 3-D Surface Matching*. Carnegie Mellon University: PhD thesis. Recuperado de: <http://cite.seerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.4190&rep=rep1&type=pdf>

Koenderink, J., & Van Doorn, A. (1992). Surface shape and curvature scales. *Image and Vision Computing*, 10, 557–56.

Leal, E., Branch, J., & Ortega, O. (2007). Estimación de curvaturas y direcciones principales en nube de puntos no organizados. *Revista Dyna*, 153, 351–362.

Lee, J.C., & Milios, E. (1990). Matching range images of human faces. *Computer Vision*, 1990. Proceedings, Third International Conference on computer vision 722–726. DOI: 10.1109/ICCV.1990.139627

Lee, J. (2005). 3D face recognition using range images. Technical report, Department of Electrical and Computer Engineering, The University of Texas, Austin, TX. Recuperado de: <http://users.ece.utexas.edu/~bevans/courses/ee381k/projects/spiring05/lee/LitSurveyReport.pdf>

Lu, X., Colbry, D., & Jain, A. (2004). Three-dimensional model based face recognition. *Proceedings of 17th International Conference on Pattern Recognition*, vol. 1, (pp 362–366).

Lu, X., Colbry, D., & Jain, A. (2006). Matching 2.5D scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1), 31–43.

MeshLab. <http://meshlab.sourceforge.net>.

Moreno-Díaz. A.B. (2004). *Reconocimiento Facial Automático mediante Técnicas de Visión Tridimensional*. Universidad Politécnica de Madrid: Tesis Doctoral. Recuperado de: <http://oa.upm.es/625/1/10200408.pdf>

Salazar, A.E., & Prieto, F.A. (2010). 3D BSM for face segmentation and landmarks detection. *Three-Dimensional Image Processing (3DIP) and Applications*.

Sun, Y., & Yin, L. (2008). Automatic pose estimation of 3d facial models. En: *Proceedings of 19th International Conference on Pattern Recognition. ICPR2008*, 1–4.

Tanaka, H., Ikeda, M., & Chiaki, H. (1998). Curvature-based face surface recognition using spherical correlation. Principal directions for curved object recognition. *Automatic Face and Gesture Recognition, Proceeding of Third IEEE International Conference on Automatic Face and Gesture Recognition*. Proceedings, 372–377. DOI: 10.1109/AFGR.1998.670977

Taubin, G. (1995). Estimating the tensor of curvature of a surface from a polyhedral approximation. *Fifth International Conference on Computer Vision (ICCV)*.

Xue, F., & Ding, X. (2006). 3D+2D face localization using boosting in multi-modal feature space. *Proceedings of 18th International Conference on Pattern Recognition (ICPR'06)*.

Yin, L., Wei, X., Sun, Y., Wang, J., & Rosato, M. (2006). A 3D facialexpression database for facial behavior research. *Automatic Face and Gesture Recognition, FGR 2006*, In 7th International Conference on Automatic Face and Gesture Recognition (FGR06). pp. 211–216. DOI: 10.1109/FGR.2006.6