

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO  
FACULTAD DE INGENIERÍA



**DETECCIÓN Y LOCALIZACIÓN AUTOMÁTICA  
DE LA SUPERFICIE FACIAL EN IMÁGENES 3D  
DE UNA PROFUNDIDAD**

**TESIS**

**QUE PARA OBTENER EL GRADO DE MAESTRO EN  
CIENCIAS DE LA INGENIERÍA**

PRESENTA:

**ING. JUAN PADUANO SALINAS**

DIRECTOR DE TESIS:

**DR. MARCELO ROMERO HUERTAS**

CIUDAD UNIVERSITARIA TOLUCA, MÉXICO

OCTUBRE 2016



A mi familia ...



# Resumen

La detección del rostro es uno de los primeros pasos en las aplicaciones de procesamiento facial, cuyo propósito es identificar y localizar todos los rostros que se encuentran en una imagen, independientemente de su posición, escala y orientaciones. En esta tesis se reporta la investigación realizada en la Maestría en Ciencias de la Ingeniería de la Universidad Autónoma del Estado de México para detectar y localizar automáticamente la superficie facial utilizando imágenes 3D de una profundidad con variaciones en número y posición de las personas en la escena. Para esto se revisó la literatura relacionada y se discute la problemática actual en la detección de la superficie facial en imágenes 3D. Se propone y evalúa experimentalmente una técnica para la detección de la superficie facial automáticamente con variaciones en el número de usuarios y su posición con respecto a la cámara. Esta técnica consta de cuatro etapas: a) Eliminación de partes planas en la imagen 3D utilizando el algoritmo MSAC, b) Análisis de curvatura para detectar puntos convexos, c) Selección automática de la punta de la nariz utilizando SpinImages y d) Clasificación rostro o no rostro. En esta investigación se utilizó el sensor Kinect One™, para coleccionar una base de datos de 1020 imágenes 2D/3D variando el número y posición en la escena de treinta y dos sujetos. Las imágenes 3D coleccionadas fueron utilizadas para la evaluación de nuestro proceso de detección de la superficie facial. Este proceso ha sido capaz de detectar con éxito el 91 % de los rostros en las imágenes coleccionadas que contienen más de un sujeto en la escena, y el 100 % y 98 % de los rostros contenidos en las bases de datos Face Recognition Grand Challenge y CurtinFaces, respectivamente. En ese sentido, se obtuvieron resultados que son alentadores para abordar el problema general en la detección de la superficie facial en condiciones no controladas, con más de una persona en la escena y variación en su posición con el objetivo de crear sistemas de procesamiento facial automáticos.

# Publicaciones

Partes de la presente investigación se han publicado en:

- **Sección 4, Capítulo 3:** J. Paduano, M. Romero, and R. M. Valdovinos, “Face detection in 3D images with more than one person” in Proceedings on the International Conference on Image Processing, Computer Vision and Pattern Recognition (ICCV’16), Jul 2016.
- **Sección 3, Capítulo 3:** J. Paduano, M. Romero, and V. Munoz, “Toward face detection in 3D data” in Proceedings on the International Conference on Image Processing, Computer Vision and Pattern Recognition (ICCV’15), Jul 2015.
- M. Romero, J. Paduano, and V. Munoz, “Point-triplet spin-images for landmark localisation in 3d face data,” in Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, 2014 IEEE Workshop on, pp. 8–14, Oct 2014.

# Índice general

|   |           |
|---|-----------|
| <b>1. Introducción</b>                      | <b>1</b>  |
| 1.1. Directriz de la tesis . . . . .        | 2         |
| 1.2. Estructura de la tesis . . . . .       | 2         |
| <b>2. Protocolo de tesis</b>                | <b>4</b>  |
| 2.1. Estado del arte . . . . .              | 4         |
| 2.2. Planteamiento del problema . . . . .   | 6         |
| 2.3. Justificación . . . . .                | 8         |
| 2.4. Objetivos . . . . .                    | 9         |
| 2.5. Hipótesis . . . . .                    | 9         |
| 2.6. Alcances y limitaciones . . . . .      | 9         |
| 2.7. Metodología . . . . .                  | 10        |
| 2.8. Infraestructura requerida . . . . .    | 11        |
| 2.9. Contribuciones esperadas . . . . .     | 12        |
| <b>2.10. Referencias</b>                    | <b>13</b> |
| <b>3. Human Face Detection in 3D images</b> | <b>15</b> |
| <b>4. Discusión</b>                         | <b>47</b> |

# Capítulo 1

## Introducción

La detección del rostro se ha considerado fundamental para la interacción persona computadora además de ser una capacidad natural del ser humano. La detección facial es una etapa primaria de las aplicaciones de procesamiento facial (Bowyer et al., 2006), tales como: reconocimiento facial, seguimiento del rostro, estimación de la posición, reconocimiento de las expresiones faciales, reconocimiento de gestos, modelado facial, diseño de interfaces hombre-máquina, la comunicación multimedia y el diagnóstico médico (Yang et al., 2002).

Sin embargo, la mayoría de los métodos existentes en estas aplicaciones utilizan imágenes o una secuencia de imágenes que contienen básicamente un rostro.

Para poder construir un sistema completamente automático primero se requiere detectar el rostro y de esa manera poder extraerlo y continuar con un análisis (Hjelmås and Low, 2001). Para lograr esto se requiere desarrollar algoritmos robustos y eficientes para detectar el rostro.

Dada una imagen o una secuencia de imágenes, el objetivo de la detección de rostros es identificar y localizar todos los rostros humanos, independientemente de sus posiciones, escalas, orientaciones, poses y las condiciones del ambiente (Yang et al., 2002).

La detección facial es un problema difícil porque los rostros humanos son altamente no rígidos con un alto grado de variabilidad en tamaño, forma, color y textura (Zafeiriou et al., 2015).

Este trabajo de investigación propone detectar y localizar la superficie facial utilizando imágenes 3D. El interés de trabajar con imágenes 3D es que proporcionan información de profundidad explícita que se considera útil para estudiar la forma de los objetos digitalizados. Además, en comparación con las imágenes 2D, las imágenes 3D son más robustas a cambios de iluminación del ambiente, posición y orientación del rostro de la persona en relación a los sensores 3D; proporcionando información más detallada sobre la forma y rasgos faciales finos.



La motivación de hacer este trabajo de investigación es apoyar la construcción de sistemas de procesamiento facial que trabajen con imágenes del rostro de manera automatizada, en ambientes no controlados, para lo cual se requiere desarrollar algoritmos robustos y eficientes para detectar la superficie facial.

Para cumplir el objetivo de esta investigación, se identificaron tres técnicas del estado del arte de la detección del rostro en imágenes 3D:(Colombo et al., 2006; Mian et al., 2006; Segundo et al., 2007) las cuales dependen de condiciones controladas y solo detectan un rostro por imagen. En este trabajo se realizó el análisis he implementación de estas técnicas, donde se identificaron debilidades y fortalezas que permitieron proponer un proceso para la detección y localización del rostro en imágenes 3D con variaciones en el número y posición de los sujetos en la escena.

Por lo tanto, la investigación aquí reportada propone un proceso para la detección de la superficie facial basada en la detección de la punta de la nariz y haciendo uso de los clasificadores PCA y SVM en imágenes 3D de una profundidad adquiridas con el sensor Kinect One™.

Acorde al Art. 60 del Reglamento de Estudios Avanzados de la Universidad Autónoma del Estado de México la estructura de esta tesis es por artículo especializado.

Por lo tanto, los resultados obtenidos en esta investigación han sido enviados para revisión y posible publicación en la revista Computer Vision and Image Understanding, la cual tiene un factor de impacto de 2.134 y esta indexada en Scopus, Computing Reviews, Current Contents/Engineering, Computing & Technology, Index to Scientific Reviews, Science Abstracts - Physics Abstracts y Science Citation Index.

## **1.1. Directriz de la tesis**

Se espera que ésta tesis se útil para todo aquel interesado en el procesamiento de imágenes faciales 3D utilizando imágenes de profundidad obtenidas con el sensor Kinect One™, específicamente se presenta una alternativa para la detección del rostro que pudiera ayudar al desarrollo de aplicaciones finales.

Este trabajo de investigación les puede servir a personas interesadas en aplicaciones procesamiento facial ya que la detección del rostro es una tarea primaria para estas aplicaciones.

## **1.2. Estructura de la tesis**

Adicionalmente a este capítulo de introducción, esta tesis contiene el siguiente capitulado:

En el **Capítulo II**: “*Protocolo de tesis*”, se muestra el protocolo de investigación desarrollado para esta tesis.

En el **Capítulo III**: “*Human Face Detection in 3D Images*”, se presenta el artículo de detección facial enviado para su revisión a la revista Computer Vision and Image Understanding.

En el **Capítulo IV**: “*Discusión*”, se puntualizan los hallazgos relevantes obtenidos en esta investigación.

# Capítulo 2

## Protocolo de tesis

### 2.1. Estado del arte

La detección facial es una tarea primaria en aplicaciones de procesamiento facial (Yang et al., 2002), puesto que es esencial detectar el rostro dentro de una imagen para proceder a su análisis.

Al paso de las décadas cuando se habla de detección facial, los investigadores entran en controversia respecto ha trabajar con imágenes de profundidad o con imágenes de intensidad. Como se sabe, las imágenes de intensidad presentan factores que retrasan o impiden el análisis de la superficie facial (Bowyer et al., 2006) . Es por eso el interés del estudio de la superficie facial en imágenes 3D. La adquisición de la imagen facial 3D es invariante a cambios en la iluminación durante la captura, además se considera que las imágenes 3D representan explícitamente la forma y dimensión del rostro humano. Trabajos como los de Bowyer et al. (2006) y Xu et al. (2004) han comprobado que si se quiere analizar la forma de los objetos, la manera natural es utilizando una imagen 3D.

El incremento del poder computacional así como la aparición de los sensores 3D de bajo costo, por ejemplo, el sensor Kinect™ (Microsoft, 2014) , han permitido incrementar el estudio en el procesamiento de imágenes digitales en particular en aplicaciones de procesamiento facial, tales como; reconocimiento facial, seguimiento del rostro, estimación de la posición, reconocimiento de las expresiones faciales y reconocimiento de gestos (Bowyer et al., 2006; Lei et al., 2014; Ortega-Garcia et al., 2004; Pears et al., 2012; Smeets et al., 2010; Xu et al., 2004).

La detección de la superficie facial es clasificada de acuerdo al tipo de imagen que se utilice (Bowyer et al., 2006):

- La detección facial 2D: es aquella que utiliza imágenes de intensidad/color para detectar el rostro.
- La detección facial 3D: es aquella que utiliza imágenes de profundidad para detectar el rostro.
- La detección facial multimodal: es aquella que utiliza tanto las imágenes 2D/3D.

En particular, este trabajo de investigación se centrará en la detección de la superficie facial en imágenes 3D.

En la literatura se encuentran estudios relacionados con aplicaciones de procesamiento facial 3D, sin embargo, muy pocos de ellos se dedican específicamente a la detección facial; Por ejemplo, [Colombo et al. \(2006\)](#) implementó un algoritmo de detección facial que fue experimentado en imágenes de posición frontal, el cual es capaz de determinar la presencia de un rostro y en caso afirmativo extraer la superficie facial. Primero las áreas que no pertenecen al rostro son descartadas, a continuación analizan las áreas cóncavas y convexas del rostro utilizando el análisis de curvatura con el fin de encontrar las esquinas de los ojos y la punta de la nariz con los cuales calculan  $n$  combinaciones de puntas de la nariz con esquinas de los ojos. Si al final de estos pasos no encuentra al menos un candidato como nariz y dos candidatos como ojos se asume que no existe un rostro en la imagen. El reporte de resultados no precisa el número de combinaciones obtenidas, sin embargo, es evidente que se obtiene una cantidad de combinaciones considerables.

Después de obtener las  $n$  combinaciones de esquinas interiores de los ojos y punta de la nariz se alinean las imágenes en una posición estándar utilizando transformaciones rígidas. Al final crean un nueva imagen 3D con la cual analizan la parte rígida del rostro utilizando eigenfaces para clasificar la imagen en rostros o no rostros.

Por su parte, [Mian et al. \(2006\)](#) presenta un sistema automático de reconocimiento facial 3D, en cuyas primeras etapas detecta y alinea la superficie facial, correspondiendo las imágenes 2D y 3D obtenidas por el sensor y utilizando la transformada de Hotelling y su algoritmo clasificador Spherical Face Representation (SFR). La experimentación aquí reportada utiliza la base de datos del Face Recognition Grand Challenge (FRGC), la cual contiene 9500 imágenes 2D/3D de 370 personas tomadas desde la parte superior del tórax a la cabeza. En este trabajo, para extraer la superficie facial de la imagen 3D se detecta la punta de la nariz, cortando la imagen horizontalmente y analizando el contenido del corte punto por punto para identificar la punta de la nariz. Los puntos que no correspondan a la región de la nariz son considerados atípicos y por lo tanto se remueven. Detectada la punta de la nariz, se utiliza un radio de ocho centímetros a partir de este punto para extraer el rostro de la imagen 3D.

En un trabajo similar, [Segundo et al. \(2007\)](#), propone una metodología para la detección de la superficie facial a través de segmentación de un mapa de profundidad obtenido de la imagen 3D. Esta técnica fue diseñada para imágenes que contienen una persona. Para esto, primero discriminan los puntos que no pertenecen a la superficie facial, utilizando el algoritmo K-Means ( $k = 3$ ) para obtener tres clusters de la imagen: fondo, cuerpo y cabeza. Sin embargo, se reporta que este paso no es suficiente porque existen otras regiones en el área de la cabeza que afectan el resultado, por ejemplo la presencia de cabello, oídos y cuello. Posteriormente, se detecta el contorno de la superficie facial utilizando la transformada de Hough, la cual se asume de forma elipsoidal. Al igual que [Mian et al. \(2006\)](#), en esta experimentación se utiliza la base de datos FRGC.

Adicionalmente se pueden identificar otras investigaciones que de forma indirecta realizan la detección de la superficie facial en imágenes 3D. Por ejemplo, [Creusot et al. \(2010\)](#) desarrolla un sistema para etiquetar los puntos antropométricos (Landmarks) de la superficie facial utilizando imágenes que solo contienen el rostro de una persona. En este sentido, [Xu et al. \(2004\)](#) propone un método para localizar la punta de la nariz utilizando un esquema de filtrado jerárquico que combina características locales. Posteriormente, en base a la punta de la nariz detectada, se estima la cresta utilizando un análisis de curva denominado Included Angle Curve (IAC). Un trabajo similar es el propuesto por [Dibeklioglu et al. \(2008\)](#), quien sugiere un método basado en análisis de curvatura para la localización de la punta de la nariz en imágenes 2D, que posteriormente corresponde a la imagen 3D.

Como se puede observar en la literatura, la detección de la superficie facial es una tarea primaria en aplicaciones de procesamiento facial, en la cual se requiere un estudio experimental detallado para la detección de la superficie facial en imágenes 3D para ser utilizado en aplicaciones de procesamiento facial 3D.

## 2.2. Planteamiento del problema

La detección de la superficie facial es un problema de alta complejidad que debido a que influyen diversos factores, por ejemplo: Los rasgos faciales que varían de un individuo a otro (e incluso para una sola persona), dichas variaciones son debido a la posición, tamaño, ángulo de visión, entre otros. La investigación de visión por computadora ha recorrido un largo camino para hacer frente a estas dificultades, pero sigue habiendo oportunidades de mejora ([Bowyer et al., 2006](#)).

La detección de la superficie facial 3D, consiste en identificar y localizar cada uno de los rostros humanos en una imagen 3D, independientemente de su posición, escala y orientación. En su forma más básica, una imagen 3D representa la superficie de los objetos mediante una nube de puntos 3D, generalmente con coordenadas cartesianas  $(x, y, z)$ , los cuales pueden ser procesados de acuerdo a su información contextual, proximidad y alineación.

Existen técnicas que permiten la detección de la superficie facial 3D, por ejemplo: [Colombo et al. \(2006\)](#), [Mian et al. \(2006\)](#), [Nair and Cavallaro \(2009\)](#) y [Segundo et al. \(2007\)](#). Lamentablemente, estos trabajos dependen de la posición de la persona y la detección se realiza con imágenes 3D que contienen básicamente un rostro humano. Además, estas imágenes fueron tomadas en condiciones controladas de posición, ubicación y expresiones faciales.

La dificultad que representa la detección de la superficie facial depende de factores tales como: La posición del sujeto, tipo de sensor 3D utilizado para la captura de imagen, número de vértices que inciden en la superficie facial, el tipo de imagen 3D, la presencia de expresiones faciales y el tiempo de procesamiento:

- La posición del sujeto juega un papel importante en la detección facial, ya que puede causar oclusión en la imagen 3D y esto complica la detección del rostro.
- El tipo de imagen 3D puede ser no estructurada (la cual contiene solamente los vértices en el espacio tridimensional), y estructurada (que además de contener los vértices proporciona la conexión de los mismos). Si las imágenes 3D son no estructuradas dificulta el análisis de la superficie facial porque no existe la conexión entre los vértices.
- El sensor 3D influye en el tipo de dato que se este utilizando. Como se mencionó en la Sección 2.1 existen dos tipos de imágenes 3D estructuradas y no estructuradas, algunos trabajos necesitan de una imagen estructurada para poder ser utilizados, lamentablemente no todos los sensores 3D proporcionan este tipo de imagen.
- El número de vértices que inciden en la superficie facial dificultan la detección del rostro, puesto que a menor número de vértices se complica el análisis de la superficie. Por el contrario, a mayor número de vértices que inciden en la superficie facial se facilita su análisis. Es importante aclarar que se esta hablando de los vértices que pertenecen a la superficie facial y no de todos los vértices contenidos en la escena.
- Las expresiones faciales son consideradas como problema fundamental para el análisis de la superficie facial, ya que cambia la superficie del rostro de una expresión a otra y además cada individuo es diferente.

- El tiempo de procesamiento es importante. Por ejemplo, si los resultados de la detección facial 3D son del 99% en un tiempo mayor de 10 minutos, no serviría en aplicaciones de seguridad o seguimiento del rostro. En este sentido, es necesario obtener mediciones cuantitativas del tiempo, el cual es un tema que no se ha analizado en la literatura.
- Es importante la precisión con la que se puede detectar la superficie facial en imágenes 3D que contienen más de un sujeto en escenarios reales.

La mayoría de los algoritmos existentes de detección de la superficie facial han sido evaluados con bases de datos 3D que fueron colectadas con sensores 3D especializados, por ejemplo la base de datos FRGC. Sin embargo, a partir del año 2010 Microsoft inicio la comercialización del sensor Kinect™ (Microsoft, 2014), que es un sensor de profundidad de bajo costo que desde su lanzamiento ha sido investigado por la comunidad científica. Lamentablemente, estos estudios se han realizado con imágenes que contienen un solo sujeto en la escena. Esta será una característica de interés para la investigación propuesta, puesto que se estudiará el problema de detección y localización de la superficie facial en imágenes de una profundidad colectadas con el sensor Kinect™, que contengan variaciones en el número y posición de sujetos en la escena.

### 2.3. Justificación

El interés de hacer este trabajo de investigación es ayudar a la construcción de sistemas de procesamiento facial que trabajen con imágenes 3D del rostro humano de manera automatizada, en diferentes variaciones en número y posición de sujetos en la escena, entonces se requiere desarrollar algoritmos robustos y eficientes para detectar la superficie facial en una imagen 3D de una profundidad puesto que en la literatura se observa como un posible punto de mejora.

Como parte de este trabajo, se evaluarán experimentalmente los cambios en el rendimiento de los algoritmos de detección facial ante variaciones como posición y número de personas en la escena. Para este propósito, se utilizará el sensor Kinect One™.

Como se puede observar en la literatura relacionada, se han publicado técnicas para detectar la superficie facial, sin embargo, éstas funcionan en condiciones controladas, por ejemplo: posiciones frontales, un solo usuario en la escena, a nivel de tórax y hasta la cabeza del usuario.

Con las nuevas tecnologías en específico el sensor Kinect™, se podrán estudiar imágenes con los siguientes factores: más de una persona por escena, variaciones de posición y escenarios no controlados. En estas condiciones los algoritmos propuestos para la detección facial que se encuentran en la literatura no funcionarían por las condiciones en las que fueron diseñadas.

Es por eso que se requiere de un estudio de la detección de la superficie facial para poder ser implementado en aplicaciones de procesamiento facial.

## 2.4. Objetivos

### Objetivo general

Detectar y localizar la superficie facial automáticamente en imágenes 3D de una profundidad con variaciones en el número de usuarios y su posición con respecto a la cámara para aplicaciones de procesamiento facial.

### Objetivos particulares

- Analizar experimentalmente las técnicas de detección facial en imágenes 3D propuestas en el estado del arte para proponer nuevos algoritmos para la detección y localización de la superficie facial.
- Colectar imágenes experimentales con variaciones en el número y posición del sujeto en la escena utilizando el sensor Kinect One™ para evaluar los algoritmos de detección y localización de la superficie facial.

## 2.5. Hipótesis

Las imágenes 3D de una profundidad adquiridas con el sensor Kinect™, permiten detectar y localizar la superficie facial de los individuos presentes en una escena con variaciones en el número y posición de sujetos.

## 2.6. Alcances y limitaciones

### Alcances

- Un proceso para detectar la superficie facial en imágenes 3D con variaciones en el número y posición de los sujetos al sensor Kinect One™, capaz de detectar el 91 % de los rostros presentes en la escena con un error menor a 16mm.



- Un conjunto de imágenes faciales experimentales con variaciones en el número y distancia de los sujetos al sensor, variando de uno a ocho sujetos en la escena de un total de treinta y dos personas.
- El proceso de detección de rostros propuesta en esta investigación solamente esta basada en la localización de la punta de la nariz.

## Limitaciones

- El sensor utilizado para la colecta de imágenes experimentales tiene limitantes técnicas como: su campo de visión y baja resolución espacial.
- Parámetros y umbrales pre-establecidos y calculados de forma experimental en el proceso de detección propuesto: detección de partes planas, cálculo curvatura Media y Gaussiana, cálculo de histograma de SpinImages.

## 2.7. Metodología

Para el desarrollo de este trabajo de investigación, se utiliza una metodología de seis etapas:

1. **Definición del problema:** En esta etapa se define el problema que se quiere resolver partiendo de lo general a lo particular.
2. **Revisión de la literatura:** Se revisa el estado del arte en la detección y localización de la superficie facial en imágenes 3D. Se evalúa experimentalmente las técnicas existentes en la literatura para detectar la superficie facial 3D.
3. **Análisis y diseño operacional:** Analizar y diseñar procedimientos, técnicas y algoritmos para realizar de detección y localización de la superficie facial en imágenes 3D. Determinar algoritmos para la detección y localización de la superficie facial en imágenes 3D considerando variaciones en el número y posición de los sujetos en una escena para aplicaciones de procesamiento facial. Diseño de función de captura de un conjunto de imágenes experimentales considerando variaciones en el número y posición de los sujetos en la escena. Entendiendo por variaciones de posición, la ubicación del sujeto con respecto al eje  $z$  del sistema de referencia del sensor de captura.

4. **Adquisición de datos experimentales:** Adquirir un conjunto de imágenes experimentales 2D/3D utilizando un módulo de captura del sensor Kinect One™.
5. **Experimentación:** Probar los procedimientos, técnicas y algoritmos de detección facial propuestos en diferentes condiciones (posición de rostro, número de sujetos en la escena) utilizando bases de datos del estado del arte (FRGC y CurtinFaces) y el conjunto de imágenes experimentales colectadas en esta investigación.
6. **Evaluación y análisis de resultados:** Evaluar y analizar la robustez de métodos, técnicas o algoritmos propuestos, además definir los puntos de mejora.

## 2.8. Infraestructura requerida

Para poder llevar a cabo este trabajo de investigación se requiere de la infraestructura siguiente:

- Bases de datos de rostros faciales 3D, estas bases permitirán implementar las técnicas existentes de detección de la superficie facial.
- Equipo de cómputo con las siguientes características mínimas: Procesador (Intel Core i7 quad core de 2.8 GHz, Turbo Boost de hasta 4.0 GHz), Memoria (16 GB SDRAM DDR3L de 1600 MHz), Espacio de almacenamiento (flash basado en PCIe de 512 GB). En este equipo de cómputo se programarán las técnicas de detección facial.
- Dispositivos de almacenamiento de 3TB de espacio para poder guardar los archivos resultantes de las pruebas, además de servir como medio de respaldo para las bases de datos.
- Un sensor Kinect One™ para implementar el escenario de experimentación y capturar las imágenes experimentales.
- Servidor con las siguientes características: Procesador (12 núcleos de 2.7 GHz con 30 MB de caché L3), Memoria (64 GB de memoria RAM de al menos 1866 MHz), Almacenamiento (flash de 1 TB basado en PCIe), Tarjetas gráficas (Dos GPUs AMD FirePro D700 con 6 GB de GDDR5 VRAM cada uno). En este servidor se realizarán evaluaciones masivas de los algoritmos propuestos para la detección de la superficie facial.
- Acceso a bibliotecas digitales para poder estudiar las tendencias de esta investigación.

## 2.9. Contribuciones esperadas

Las contribuciones originales que se espera derivar de este trabajo de investigación son las siguientes:

- Algoritmo para la detección y localización de la superficie facial en imágenes 3D de una profundidad con variaciones en el número y posición de los sujetos en la escena.
- Conjunto de imágenes 3D capturadas con sensores de una profundidad en escenas no controladas.

## 2.10. Referencias

- Bowyer, K. W., Chang, K., Flynn, P., 1 2006. A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding* 101 (1), pp. 1–15.
- Colombo, A., Cusano, C., Schettini, R., Mar 2006. 3d face detection using curvature analysis. *Pattern Recognition* 39 (3), pp. 444 – 455.
- Creusot, C., Pears, N., Austin, J., October 2010. 3d face landmark labelling. In: *Proceedings of the ACM workshop on 3D object retrieval*. ACM, Association for Computing Machinery, pp. 27–32.
- Dibeklioglu, H., Salah, A. A., Akarun, L., 2008. 3d facial landmarking under expression, pose, and occlusion variations. In: *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*. pp. 1–6.
- Hjelmås, E., Low, B. K., 2001. Face detection: A survey. *Computer Vision and Image Understanding* 83 (3), pp. 236 – 274.
- Lei, Y., Bennamoun, M., Hayat, M., Guo, Y., 2 2014. An efficient 3d face recognition approach using local geometrical signatures. *Pattern Recognition* 47 (2), pp. 509–524.
- Mian, A., Bennamoun, M., Owens, R., 2006. Automatic 3d face detection, normalization and recognition. In: *3D Data Processing, Visualization, and Transmission, Third International Symposium on*. pp. 735–742.
- Microsoft, C., Microsoft Kinect for Windows, 2014. Nov. 2014.  
URL <http://www.microsoft.com/en-us/kinectforwindows/>
- Nair, P., Cavallaro, A., 2009. 3-d face detection, landmark localization, and registration using a point distribution model. *Multimedia, IEEE Transactions on* 11 (4), pp. 611–623.

- Ortega-Garcia, J., Bigun, J., Reynolds, D., Gonzalez-Rodriguez, J., 2004. Authentication gets personal with biometrics. *Signal Processing Magazine, IEEE* 21 (2), pp. 50–62.
- Pears, N., Liu, Y., Bunting, P., 2012. *3D Imaging, Analysis and Applications*, 1st Edition. Vol. 1. Springer-Verlag London.
- Segundo, M. P., Queirolo, C., Bellon, O. R. P., Silva, L., 2007. Automatic 3d facial segmentation and landmark detection. In: *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. pp. 431–436.
- Smeets, D., Claes, P., Vandermeulen, D., Clement, J. G., 9 2010. Objective 3d face recognition: Evolution, approaches and challenges. *Forensic Science International* 201 (1–3), pp. 125–132.
- Xu, C., Wang, Y., Tan, T., L., Q., 2004. Depth vs. intensity: which is more important for face recognition? *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* 1, pp. 342–345 Vol.1.
- Yang, M., Kriegman, D., Ahuja, N., Jan 2002. Detecting faces in images: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24 (1), pp. 34–58.
- Zafeiriou, S., Zhang, C., Zhang, Z., 2015. A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding* 138, pp. 1 – 24.

## Capítulo 3

# Human Face Detection in 3D images

En este capítulo se presenta un artículo que reporta los resultados obtenidos en esta investigación, el cual ha sido enviado para su revisión y posible publicación en la Revista Científica *Computer Vision and Image Understanding*, la cual tiene un factor de impacto de 2.1 y se encuentra indexada en Scopus.

Manuscript Number:

Title: Human Face Detection in 3D Images

Article Type: Research paper

Keywords: 3D face detection; 3D face processing; Biometrics

Corresponding Author: Dr. Marcelo Romero, Ph.D.

Corresponding Author's Institution: Universidad Autonoma del Estado de Mexico

First Author: Juan Paduano

Order of Authors: Juan Paduano; Marcelo Romero, Ph.D.

Abstract: Face detection is one of the most studied topics in computer vision; however, current 3D face detection techniques depend on the position of the person and they are performed basically on images that contain mainly the human face. In particular, three key face detection techniques were experimentally investigated to begin this research. The novelty of this research is to detect and locate every face present into a 3D image regardless of the number of subjects and their distance to the camera. For this purpose, we are introducing an effective straight forward approach that consists of four steps. We populated a 3D image database using the Kinect One sensor, varying positions from one to eight different subjects in the scene. In addition to our 3D image database, this face detection approach has been evaluated with state of the art databases, Face Recognition Grand Challenge (FRGC) and CurtinFaces that contain one subject per image. With localisation errors from 0 to 16 mm, our experimental results are 100%, 98% and 91% for the FRGC, CurtinFaces and our database, respectively.

Suggested Reviewers: Felipe Orihuela PhD  
Full Time Researcher, Ciencias Computacionales, INAOE  
f.orihuela-espina@ccc.inaoep.mx  
Active researcher in the field

Fernando Arambula PhD  
Full Time Researcher, CCADET, Universidad Nacional Autonoma de Mexico  
fernando.arambula@ccadet.unam.mx  
Active researcher in the field

Juan Carlos Cuevas PhD  
Full Time Researcher, Facultad de Ingenieria, Universidad Autonoma de San Luis Potosi  
cuvastello@gmail.com  
Active researcher in the field

**\*1. Cover Letter**

TITLE OF THE SUBMITTED MANUSCRIPT:

**Human Face Detection in 3D Images**

LIST OF ALL AUTHORS' NAMES AND AFFILIATIONS:

**Juan Paduano and Marcelo Romero**  
**Facultad de Ingeniería**  
**Universidad Autónoma del Estado de México**

A SHORT STATEMENT (<50 words) OF THE PRECISE PROBLEM OR OBJECTIVE ADDRESSED IN THE PAPER:

**A simple four-steps process for face detection in 3D images with more than one person is presented.**

A VERY BRIEF (<100 words) DESCRIPTION OF THE ESSENCE OF YOUR APPROACH:

A LIST OF THE SPECIFIC MAJOR NOVEL CONTRIBUTIONS (up to 3) REPORTED HERE:

- 1. A novel face detection approach**
- 2. A state of the art dataset**
- 3. Experimental results**

LIST OF OTHER JOURNAL OR CONFERENCE PAPERS (if any) PUBLISHED OR SUBMITTED BY YOU OR ANY CO-AUTHOR THAT HAVE A SIGNIFICANT OVERLAP WITH THE CONTRIBUTION SUBMITTED HERE ACCOMPANIED BY A BRIEF EXPLANATION OF THE NATURE OF THIS OVERLAP POINTING OUT CLEARLY WHICH NOVEL IDEAS SUBMITTED HERE HAVE NOT BEEN DISCUSSED IN THESE PRIOR PUBLICATIONS:

- 1. IPCV 2015, WorldComp 2015, <http://worldcomp-proceedings.com/proc/proc2015/>**
- 2. IPCV 2016, WorldComp 2016, <http://worldcomp-proceedings.com/proc/p2016/>**

A REFERENCE TO THE CLOSEST PRIOR ARTICLE (by others) UPON WHICH YOUR CONTRIBUTIONS IMPROVE:

**None.**

THE NAMES, EMAILS, AND HOMEPAGE URLS OF FOUR EXPERTS COVERING THESE AREAS AND FIELDS (by expert, we refer to someone who has published several high quality papers in that technical area in the last five years and who is recognized internationally as an expert in one of the fields you listed above):

- 1. Kevin W. Bowyer, [kwb@cse.nd.edu](mailto:kwb@cse.nd.edu), <https://engineering.nd.edu/profiles/kbowyer>**
- 2. Ajmal Saeed Mian, [ajmal.mian@uwa.edu.au](mailto:ajmal.mian@uwa.edu.au), <http://www.web.uwa.edu.au/people/ajmal.mian>**
- 3. Nick Pears, [nick.pears@york.ac.uk](mailto:nick.pears@york.ac.uk), <https://www-users.cs.york.ac.uk/nep/>**
- 4. Ioannis A. Kakadiaris, [ioannisk@uh.edu](mailto:ioannisk@uh.edu), <http://www.ee.uh.edu/faculty/kakadiaris>**



THE NAMES OF THE TWO GMOD ASSOCIATE EDITORS WHO YOU BELIEVE ARE THE MOST QUALIFIED TO HANDLE YOUR PAPER (please go to [http://www.elsevier.com/wps/find/journaleditorialboard.cws\\_home/622839/editorialboard](http://www.elsevier.com/wps/find/journaleditorialboard.cws_home/622839/editorialboard) to see list of associate editors:

TECHNICAL AREAS AND FIELDS OF EXPERTISE NECESSARY TO FULLY UNDERSTAND YOUR CONTRIBUTION AND TO EVALUATE ITS POTENTIAL AND NOVELTY (see list of areas and topics posted at <http://ees.elsevier.com/gmod/> and feel free to add areas and fields as needed):

- 1. 3D image processing**

# Human Face Detection in 3D Images

Juan Paduano, Marcelo Romero\*

*Facultad de Ingeniería, Universidad Autónoma del Estado de México. Cerro de Coatepec  
s/n. Ciudad Universitaria, 50100, Toluca, Estado de México, México*

---

## Abstract

Face detection is one of the most studied topics in computer vision; however, current 3D face detection techniques depend on the position of the person and they are performed basically on images that contain mainly the human face. In particular, three key face detection techniques were experimentally investigated to begin this research. The novelty of this research is to detect and locate every face present into a 3D image regardless of the number of subjects and their distance to the camera. For this purpose, we are introducing an effective straight forward approach that consists of four steps. We populated a 3D image database using the Kinect One™ sensor, varying positions from one to eight different subjects in the scene. In addition to our 3D image database, this face detection approach has been evaluated with state of the art databases, Face Recognition Grand Challenge (FRGC) and CurtinFaces that contain one subject per image. With localisation errors from 0 to 16 mm, our experimental results are 100%, 98% and 91% for the FRGC, CurtinFaces and our database, respectively.

*Keywords:* 3D face detection, 3D face processing, Biometrics

---

## 1. Introduction

Face detection is the first step in almost every face processing application, where the face is localised and extracted from an incoming image prior to specific analysis (Hjelmås and Low, 2001; Bowyer et al., 2006; Zafeiriou et al., 2015). However, face detection from a single image is a challenging

---

\*Corresponding author: Tel. +52 722 214 08 55  
Email address: mromeroh@uaemex.mx (Marcelo Romero)

6 task because of variability in scale, location, orientation (up-right, rotated),  
7 and pose (frontal, profile); also facial expressions, occlusion, and lighting  
8 conditions change the overall face appearance (Yang et al., 2002; Yang and  
9 Ahuja, 2001; Bowyer et al., 2006; Pears et al., 2012; Paduano et al., 2015).  
10 Evidently, face detection is the first step in any automated system that solves  
11 problems about face recognition areas. However, most of the existing face  
12 processing methods, assume that only one human face is present in an image  
13 or into a video (Yang and Ahuja, 2001; Bowyer et al., 2006; Zafeiriou et al.,  
14 2015; Ariz et al., 2016).

15 Although 2D detection algorithms have reached an acceptable level in face  
16 detection, they have a lot of difficulties because the 3D world is projected  
17 down onto a 2D image, resulting on ambiguities and losing depth informa-  
18 tion (Pears et al., 2012). On the other hand, 3D images provide explicit  
19 shape information and they are considered robust to illumination and pose  
20 variations (Bowyer et al., 2006).

21 Pan et al. (2013) used heterogeneous feature descriptors and feature se-  
22 lection in combination with Adaboost (Yin et al., 2015; Hu et al., 2008),  
23 although this is a robust and efficient face detection method, it depends on  
24 frontal position. Other pose-dependent face detection investigations using 2D  
25 image are: Subburaman and Marcel (2013); Aghaei et al. (2016); Bhandarkar  
26 and Luo (2009); Wang and Ji (2007).

27 Face detection is necessary for face processing application, for example:  
28 face recognition (Wagner et al., 2012; Kumar et al., 2015; Kakadiaris et al.,  
29 2016), face tracking (Smeulders et al., 2014; Aghaei et al., 2016; Bhan-  
30 darkar and Luo, 2009), pose estimation (Wohlhart and Lepetit, 2015), facial  
31 expressions and facial gestures recognition (Valstar et al., 2012). Unfortu-  
32 nately, different factors like position of the subject, 3D sensor, structural  
33 components, occlusion, facial surface resolution and facial expression, make  
34 face detection a complicated problem (Bowyer et al., 2006; Pears et al., 2012;  
35 Yang et al., 2014; Kumar et al., 2015; Chellappa et al., 1995).

36 Although face detection is essential for many face processing applications,  
37 it has received little attention, especially when using 3D data. In this paper,  
38 we propose an innovative approach for 3D face detection, that deals with  
39 face detection using one view 3D images of scenes with variations in number  
40 of persons and the distance to the sensor.

41 The rest of this paper is as follows. Section 2 describes previous work on  
42 3D face detection. Section 3 reports our experimental research in 3D images  
43 with one person. Section 4 introduces our face detection approach in 3D

44 images with more than one person. Finally, Section 5 concludes this paper.

## 45 2. Related work

46 There is a number of papers in face processing applications, however,  
47 only few of them are specifically dedicated to face detection. This section  
48 discusses relevant related work on face detection in 3D images.

49 Mian et al. (2006) proposed a nose tip detection method by using hori-  
50 zontally sliced contours along a face image. Each sliced contour is formed by  
51  $n$  vertex. For each vertex in each sliced contour, circles of constant radius  
52 were estimated with centre in each vertex. Next, circumscribed triangles were  
53 formed in all the circles by using three vertices, one vertex is the centre of the  
54 circle and two intersection vertices located by coinciding the circle contour  
55 and points in the slice. After triangle altitude comparison, only one triangle  
56 is selected in each slice as the nose tip candidate, the one with maximum alti-  
57 tude from the centre of the circle. Random Sample Consensus (RANSAC)  
58 (Trucco and Verri, 1998) was used next to further outliers removal and from  
59 all the remaining nose tip candidates, the one with the maximum altitude is  
60 taken as the nose tip. Finally, the face is extracted using a pre-established  
61 radius. They reported 98.3% accuracy in face detection using the FRGC  
62 database.

63 Colombo et al. (2006) proposed a technique based on curvature analysis  
64 for face detection, they started the process using a range image containing  
65 a single face that is a representation of the  $(x, y, z)$  coordinates of the 3D  
66 image in a location  $(i, j)$ . Every 3D point's location was expressed regarding  
67 to the 3D sensor reference system. Next, from the range images, the Mean  
68 and Gaussian curvature were computed by the first and second derivatives  
69 in  $x$  and  $y$ . They considered a bi-quadratic polynomial approximation of the  
70 surface and the coefficients were obtained using least squares by fitting the  
71 points in a neighborhood of  $x$  and  $y$ . The objective for calculating Mean  
72 and Gaussian curvatures, is to use *HK classification* to obtain four segment  
73 types: convex, concave, and two saddle regions. The output may contain  
74 many candidates for facial features. If no nose or less than two candidate  
75 eyes are detected, they assume that no face is present in the scene. Finally, if  
76 a face is detected, the output is a list containing its location and facial area.  
77 They experimented with their own collected data using a Minolta<sup>TM</sup> camera  
78 and they achieved 96.85% successful face detection.

79 [Segundo et al. \(2007\)](#) presented a method to extract the face region of  
80 a single person. They also used range images generated from a point cloud.  
81 Their face segmentation algorithm consists of two main stages: First, the  
82 homogeneous regions localization in the input image using clustering and  
83 edge detection is used. Second, the candidates are identified when they  
84 belong to the face region, then this region is delimited by an ellipse using  
85 the *Hough Transform*. They applied the K-Means algorithm with  $k = 3$   
86 for image segmentation. So, the input image was segmented in three main  
87 regions identified as: background, body, and face. However, this step by  
88 itself is not enough for face segmentation, then it is necessary to extract the  
89 face region by edge detection. After performing region and edge detection,  
90 which can be made in parallel, they combine the two resulting images and  
91 finally, by an *and* operation, the facial region was obtained. They reported  
92 a 99.95% face detection using the FRGC database.

93 [Nair and Cavallaro \(2009\)](#) present an accurate and robust framework  
94 for face detection and segmentation using landmark localization and doing  
95 fine registration of face meshes by fitting a facial model (polygons and ver-  
96 tices). This model is based on a 3D Point Distribution Model (PDM) that  
97 is fitted without relying on texture, pose, or orientation information. Fit-  
98 ting is initialized using candidate locations on the mesh, which are extracted  
99 from low-level curvature-based feature maps. Face detection is performed by  
100 classifying the transformations between model points and candidate vertices  
101 based on the upper-bound of the deviation of the parameters from the mean  
102 model. The performance of face detection is evaluated on a database of face  
103 and no-face images achieving an accuracy of 99.6%. They also demonstrated  
104 face detection and segmentation with different scale and pose.

105 From literature reviewed above, we have experimentally investigated three  
106 key face detection techniques ([Paduano et al., 2015](#)) using state of the art  
107 databases: Face Recognition Grand Challenge and CurtinFaces for a deep  
108 understanding. Findings and knowledge from this experimental research al-  
109 low us to propose our face detection approach ([Paduano et al., 2016](#)).

110 Additionally, one may identify a number of related face processing re-  
111 search using 2D and 3D images. For example, [Peng et al. \(2011\)](#) propose a  
112 nose tip detection method that has three main characteristics. First, it does  
113 not require training and does not rely on any particular model. Second, it  
114 can deal with both frontal and non-frontal poses. Finally, it is quite fast, re-  
115 quiring only seconds to process an image of 100 to 200 pixels (in both x and  
116 y dimensions). [Chang et al. \(2006\)](#) set two thresholds for the two curvature

117 maps to search for the nose tip, the two eyes and the nose ridge, they used  
118 a database of 4000 images of 449 different subjects. [Hutton et al. \(2003\)](#) use  
119 a hybrid *iterative closest-point* (ICP) and *active shape model* (ASM) fitting  
120 for non-rigid registration of a dense surface model on 3D faces. This method  
121 does not require texture so it imposes constraints on the orientation of the  
122 face and it is not scale invariant. On the other hand, [Nanni et al. \(2014\)](#)  
123 presents a face detector based on Viola and Jones algorithm. [Soldara et al.](#)  
124 [\(2015\)](#) proposes a method of facial recognition in 2D images based on pro-  
125 jections of high dimensionality face images into low dimensions. In his face  
126 recognition study, it is explained that remains a difficult task since 2D images  
127 of the face can be affected by changes in the scene, for example: position of  
128 the face, facial expressions, or lighting. Their recognition method is a mod-  
129 ification of the technique introduced by [Cai et al. \(2006\)](#), which proposes a  
130 face recognition method based on appearance, called orthogonal Laplacian-  
131 faces (OLPP). [Zhang et al. \(2008\)](#) proposes a method for detecting faces in  
132 2D images based on *principal component analysis* (PCA) and *support vec-*  
133 *tor machine* (SVM). First, they filter areas that are potentially a face using  
134 a statistical distribution function of the local histogram. Then, Candidate  
135 faces are reduced using a SVM classifier. [Chen et al. \(2009\)](#) proposes a face  
136 detection method using a 2D half-template face model. In this research they  
137 explain that face detection is an important process for facial recognition.  
138 Their model is an average face template containing: eyes, nose, mouth and  
139 part of the cheek. The face detection was made using a template matching  
140 method to determine the position of the face in the image. Their theoretical  
141 analysis shows that the average face template can reduce the chances of local  
142 density and that it can be adapted to side face images at an angle, which  
143 improves the accuracy of face detection in lateral position. [Belahcene et al.](#)  
144 [\(2014\)](#) proposes a five stages face recognition system (FRS): preprocessing  
145 and detection, branch, fusion features, feature extraction and classification.  
146 The first stage, is to detect the face using horizontal and vertical curves inte-  
147 grated projection. In the second stage, they divide the face using face patches  
148 consisting of the nose and eyes. In the third stage, they combine the image  
149 information of the 2D image depth, which is used by PCA and *Enhanced*  
150 *model Fisher* (EFM) to extract characteristic features. These were used to  
151 the classification stage using two methods, measures based on distance and  
152 a *support vector machine* (SVM). [Maes et al. \(2010\)](#) presents an adapted  
153 SIFT algorithm for 3D surfaces (called meshSIFT) and its applications to  
154 3D face pose normalization and recognition. Their algorithm allows reliable

155 detection of scale space as local feature locations. The meshSIFT algorithm  
 156 then describes the neighbourhood of every scale space in a feature vector  
 157 consisting of concatenated histograms of shape indices and angles. Then,  
 158 feature vectors are reliably matched by comparing the angle in feature space.

### 159 **3. Face detection in 3D images with one person**

160 This section presents an experimental analysis on three key face de-  
 161 tection approaches using 3D images with one person. Subsection 3.1 ex-  
 162 plains Colombo et al. (2006) curvature analysis. Subsection 3.2 details Mian  
 163 et al. (2006) slicing approach. Subsection 3.3 describes Segundo et al. (2007)  
 164 segmentation technique. Finally, Subsection 3.4 shows experimental results.

#### 165 *3.1. Curvature analysis*

166 Colombo et al. (2006) curvature analysis technique was experimentally  
 167 analysed and implemented considering some particularities, as follows:

168 Step 1.- Depth images were computed from its respective 3D point cloud using  
 169 equations:

$$f \rightarrow U \text{ defined on an open set } U \subseteq R^2 \quad (1)$$

$$S = (x, y, z) | (x, y) \in U; z \in R; f(x, y) = z \quad (2)$$

170 Step 2.- A Gaussian filter was applied (Equation 3) to the input depth image to  
 171 discard high-frequency fluctuations in the surface, while salient facial  
 172 features, such as the eyes and nose, are still clearly distinguishable.  
 173 When the filter is applied to the whole image some regions as corners  
 174 or face contours are not discarded due radial distortion.

$$\left(\frac{1}{16}\right) * \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

175 Step 3.- Mean (4) and Gaussian (5) curvature were calculated by the first (6)  
 176 and second (7) derivatives from every point of the depth map as in [Trucco](#)  
 177 and [Verri \(1998\)](#).

$$H(x, y) = \frac{(1 + f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1 + f_x^2)f_{yy}}{2(1 + f_x^2 + f_y^2)^{\frac{3}{2}}} \quad (4)$$

$$K(x, y) = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2} \quad (5)$$

178 where  $f_x, f_y, f_{xy}, f_{xx}, f_{yy}$  are the first and second derivatives of  $f$  in  
 179  $(x, y)$ . Then, a face is initially represented as a range image. Since we  
 180 have only a discrete representation of  $S$ , the partial derivatives were  
 181 also calculated for each point of the depth map:

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} + O(h^2) \quad (6)$$

$$f''(x_0) = \frac{f(x_0 + h) - 2f(x_0) + 2f(x_0 - h)}{h^2} + O(h^2) \quad (7)$$

182 considering that  $x_0$  is a  $(x, y)$  point in the range image,  $h$  is an integer  
 183 greater than zero, and  $O(h^2)$  is the *truncation error*, caused by stopping  
 184 the polynomial approximation to second order, which tends to zero.

185 Step 4.- HK classification were performed according to the Mean and Gaussian  
 186 curvature signs. Then, elliptical convex points can be identified when  
 187  $K > 0$  and  $H < 0$ , and those points could represent a nose (see [Table](#)  
 188 [1](#)).

Table 1: HK classification was introduced by [Besl and Jain \(1986\)](#).

|       | K < 0                | K = 0               | K > 0              |
|-------|----------------------|---------------------|--------------------|
| H < 0 | Hyperbolic concave   | Cylindrical concave | Elliptical concave |
| H = 0 | Hyperbolic symmetric | Planar              | Impossible         |
| H > 0 | Hyperbolic convex    | Cylindrical convex  | Elliptical convex  |



189 Step 5.- A thresholding process (Equation 8) by curvature values is now per-  
 190 formed. Regions with highest curvature are isolated as shown in Fig-  
 191 ure 3 as nose candidates, while areas with low curvature values are  
 192 discarded.

$$|H(u, v)| \geq T_h \quad |k(u, v)| \geq T_k \quad (8)$$

193 where  $T_h$  and  $T_k$  are our estimated thresholds:

$$T_k = (k_{max} - k_{min})/4 \quad (9)$$

$$T_h = (h_{max} - h_{min})/4 \quad (10)$$

194 Step 6.- Finally, in the last step the number of candidates was reduced by fil-  
 195 tering each elliptical convex candidate, considering average Mean ( $\overline{H}_l$ )  
 196 and Gaussian curvature ( $\overline{K}_l$ ):

$$\overline{H}_l \geq \overline{H}_{min} \quad (11)$$

$$\overline{K}_l \geq \overline{K}_{min} \quad (12)$$

### 197 3.2. Slicing analysis

198 Mian et al. (2006) slicing technique was also studied in detail and our  
 199 implementation is as follows:

200 Step 1.- A 3D image is horizontally sliced in multiple steps using a threshold  
 201  $dv$ . For each slice, spikes are removed using  $dt$ . The  $dt$  is automatically  
 202 calculated by using equation 13.

$$dt = \mu + 0.6\sigma \quad (13)$$

203 where  $\mu$  is the mean distance between eight neighbouring points with  
 204 standard deviation  $\sigma$ .

205 Step 2.- After removing spikes, the 3D image may result in holes. Those holes  
206 were refilled by cubic interpolation. Then, circles were created centred  
207 at multiple horizontal intervals  $dh$  on each slice for a region segmen-  
208 tation. In each segmented area, a triangle was inscribed from the centre  
209 of the circle and two intersection points coinciding by the circle and  
210 slice segmented points. The circles were created with a 2cm radius  
211 associated to the mean nose size from the tip to the end of the nose.

212 Step 3.- In each slice must be a maximum altitude triangle. This triangle can  
213 be considered as a potential nose tip. A confidence value, equal to  
214 triangle's altitude is assigned to it for further comparison with the  
215 other maximum altitude triangles from other slices.

216 Step 4.- After getting the confidence values from Step 3.- associated to all the  
217 slices, one nose tip candidate point per slice is identified. Those candi-  
218 dates were analysed for nose ridge identification. The candidate points  
219 that do not correspond to the nose ridge were considered as outliers  
220 and were removed using *RANSAC* (Trucco and Verri, 1998).

221 Step 5.- Finally, the few candidate points were evaluated and the one with max-  
222 imum altitude was defined as the nose tip. Then, the whole face can  
223 be extracted from detected nose tip.

### 224 3.3. Segmentation technique

225 Our Segundo et al. (2007) slicing technique implementation was as follows:

226 Step 1.- The input depth images (see Figure 1-a) for this algorithm were gen-  
227 erated from a 3D point cloud by using equations (1) and (2). The  
228 complete process consists of six stages: K-Mean segmentation, face re-  
229 gion extraction, Sobel operation, closing, ellipse detection, binarization  
230 and segmentation (see Figure 1).

231 Step 2.- K-Means algorithm Krishna and Murty (1999) is applied by setting  $k =$   
232 3, so the input image was segmented in three main regions: background,  
233 body, and face (see Figure 1-b). The whole background is in black,  
234 the body is in gray and the region of interest, the face, is in white (see  
235 Figure 1-c). Only this white region is extracted and is used in following  
236 steps.

237 Step 3.- In this step, edge detection is applied using Sobel operator and the face  
238 region is now represented by edges (see Figure 1-d).

239 Step 4.- Then, only the bigger edge is kept (see Figure 1-e) to which an ellipse  
240 is fit using Hough Transform. The resulting image is a clean binary  
241 image (see Figure 1-f) with the possible face shape.

242 Step 5.- Finally, to obtain an edge map with the detected and segmented face, a  
243 closing process was applied to the resulting image in the last step. By  
244 using an *and* operation with this image and the original input depth  
245 image, the face region is obtained.

### 246 3.4. Experimental Results

247 We have performed an experimental procedure for comparison and analy-  
248 sis of those techniques using *CurtinFaces* (Li et al., 2013) and *FRGC* (Phillips  
249 et al., 2005) databases. These databases were considered because, the FRGC  
250 database is the most commonly used in face processing research, and Curtin-  
251 Faces database was collected with Kinect 360 sensor and images are similar  
252 to those collected in this investigation. A performance summary is shown in  
253 Table 2.

Table 2: Performance summary: Face detection in 3D images with one person.

| Technique          | Overall performance FRGC [%] | Overall performance CurtinFaces [%] |
|--------------------|------------------------------|-------------------------------------|
| Curvature Analysis | 98.48%                       | 75.18%                              |
| Slicing analysis   | 92.74%                       | 57.47%                              |
| Face segmentation  | 97.19%                       | 46.88%                              |

254 We implement every face detection approach by developing specific pur-  
255 pose functions or using standard available libraries in MATLAB version 15.  
256 For the three face detection algorithms, their performance is calculated by  
257 counting those facial areas that contain the tip of the nose landmark ac-  
258 cording reference ground truth Creusot et al. (2010). For slicing technique  
259 we used only the tip of the nose to detect a face. For the curvature and  
260 segmentation techniques we used the corners of the eyes and the tip of the  
261 nose.

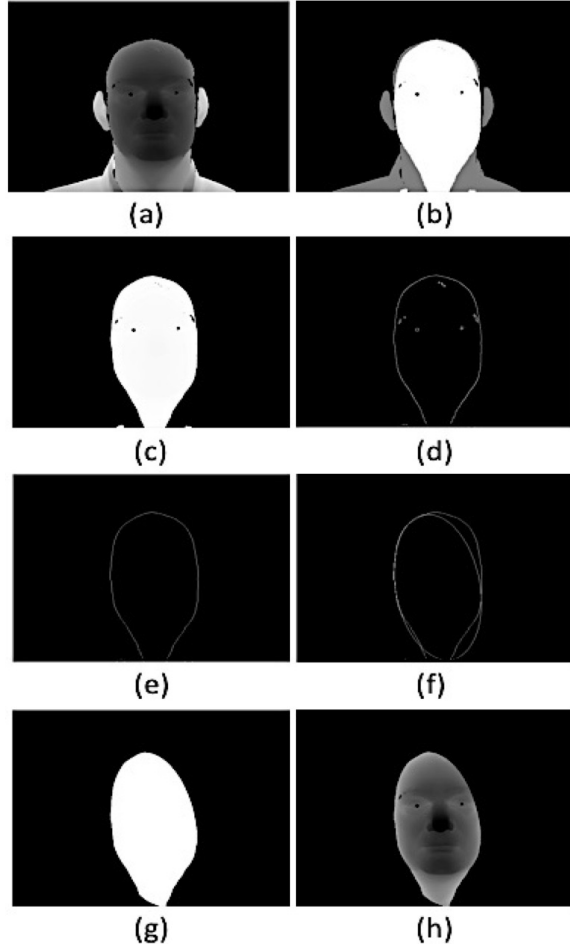


Figure 1: Illustration of our [Segundo et al. \(2007\)](#) implementation: (a) range image, (b) K-Means algorithm result, (c) face region extracting, (d) Sobel operator result, (e) closing process result, (f) ellipse detection with Hough Transform (g), binary image after normalization, and (h) final segmentation.

262 **4. Face detection in 3D images with more than one person**

263 This section investigates the face detection problem with images that  
 264 contains more than one person in the scene. Subsection 4.1 introduces our  
 265 face detection approach. Subsection 4.2 describes our experimental image

266 acquisition. Finally, Subsection 4.3 shows our experimental results.

#### 267 4.1. Face Detection Approach

268 Taking in consideration results and knowledge obtained from the experi-  
 269 mental analysis described in Section 3, we propose a novel approach that is  
 270 able to detect every present face within an image that contains more than  
 271 one person in the scene. As illustrated in Figure 2 this approach consists of  
 272 four steps: a) Elimination of planar regions, b) Curvature analysis, c) Nose  
 273 tip candidate selection, and d) Face/No-Face classification. A description in  
 274 detail of this steps is as follows.

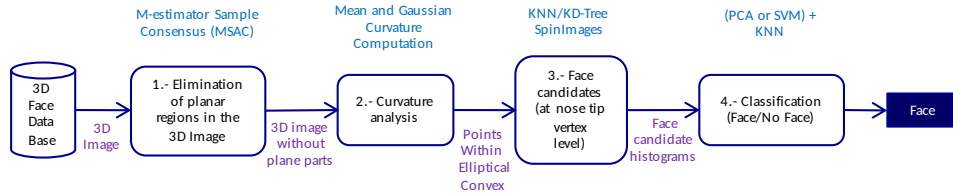


Figure 2: Experimental procedure to detect every face in a 3D image consisting of four main steps.

275 Step 1.- For every 3D image, planar areas were eliminated using M-estimator  
 276 Sample Consensus (MSAC). MSAC is a modification of the RANSAC  
 277 technique, where the aim is to adjust planes in the point cloud using a  
 278 distance  $d$ , when set the plane the distance of the points is calculated  
 279 to it in order to obtain two sets of points, inliers and outliers.

280 The inliers are the vertices that fit the calculated plane and the outliers  
 281 are those that distance to the plane is greater that  $d$ . Our MSAC  
 282 technique implementation consists of five steps:

- 283 (a) The input of the MSAC algorithm is a point cloud and a distance  
 284  $d$ .

- 285 (b) A plane  $P1 \cdot X + P2 \cdot Y + P3 \cdot Z + P4 = 0$  is adjusted to a part of  
 286 the 3D image with the operation:

$$\begin{bmatrix} X1 & Y1 & Z1 & 1 \\ X2 & Y2 & Z2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ Xn & Yn & Zn & 1 \end{bmatrix} \begin{bmatrix} P1 \\ P2 \\ P3 \\ P4 \end{bmatrix} = A\bar{p} \quad (14)$$

- (c) The optimal parameters are calculated. We defined the optimal parameters as the minimum values of plane adjustment:

$$\bar{p}_{opt} = \operatorname{argmin} \|A\bar{p}\|^2 \quad (15)$$

- (d) Calculate the estimation error of the distance of each point to the plane:

$$e^2 = \frac{([x \ y \ z \ 1]\bar{p}_{opt})^2}{P_1^2 + P_2^2 + P_3^2} \quad (16)$$

- 287 (e) Select the inliers using a distance  $d$ . If the distance of a point is  
 288 less or equal than the value of  $d$  it is considered inliers, otherwise  
 289 are outliers.

290 We selected the outlier because they are points that are not planar  
 291 regions.

292 **Step 2.-** Then, we used curvature analysis for detecting convex elliptical points,  
 293 the steps for implementation of the curvature analysis are:

- 294 (a) We started with a range image that is a representation of the  
 295 coordinates  $(x, y, z)$  of the 3D image in a location  $(i, j)$ . The  
 296 depth images are computed from respective 3D point cloud using  
 297 equations (1) and (2).  
 298 (b) For every point within the depth map, Mean (4) and Gaussian (5)  
 299 curvature are computed.  
 300 (c) Then, by analyzing the signs of the Mean and the Gaussian cur-  
 301 vature, we perform what is called an HK classification (Besl and  
 302 Jain, 1986). For this case, we have only selected the convex ellip-  
 303 tical points as illustrated in Figure 3.



Figure 3: The black vertices are convex elliptical points

304 At this point, we mostly got those points around the subjects body and  
 305 head.

306 Step 3.- For every vertex in the 3D point cloud, *Spin Images*, as prescribed by  
 307 [Johnson and Hebert \(1998\)](#) were computed using the following steps:

- 308 (a) Select a point  $P$  in the point cloud.  
 309 (b) Calculate the normal of  $P$  using a radius  $r$  to select neighbours of  
 310  $P$   
 (c) Compute  $\alpha$  and  $\beta$  values:

$$\alpha = \sqrt{||x - p||^2 - (n \cdot (x - p))^2} \quad (17)$$

$$\beta = (n \cdot (x - p)) \quad (18)$$

- (d) Every alpha and beta coordinates contribute to its respective (i, j) spin image histogram:

$$i = \left[ \frac{\beta_{max} - \beta}{bin} \right] \quad , \quad j = \left[ \frac{\alpha}{bin} \right] \quad (19)$$

311 where bin is a predefined constant, which in this research was  
 312 experimentally calculated as 0.2.

313 To speed up the selection process to create a spin image histogram, we  
 314 are using *Nearest Neighbour* and *KD-Tree* algorithms. To calculate the  
 315 nearest neighbours of each vertex KD-tree was used for partitioning  
 316 the space that organizes the points in a k-dimensional Euclidean space.  
 317 With this partitioning, we can get the neighbours of each vertex using  
 318 KNN, an 8cm distance was used for neighbours selection to generate the  
 319 spin image histogram. Figure 4 shows spin image histograms computed  
 320 from three different databases at the nose tip level.

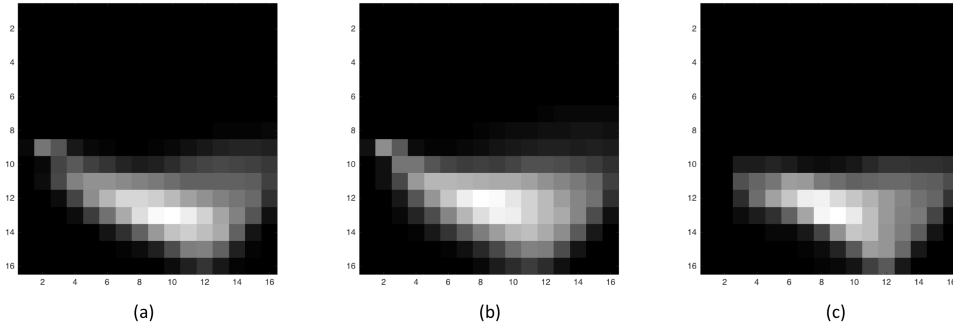


Figure 4: Spin image histogram samples: (a) FRGC (b) Curtin Faces (c) BDUAEMex.

321 Step 4.- Face/No-face classification was done using PCA/SVM classifiers and  
 322 KNN.

323 For PCA classification we compare spin image histograms of remaining  
 324 nose tip candidates. For this, we see a  $[p \times q]$  histogram as an  $[m \times 1]$   
 325 feature vector in an  $m = pq$  dimensional space. Separate training and  
 326 testing sets were defined and we used a one-nearest neighbour scheme  
 327 as proposed by Pears et al. (2012) based on Mahalanobis distance as  
 328 follows:

329 ***PCA training***

- (a) For the set of  $n$  training images,  $x_i$ ,  $i = 1 \dots n$ , where each training face is represented as an  $m$ -dimensional column vector:

$$x = [x_1, \dots, x_m]^T \quad (20)$$



stack the  $n$  training face vectors together to construct the  $[n \times m]$  training data matrix:

$$X = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \end{bmatrix} \quad (21)$$

330 where each column vector represents Spin Image histogram . In  
 331 this experiment,  $[16 \times 16]$  histograms from 200 tips of the nose were  
 332 used.

333 (b) The mean of the training set is calculated:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (22)$$

334 to form the zero-mean training data matrix:

$$X_{zmean} = X - J_{(n,1)} \bar{x}^T \quad (23)$$

335 where  $J_{(n,1)}$  is an  $n \times 1$  matrix of ones.

336 (c) Perform a standard eigen decomposition. Several variants of PCA  
 337 based 3D face recognition or classification exist in the literature  
 338 (Colombo et al., 2006; Bouzalmat et al., 2014; Gumus et al., 2010;  
 339 Pears et al., 2012). We used Singular Value Decomposition (SVD)  
 340 directly on the  $n \times m$  zero-mean training data matrix,  $X_{zmean}$ .  
 341 The advantage of using SVD is that it can often provide superior  
 342 numerical stability compared to eigen decomposition algorithms,  
 343 additionally, the storage required for a data matrix is often much  
 344 less than a covariance matrix (Pears et al., 2012). The *SVD* is:

$$USV^T = X_{zmean} \quad (24)$$

345 where  $U$  and  $V$  are orthogonal matrices of dimension  $n \times n$  and  
 346  $m \times m$  respectively and  $S$  is a  $n \times m$  matrix of singular values along  
 347 its diagonal.

348 (d) Select the number of subspace dimensions for projection. This is  
 349 the *dimensionality reduction* step and is usually done by analysing  
 350 the ratio of cumulative variance associated with the first  $k$  dimen-  
 351 sions of the rotated image space to the total variance associated  
 352 with the full set of  $m$  dimensions in that space.

353 (e) Project the training data set into the k-dimensional subspace:

$$\tilde{X} = X_{zmean} V_k \quad (25)$$

354 where  $V_k$  is a  $m \times k$  matrix containing the first  $k$  eigenvectors of  
355  $V$  and  $\tilde{X}$  is a  $n \times k$  matrix of  $n$  training faces in the k-dimensional  
356 subspace.

### 357 *PCA testing*

358 Once the above PCA training phase is completed, it is straightforward  
359 to implement a simple nearest neighbour face identification scheme,  
360 within the reduced k-dimensional space.

(a) Project the testing set by subtracting the mean vector times the  
derived subspace PCA  $V_k$ :

$$\tilde{X}_t^T = (X_t - \bar{X})^T V_k \quad (26)$$

361 (b) Compute the Euclidean distance between every face in the training  
362 data set against the testing set. The testing candidate with the  
363 smallest distance is considered the face nose tip.

### 364 *Support Vector Machine (SVM)*

365 Support vector machine (*SVM*) is considered a supervised learning  
366 method. The basics of *SVM* are rooted in the work on statistical  
367 learning theory and were introduced in the 90s by [Vapnik et al. \(1996\)](#).  
368 Within the task of classification, *SVM* belong to the category of linear  
369 classifiers, since they induce linear or hyperplanes separators: separable,  
370 quasi-separable, transformed space ([Gumus et al., 2010](#); [Bouzal-  
371 mat et al., 2014](#); [Tello et al., 2013](#)). While most learning methods focus  
372 on minimizing the errors computed by the generated model from the  
373 training sample, the inductive bias associated with the *SVM* lies in  
374 minimizing the structural risk ([Vapnik et al., 1996](#)).

375 For the implementation of *SVM*, the 3D vertexes have been filtered  
376 previously as candidate to nose tip. The idea is to select a separation  
377 hyperplane equidistant from the closest examples of every class,  
378 in other words, to get what is called a margin maximum on either side

379 of the hyperplane margin. Moreover, when defining the hyperplane,  
 380 only the training examples of each class that lay in the border of these  
 381 margins are considered (support vectors). Therefore, support vectors  
 382 are subsets of training observations that are used to support an optimal  
 383 location decision surface (Vapnik et al., 1996).

384 We train SVM with 200 correct nose tip histograms and 200 nose tip  
 385 shaped histograms that were actually: clothes, eyebrows or chins. As  
 386 we can see in Table 3, correct and incorrect nose tip histograms are la-  
 387 belled as 1 and -1, respectively. The SVM implementation were carried  
 388 out using Matlab Support Vector Machines toolbox for Classification.

Table 3: Data encoding matrix exemplifying the distribution of the SVM training.

|           | $c_1$   | $c_2$   | $c_3$   | $c_4$   | $c_5$   | $c_6$   | $\dots$ | $c_{256}$ | $d_1$   |
|-----------|---------|---------|---------|---------|---------|---------|---------|-----------|---------|
| $P_1$     | 1       | 0       | 1       | 1       | 0       | 1       | $\dots$ | 1         | 1       |
| $P_2$     | 0       | 1       | 0       | 1       | 0       | 1       | $\dots$ | 0         | -1      |
| $P_3$     | 1       | 0       | 1       | 1       | 0       | 1       | $\dots$ | 1         | 1       |
| $\dots$   | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$   | $\dots$ |
| $P_{400}$ | 1       | 0       | 1       | 1       | 0       | 1       | $\dots$ | 1         | 1       |

389 We used an optimized Matlab function (*svmtrain*) to identify support  
 390 vectors  $s_i$ , weights  $\alpha_i$ , and bias  $b$  for a given kernel  $k$  to classify a testing  
 391 set  $x$ :

$$c = \sum_i \alpha_i k(s_i, x) + b \quad (27)$$

392 We use KNN to calculate the distances of our training set and candi-  
 393 dates to nose tip, the point with the shortest distance is selected.

394 Finally, from that selected nose tip the facial surface is extracted from  
 395 the image using an 8cm radio.

#### 396 4.2. Experimental data acquisition

397 As we know, state of the art 3D databases have only one person per  
 398 image and they were collected in controlled illumination, mostly neutral facial  
 399 expressions and front pose (Senthilkumar and Gnanamurthy, 2014). Hence,  
 400 there are no 3D experimental images for the purpose of this research.

401 To investigate face detection in 3D images with one person, we are using  
 402 two state of the art databases: Face Recognition Grand Challenge (Phillips

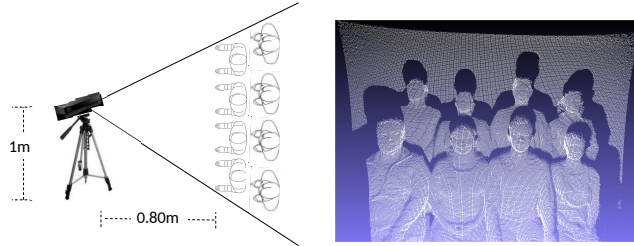


Figure 5: A one-depth image with eight people in the scene

403 [et al., 2005](#)) and CurtinFaces ([Li et al., 2013](#)). From the FRGC and Curtin-  
 404 faces databases, we are using 4950 and 1437 3D images, respectively.

405 For our face detection experiment in 3D images with more than one per-  
 406 son, a capture session was performed using the Kinect One<sup>TM</sup> sensor, to  
 407 collect 2D/3D images varying the number and position of eight subjects in  
 408 the scene. Figure 5 shows how people were placed within the field of view of  
 409 the 3D sensor.

410 Our data collection is limited to eight persons per scene, because the  
 411 Kinect One<sup>TM</sup> sensor field of view is not able to capture more people without  
 412 reducing the number of vertices per face. This is well known as a challenge  
 413 in 3D face processing [Bowyer et al. \(2006\)](#).

414 Four sets of images were collected varying in each scene the number and  
 415 the position of eight subjects, giving  $2^8$  different positions and counting 1020  
 416 images in total not considering images with no subject.

417 For training and performance evaluation it was necessary to collect ground  
 418 truth data. As we are assuming that a face exists into an image if its nose is  
 419 present. Therefore, we meticulously collect ground truth values at nose tip  
 420 vertex level, by manually identifying the nose as the most prominent part in  
 421 a face in a MATLAB 3D plot.

422 To implement and evaluate our experimental procedure to detect any  
 423 present face into a 3D image, using our set of 1020 depth images we defined  
 424 different training and testing sets as indicated in Table 4.

Table 4: Training and testing sets using depth images containing from one to eight different people.

| Set      | Number of images | Number of faces |
|----------|------------------|-----------------|
| Training | 100              | 200             |
| Testing  | 920              | 3896            |
| Total    | 1020             | 4096            |

### 4.3. Experimental Results

Using our face data set containing from one to eight persons in each scene we have experimentally evaluated our face detection procedure. In this case, from the resulting spin image classification we find first the local maximum vertex for every cluster of candidates. Then, localization errors were computed as the Euclidean distance between the localised nose tip and its respective ground truth value. Table 5 summarises our face detection performance. As we can see, we are successfully detecting 91% of faces in our experimental face images, 100% of faces in the FRGC database and 98% of faces in the CurtinFaces database within a localization error lower or equal to 16mm (see Figure 6 and 7).

## 5. Conclusions

In this paper we have presented a novel pose invariant experimental procedure for automatic face detection in 3D images, which consist of four steps: Planar region elimination, curvature analysis, spin image analysis and classification. This localisation approach was evaluated using our 3D image data set (that contains from one to eight different subjects) and state of the art databases (FRGC and CurtinFaces) that were collected considering only one person per image.

To begin our research, three key approaches in related literature were experimentally investigated toward our final aim. The first approach uses a curvature analysis to detect the most rigid facial area where the inner-eyes corners and the tip of the nose are expected. The second approach horizontally slices an image to localize the most likely facial area that contains the tip of the nose. The third approach uses a segmentation procedure to detect an elliptical area where a human face is expected.

For the three experimented datasets (Our database with images with more than one subject, FRGC and CurtinFaces) we have gathered location errors by computing the Euclidean distance from the selected nose tip and

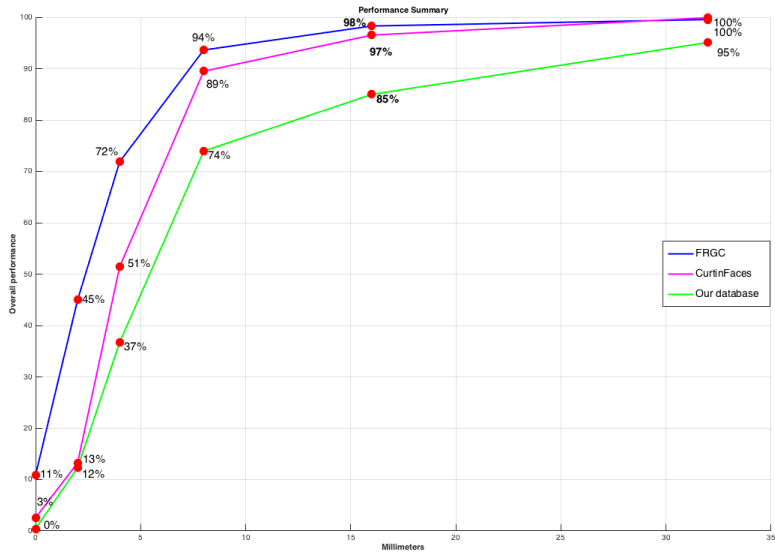


Figure 6: Performance summary in our face detection approach using PCA

454 their respective ground-truth. Our best experimental results, using a SVM  
 455 classifier, shows that within a localization error between 0 to 16 mm, 100%,  
 456 98% and 91% of faces in every image in the FRGC, CurtinFaces and our  
 457 dataset are located.

458 Despite of the Kinect One<sup>TM</sup> low resolution and limited field of view,  
 459 our face detection approach has been shown robust for nose tip localisation.  
 460 Then, our approach assume the presence of a face by locating its nose tip.

461 On the other hand, our face detection approach as been experimented  
 462 using pre-stablish thresholds, which were experimentally calculated for each  
 463 step.

464 As part of the future work, we are studying alternate techniques to im-  
 465 prove our face detection approach performance. Additionally, it is mandatory  
 466 to consider a face extraction technique for any counted face within an image  
 467 to complete our process.

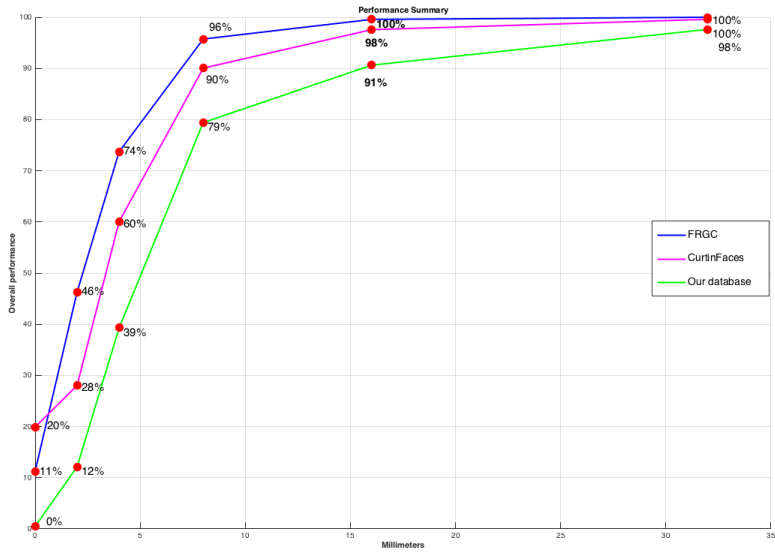


Figure 7: Performance summary in our face detection approach using SVM

468 **Acknowledgments**

469 Authors thank to the Research Department of the Autonomous Univer-  
 470 sity of the State of Mexico (UAEMex) for their financial support, grant  
 471 3720/2014/CID. Also, our gratitude goes to CONACYT for Mr. Paduano’s  
 472 scholarship 634473.

Table 5: Face detection summary using FRGC, CurtinFaces and our point cloud images containing from one to eight subjects at 16mm.

| Data Bases   | Performance using PCA [%] | Performance using SVM [%] |
|--------------|---------------------------|---------------------------|
| FRGC         | 98%                       | 100%                      |
| CurtinFaces  | 97%                       | 98%                       |
| Our DataBase | 85%                       | 91%                       |

473 **References**

- 474 Aghaei, M., Dimiccoli, M., Radeva, P., 2016. Multi-face tracking by extended  
475 bag-of-tracklets in egocentric photo-streams. *Computer Vision and Image*  
476 *Understanding* 149, pp. 146 – 156, special issue on Assistive Computer  
477 *Vision and Robotics - Assistive Solutions for Mobility, Communication*  
478 *and HMI*.
- 479 Ariz, M., Bengoechea, J. J., Villanueva, A., Cabeza, R., 2016. A novel 2d/3d  
480 database with automatic face annotation for head tracking and pose esti-  
481 mation. *Computer Vision and Image Understanding* 148, pp. 201 – 210,  
482 special issue on Assistive Computer Vision and Robotics - Assistive Solu-  
483 tions for Mobility, Communication and HMI.
- 484 Belahcene, M., Chouchane, A., Mokhtari, N., Nov 2014. 2d and 3d face  
485 recognition based on ipc detection and patch of interest regions. In: *Con-*  
486 *connected Vehicles and Expo (ICCVE), 2014 International Conference on*. pp.  
487 627–628.
- 488 Besl, P. J., Jain, R. C., 1986. Invariant surface characteristics for 3d ob-  
489 ject recognition in range images. *Computer Vision, Graphics, and Image*  
490 *Processing* 33 (1), pp. 33 – 80.
- 491 Bhandarkar, S. M., Luo, X., 2009. Integrated detection and tracking of mul-  
492 tiple faces using particle filtering and optical flow-based elastic matching.  
493 *Computer Vision and Image Understanding* 113 (6), pp. 708 – 725.
- 494 Bouzalmat, A., Kharroubi, J., Zarghili, A., 2014. Comparative study of pca,  
495 ica, lda using svm classifier. *Journal of Emerging Technologies in Web*  
496 *Intelligence* 6 (1), pp. 64–68.
- 497 Bowyer, K. W., Chang, K., Flynn, P., 1 2006. A survey of approaches and  
498 challenges in 3d and multi-modal 3d + 2d face recognition. *Computer*  
499 *Vision and Image Understanding* 101 (1), pp. 1–15.
- 500 Cai, D., He, X., Han, J., Zhang, H. J., Nov 2006. Orthogonal laplacianfaces  
501 for face recognition. *IEEE Transactions on Image Processing* 15 (11), pp.  
502 3608–3614.
- 503 Chang, K. I., Bowyer, K. W., Flynn, P. J., Oct 2006. Multiple nose region  
504 matching for 3d face recognition under varying facial expression. *IEEE*



- 505 Transactions on Pattern Analysis and Machine Intelligence 28 (10), pp.  
506 1695–1700.
- 507 Chellappa, R., Wilson, C., Sirohey, S., May 1995. Human and machine recog-  
508 nition of faces: a survey. Proceedings of the IEEE 83 (5), pp. 705–741.
- 509 Chen, W., Sun, T., Yang, X., Wang, L., Aug 2009. Face detection based on  
510 half face-template. In: Electronic Measurement Instruments, 2009. ICEMI  
511 '09. 9th International Conference on. pp. 4–54–4–58.
- 512 Colombo, A., Cusano, C., Schettini, R., Mar 2006. 3d face detection using  
513 curvature analysis. Pattern Recognition 39 (3), pp. 444 – 455.
- 514 Creusot, C., Pears, N., Austin, J., October 2010. 3d face landmark labelling.  
515 In: Proceedings of the ACM workshop on 3D object retrieval. ACM, As-  
516 sociation for Computing Machinery, pp. 27–32.
- 517 Gumus, E., Kilic, N., Sertbas, A., Ucan, O. N., 2010. Eigenfaces and support  
518 vector machine approaches for hybrid face recognition. pattern recognition  
519 8, pp. 9.
- 520 Hjelmås, E., Low, B. K., 2001. Face detection: A survey. Computer Vision  
521 and Image Understanding 83 (3), pp. 236 – 274.
- 522 Hu, W., Hu, W., Maybank, S., April 2008. Adaboost-based algorithm for  
523 network intrusion detection. IEEE Transactions on Systems, Man, and  
524 Cybernetics, Part B (Cybernetics) 38 (2), pp. 577–583.
- 525 Hutton, T. J., Buxton, B. F., Hammond, P., 2003. Automated registration of  
526 3d faces using dense surface models. In: In Proc. British Machine Vision  
527 Conference. pp. 439–448.
- 528 Johnson, A. E., Hebert, M., 7 1998. Surface matching for object recognition  
529 in complex three-dimensional scenes. Image and Vision Computing 16 (9–  
530 10), pp. 635–651.
- 531 Kakadiaris, I. A., Toderici, G., Evangelopoulos, G., Passalis, G., Chu, D.,  
532 Zhao, X., Shah, S. K., Theoharis, T., 2016. 3d-2d face recognition with  
533 pose and illumination normalization. Computer Vision and Image Under-  
534 standing, –.

- 535 Krishna, K., Murty, M. N., 1999. Genetic k-means algorithm. *Systems, Man,*  
536 *and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 29 (3), pp.  
537 433–439.
- 538 Kumar, A., Datta, M., Kumar, P. B., 2015. *Face Detection and Recognition:*  
539 *Theory and Practice.* Chapman and Hall/CRC.
- 540 Li, B. Y. L., Mian, A. S., Liu, W., Krishna, A., Jan 2013. Using kinect for face  
541 recognition under varying poses, expressions, illumination and disguise. In:  
542 *Applications of Computer Vision (WACV), 2013 IEEE Workshop on.* pp.  
543 186–192.
- 544 Maes, C., Fabry, T., Keustermans, J., Smeets, D., Suetens, P., Vander-  
545 meulen, D., 2010. Feature detection on 3d face surfaces for pose normali-  
546 sation and recognition. In: *Biometrics: Theory Applications and Systems*  
547 *(BTAS), 2010 Fourth IEEE International Conference on.* pp. 1–6.
- 548 Mian, A., Bennamoun, M., Owens, R., 2006. Automatic 3d face detection,  
549 normalization and recognition. In: *3D Data Processing, Visualization, and*  
550 *Transmission, Third International Symposium on.* pp. 735–742.
- 551 Nair, P., Cavallaro, A., 2009. 3-d face detection, landmark localization, and  
552 registration using a point distribution model. *Multimedia, IEEE Transac-*  
553 *tions on* 11 (4), pp. 611–623.
- 554 Nanni, L., Lumini, A., Dominio, F., Zanuttigh, P., 1 2014. Effective and  
555 precise face detection based on color and depth data. *Applied Computing*  
556 *and Informatics* 10 (1–2), pp. 1–13.
- 557 Paduano, J., Romero, M., Muñoz, V., Jul 2015. Toward face detection in 3d  
558 data. In: *International Conference on Image Processing, Computer Vision,*  
559 *& Patter recognition. Vol. 15.* pp. 473–479.
- 560 Paduano, J., Romero, M., Valdivinos, R., Jul 2016. Face detection in 3d  
561 images with more than one person. In: *International Conference on Image*  
562 *Processing, Computer Vision, & Patter recognition.* pp. 287–293.
- 563 Pan, H., Zhu, Y., Xia, L., 2013. Efficient and accurate face detection using  
564 heterogeneous feature descriptors and feature selection. *Computer Vision*  
565 *and Image Understanding* 117 (1), pp. 12 – 28.

- 566 Pears, N., Liu, Y., Bunting, P., 2012. 3D Imaging, Analysis and Applications,  
567 1st Edition. Vol. 1. Springer-Verlag London.
- 568 Peng, X., Bennamoun, M., S., M. A., 2011. A training-free nose tip detection  
569 method from face range images. *Pattern Recognition* 44 (3), pp. 544 – 558.
- 570 Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman,  
571 K., Marques, J., Min, J., Worek, W., 2005. Overview of the face recognition  
572 grand challenge. *Computer Vision and Pattern Recognition, 2005. CVPR*  
573 2005. IEEE Computer Society Conference on 1, pp. 947–954.
- 574 Segundo, M. P., Queirolo, C., Bellon, O. R. P., Silva, L., 2007. Automatic  
575 3d facial segmentation and landmark detection. In: *Image Analysis and*  
576 *Processing, 2007. ICIAP 2007. 14th International Conference on*. pp. 431–  
577 436.
- 578 Senthilkumar, R., Gnanamurthy, R. K., April 2014. A detailed survey on 2d  
579 and 3d still face and face video databases part i. In: *Communications and*  
580 *Signal Processing (ICCSP), 2014 International Conference on*. pp. 1405–  
581 1409.
- 582 Smeulders, A. W. M., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A.,  
583 Shah, M., 2014. Visual tracking: An experimental survey. *Pattern Analysis*  
584 *and Machine Intelligence, IEEE Transactions on* 36 (7), pp. 1442–1468.
- 585 Soldera, J., Behaine, C. A. R., Scharcanski, J., Sept 2015. Customized or-  
586 thogonal locality preserving projections with soft-margin maximization for  
587 face recognition. *IEEE Transactions on Instrumentation and Measurement*  
588 64 (9), pp. 2417–2426.
- 589 Subburaman, V. B., Marcel, S., 2013. Alternative search techniques for face  
590 detection using location estimation and binary features. *Computer Vision*  
591 *and Image Understanding* 117 (5), pp. 551 – 570.
- 592 Tello, J. C. C., Hernandez-Ramirez, D., Garcia-Sepulveda, C. A., 2013. Sup-  
593 port vector machine algorithms in the search of kir gene associations with  
594 disease. *Computers in Biology and Medicine* 43 (12), pp. 2053 – 2062.
- 595 Trucco, E., Verri, A., 1998. *Introductory Techniques for 3-D Computer Vi-*  
596 *sion*. Prentice Hall.

- 597 Valstar, M. F., Mehu, M., Jiang, B., Pantic, M., Scherer, K., 2012. Meta-  
598 analysis of the first facial expression recognition challenge. *Systems, Man,  
599 and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 42 (4), pp.  
600 966–979.
- 601 Vapnik, V., Golowich, S. E., Smola, A., 1996. Support vector method for  
602 function approximation, regression estimation, and signal processing. In:  
603 *Advances in Neural Information Processing Systems 9*. Citeseer.
- 604 Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y., 2012. To-  
605 ward a practical face recognition system: Robust alignment and illumina-  
606 tion by sparse representation. *Pattern Analysis and Machine Intelligence,  
607 IEEE Transactions on* 34 (2), pp. 372–386.
- 608 Wang, P., Ji, Q., 2007. Multi-view face and eye detection using discriminant  
609 features. *Computer Vision and Image Understanding* 105 (2), pp. 99 – 111.
- 610 Wohlhart, P., Lepetit, V., feb 2015. Learning descriptors for object recogni-  
611 tion and 3d pose estimation. In: *Computer Science: Computer Vision and  
612 Pattern Recognition*.
- 613 Yang, M., Ahuja, N., 2001. *Face Detection and Gesture Recognition for  
614 Human-Computer Interaction, 1st Edition. The International Series in  
615 Video Computing 1*. Springer US.
- 616 Yang, M., Feng, Z., Shiu, S. C. K., Zhang, L., 2 2014. Fast and robust  
617 face recognition via coding residual map learning based adaptive masking.  
618 *Pattern Recognition* 47 (2), pp. 535–543.
- 619 Yang, M., Kriegman, D., Ahuja, N., Jan 2002. Detecting faces in images: a  
620 survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*  
621 24 (1), pp. 34–58.
- 622 Yin, S., Ouyang, P., Dai, X., Liu, L., Wei, S., 2015. An adaboost-based face  
623 detection system using parallel configurable architecture with optimized  
624 computation. *IEEE Systems Journal* PP (99), pp. 1–12.
- 625 Zafeiriou, S., Zhang, C., Zhang, Z., 2015. A survey on face detection in the  
626 wild: Past, present and future. *Computer Vision and Image Understanding*  
627 138, pp. 1 – 24.

628 Zhang, J., d. Zhang, X., w. Ha, S., Oct 2008. A novel approach using pca and  
629 svm for face detection. In: Natural Computation, 2008. ICNC '08. Fourth  
630 International Conference on. Vol. 3. pp. pp. 29–33.

# Capítulo 4

## Discusión

En esta investigación se propuso un proceso para la detección de la superficie facial en imágenes 3D con variaciones en el número y posición de las personas con respecto al sensor Kinect One™. Este proceso consiste en cuatro etapas; La primera es la eliminación de parte planas de la imagen. La segunda es un análisis de curvatura para seleccionar puntos 3D convexos. La tercera es la selección de candidatos a punta de la nariz, y la cuarta es una clasificación de rostro/No rostro. Para esta clasificación se experimento el uso de dos técnicas; PCA y SVM.

Las contribuciones originales que se derivaron de este trabajo de investigación son: un proceso para la detección y localización de la superficie facial en imágenes 3D de una profundidad y un conjunto de 1020 imágenes 3D capturadas con el sensor Kinect One™ variando el número y la posición de treinta y dos sujetos. Con éstas contribuciones se cumple el objetivo general de ésta tesis.

Se evaluó el proceso de detección propuesto en este trabajo de investigación utilizando dos bases de datos del estado del arte (Face Recognition Grand Challenge y CurtinFaces) y nuestra base de datos, obteniendo resultados competitivos contra los del estado del arte.

Para confirmar los resultados obtenidos con los clasificadores *PCA* y *SVM*, se corrieron tres pruebas experimentales variando los doscientos rostros de entrenamiento de nuestra base de datos considerando las tres resoluciones con las que fue colectada (alta, media y baja). La alta resolución la etiquetamos a las imágenes donde solo se encuentra una persona, la de media resolución es donde están cuatro personas y finalmente, la baja resolución es donde existen ocho personas en la escena.

**Experimento 1:** Esta prueba se realizó utilizando imágenes que contienen una, dos y tres personas. Esta selección se realizó de manera manual, con el propósito en que el algoritmo aprendiera la resolución de las imágenes. Entonces, la ideología de esta prueba es que el algoritmo

conozca altas resoluciones. Los resultados fueron un 85 % (ver Figura 6 Capitulo 3, Sección 4.3) utilizando PCA y un 91 % (ver Figura 7 Capitulo 3, Sección 4.3) utilizando SVM como se pueden observar en la Tabla 4.1.

**Experimento 2:** Esta prueba se realizó utilizando imágenes que contienen una, dos y tres personas pero ahora la selección de las imágenes fue aleatoriamente. Esta prueba se realizó para confirmar los resultados del experimento uno. Aunque los resultados disminuyeron el 3 % como se puede observar en las Figuras 4.1. En este caso no se ve muy afectado al utilizar SpinImages en la generación de los histogramas, porque, no se esta modificando significativamente la resolución de las imágenes que se utilizan para generar el modelo medio.

**Experimento 3:** Esta prueba fue realizada utilizando imágenes 3D aleatoriamente, hasta tener un total de 200 rostros como las dos anteriores pruebas. En esta prueba se ve una disminución significativa en el rendimiento mostrado en la Figura 4.2, los motivos que nosotros estamos considerando del porque esta disminución en el rendimiento son los siguientes:

- Se ve que se esta sesgando los resultados por la técnica de SpinImages, esta técnica depende de la posición y de la resolución de las imágenes; al modificar la distancia del sujeto al sensor se ve afectada el número de vértices que inciden en el rostro.
- Existencia de oclusión en las imágenes; algunas imágenes de nuestra base de datos tienen perdida de información, esto es por el tipo de sensor utilizando ya que se quiere estudiar el número y la distancia de los sujetos al sensor.
- El número de imágenes seleccionadas para el conjunto de entrenamiento; entre mayor cantidad de imágenes para entrenar se mejora la detección del rostro. Estos experimentos se realizaron con 200 rostros, pero el número de imágenes varía.

Tabla 4.1 Porcentaje de detección facial exitosa con un error máximo de 16mm.

| Experimento | PCA [%] | SVM [%] |
|-------------|---------|---------|
| #1          | 85 %    | 91 %    |
| #2          | 82 %    | 89 %    |
| #3          | 62 %    | 74 %    |

Con la experimentación realizada en esta investigación al utilizar las imágenes de nuestra base de datos, se ha mostrado que es posible detectar y localizar la superficie facial de los individuos

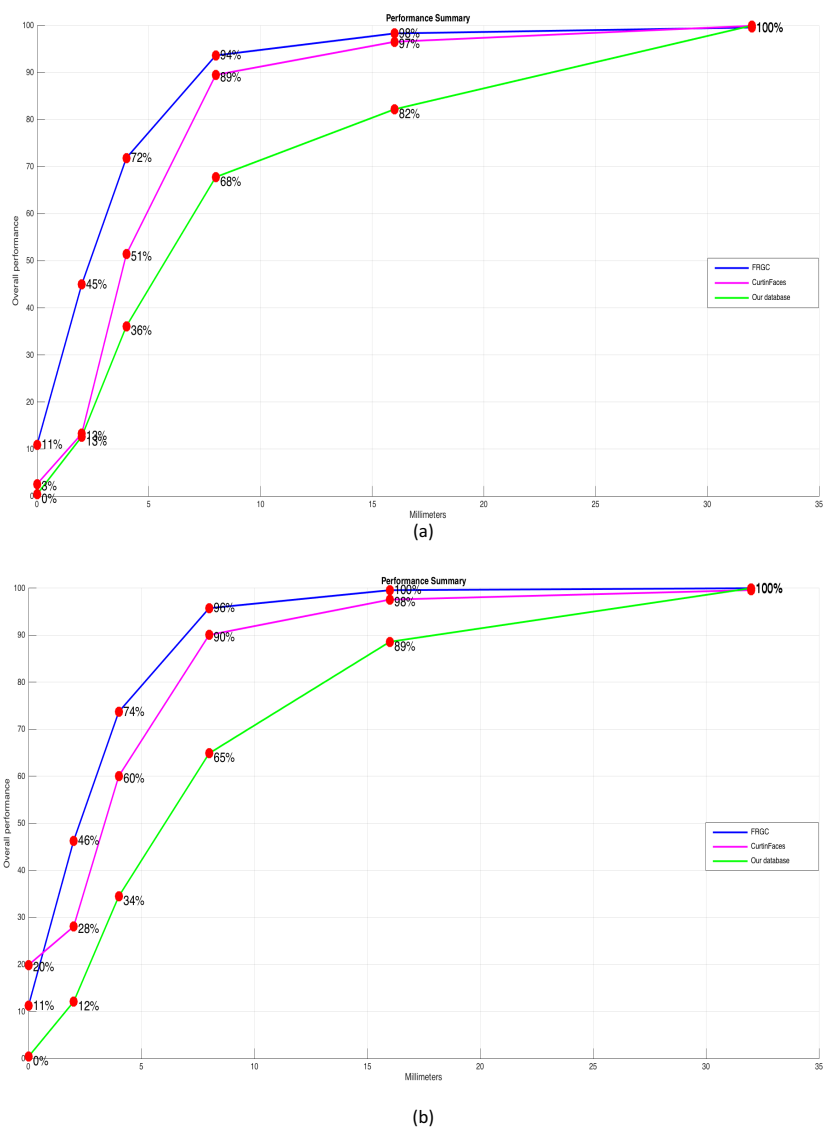


Figura 4.1 Resultados seleccionando imágenes que tienen una, dos y tres personas en la escena aleatoriamente: La imagen a) muestra los resultados utilizando PCA en tres bases de datos, mientras que en la imagen b) muestra los resultados de las mismas bases de datos pero utilizando el clasificador SVM



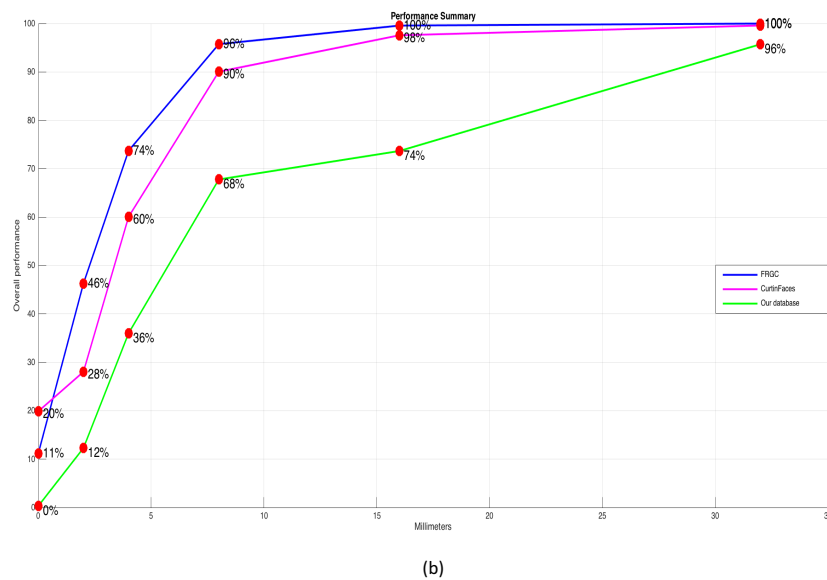
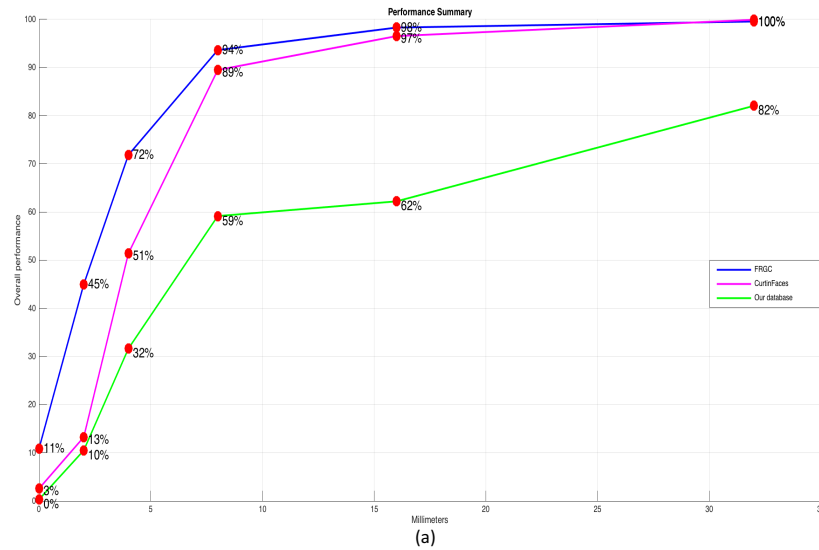


Figura 4.2 Resultados seleccionando imágenes aleatoriamente: La imagen a) muestra los resultados utilizando PCA en tres bases de datos, mientras que en la imagen b) muestra los resultados de las mismas bases de datos pero utilizando el clasificador SVM

presentes en una escena con variaciones en el número y posición de hasta ocho sujetos. Por lo tanto, se acepta la hipótesis de esta tesis.