



TELECOMUNICACION

Campus Sur
POLITÉCNICA

MEMORIA DE PRÁCTICA EXTERNA
ETS DE INGENIERÍA Y SISTEMAS DE TELECOMUNICACIÓN
UPM

<Alumno>



ÍNDICE DE CONTENIDOS

DATOS IDENTIFICATIVOS DE LAS PRÁCTICAS EXTERNAS:	1
1. INTRODUCCIÓN:	2
2. INFORMACIÓN DE LA ENTIDAD COLABORADORA.....	3
3. ENMARCAR LAS PRÁCTICAS EN EL CONTEXTO DE LA ENTIDAD	4
4. OBJETIVOS DE LAS PRÁCTICAS, TAREAS Y ACTIVIDADES REALIZADAS	5
5. TECNOLOGÍAS Y MEDIOS TÉCNICOS UTILIZADOS.....	7
6. COMPETENCIAS Y HABILIDADES ADQUIRIDAS CON LAS PRÁCTICAS	8
7. CONCLUSIONES.....	9
8. DIARIO DE PRÁCTICAS,.....	10

INSTRUCCIONES:

Se recomienda completar una cara por cada uno de los epígrafes a desarrollar, a excepción del nº4, “Objetivos de las prácticas...”, que debe tener mayor contenido (dos caras)

1. INTRODUCCIÓN:

Las prácticas se han desarrollado en la empresa CITSEM (Centro de Investigación en Tecnologías Software y Sistemas Multimedia para la Sostenibilidad). En concreto, el trabajo realizado está englobado en el área de investigación: Grupo de Redes y Servicios de Próxima Generación (GRyS) que abarcan distintas temáticas: Tecnologías de Imagen y Vídeo, Redes y servicios y Tecnologías de software y servicios.

Las actividades llevadas a cabo en estas prácticas se han orientado en torno al reconocimiento de expresiones faciales de una persona a tiempo real, utilizando el dispositivo comercial, sensor KINECT v2 modelo 1520 de Microsoft, se ha utilizado Visual Studio 2013 para poder realizar los códigos que controlaran el sensor con lenguaje C#.

El objetivo de esta práctica es entender los avances posteriores con KINECT v1, entender el estudio previo patrones de reconocimiento facial y de expresiones, empleando los trabajos de investigación realizados en este centro anteriormente, para poder finalmente trasladarlo en medida de lo posible a una implementación para detectar emociones en tiempo real.

Se procedió a realizar un estudio sobre la programación en lenguaje C# para posteriormente entender las funcionalidades de la SDK 1.8 (para la Kinect v1), entender su funcionamiento y características para posteriormente entrar al entendimiento de SDK 2.0 (para Kinect v2) y poder extraer puntos faciales. Las SDK suministradas por Microsoft para controlar los sensores Kinect son lo más importante de este trabajo.

Tras los análisis de los resultados obtenidos se proponen conclusiones y posibles trabajos futuros.

2. INFORMACIÓN DE LA ENTIDAD COLABORADORA

La entidad donde se han llevado las prácticas ha sido CITSEM (Centro de Investigación en Tecnologías Software y Sistemas Multimedia para la Sostenibilidad). Este centro de investigación tiene como objetivo el aprovechamiento de las tecnologías básicas en las que sus miembros tienen experiencia, tecnologías software y multimedia, en un contexto de sostenibilidad.

Presenta cinco misiones principales, en las que se enfocarán todas las distintas áreas de trabajo que se llevan a cabo en dicha entidad. Estas misiones son:

- La realización por propia iniciativa de proyectos y otras tareas de investigación, desarrollo e innovación en el ámbito de las tecnologías software y multimedia con especial orientación hacia la sostenibilidad, así como la difusión de los resultados alcanzados con el fin de incrementar el conocimiento científico y los recursos tecnológicos al alcance de su entorno socio-económico.
- Colaborar con otras entidades (públicas o privadas) en la realización de proyectos de investigación, desarrollo y/o innovación en el ámbito de las tecnologías software y multimedia que contribuyan al logro de sus objetivos empresariales o de servicio público.
- Realizar trabajos específicos, informes o labores de asesoría para empresas privadas o instituciones públicas en el ámbito de las tecnologías software y multimedia.
- Impartir formación de postgrado tanto para salida profesional como investigadora (doctorado).
- Organizar e impartir cursos de especialización y perfeccionamiento, seminarios, conferencias y otras actividades de análoga naturaleza, en el ámbito de las tecnologías software y multimedia.

Las áreas de trabajo en las que está dividido el CITSEM para el desarrollo de su investigación son: Grupo de Tecnología Software y Sistemas (SYST), Grupo de Diseño Electrónico y Microelectrónico (GDEM), y Grupo de Redes y Servicios de Próxima Generación (GRyS); el primero de la ETSISI y los otros dos de la ETSIST. En total cuenta con más de cuarenta investigadores a tiempo completo y se integra, como un actor más, en la estrategia de promover y fortalecer la I+D+i que se hace en el Campus Sur de la UPM.

3. ENMARCAR LAS PRÁCTICAS EN EL CONTEXTO DE LA ENTIDAD

La entidad donde se han desarrollado las prácticas, cubre tres áreas de investigación:

- Grupo de Tecnología Software y Sistemas (SYST).
- Grupo de Diseño Electrónico y Microelectrónico (GDEM).
- Grupo de Redes y Servicios de Próxima Generación (GRyS) que abarcan distintas temáticas: Tecnologías de Imagen y Vídeo, Redes y servicios y Tecnologías de software y servicios.

En el presente caso, se ha trabajado en el área GRyS. En esta área se ha realizado un estudio y pruebas de reconocimiento facial utilizando dos de los sensores que posee el centro.

El espacio de trabajo se ha situado en un laboratorio del centro de investigación CITSEM. Para la puesta en común con el resto del grupo de trabajo y tutora, se realizó en el mismo centro.

Las personas con las que se han trabajado en dicha entidad han sido:

- Juan Manuel Meneses Chaus: Director del CITSEM.
- Martina Eckert: Doctora del CITSEM, tutora de prácticas.

4. OBJETIVOS DE LAS PRÁCTICAS, TAREAS Y ACTIVIDADES REALIZADAS

En las prácticas realizadas, el objetivo principal se ha centrado en conseguir un reconocimiento de expresiones faciales a través de la información recogida mediante un sensor Kinect v2.

Para conseguir este objetivo principal, se han presentado una serie de objetivos a realizar:

- Estudio de reconocimiento de expresiones faciales.
- Estudio del funcionamiento de Kinect v1 y v2.
- Estudio del entorno de desarrollo SDK v1.8 y v2.0
- Estudio de las implementaciones realizadas para reconocimiento de expresiones faciales previamente con SDK v1.8
- Análisis del reconocimiento de expresiones faciales de Kinect.
- Estudio de lenguaje C#.
- Estudio de posibles implementaciones utilizando el sensor Kinect v1 y sensor Kinect v2.

En primer lugar, se procedió a realizar un estudio y revisión de los métodos que se han utilizado para poder hacer reconocimiento facial, esto se lo ha realizado gracias a la información previa generado por el grupo de investigación, la misma que se encuentra compartida en una carpeta de Moodle de la institución.

Revisando esta información lo más importante a destacar es que se puede realizar reconocimiento facial mediante el uso de AU (Action Units) los cuales se conforman de grupos de puntos detectados en la cara, los mismos que al agruparse y formar movimientos comparado con patrones previamente estudiados se puede determinar expresiones en el rostro de la persona detectada.

En la Figura 1 se pueden observar los puntos de la cara que se pueden obtener mediante un tratamiento de imágenes.

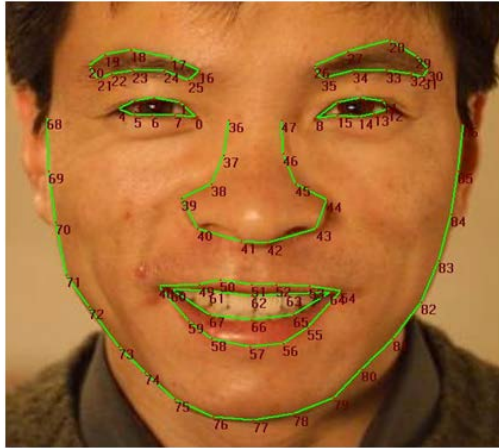


Figura 1 Puntos en el rostro para reconocimiento de emociones

Sin embargo, para un efecto práctico analizar alrededor de 90 puntos implica un procesamiento alto tomando en cuenta que los valores de los puntos varían constantemente en función del tiempo debido a que el rostro humano no es estático. En base a ello, los desarrolladores han planteado que se puede determinar las emociones con menos puntos como se ilustra en la Figura 2.

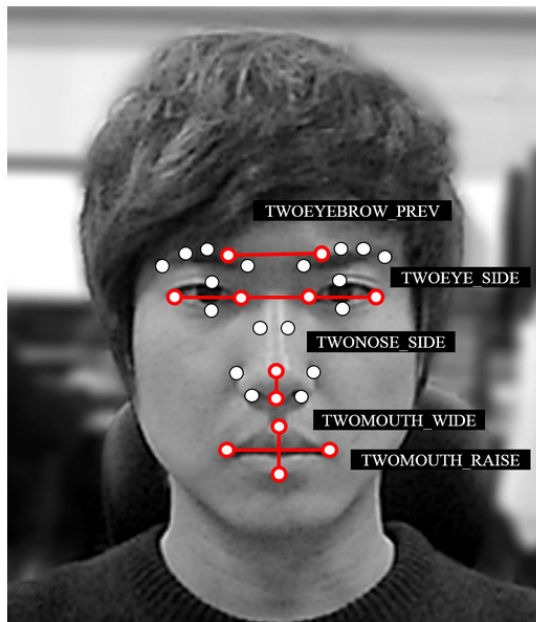


Figura 2 Puntos para reconocimiento facial en 2D

Estos puntos serán los que se intentaran obtener en Kinect para poder en un futuro procesarlos.

Adicionalmente, utilizando estos puntos se pueden detectar las emociones indicadas en Figura 3, a través de los AU definidos en Figura 4.



Figura 3 Emociones

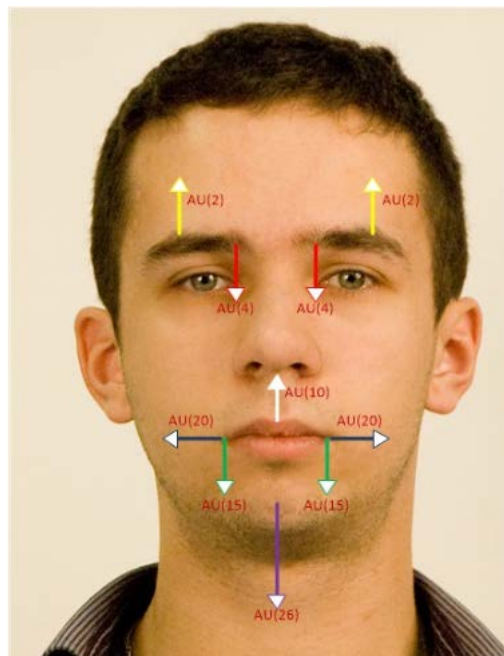


Figura 4 AU definidos para detección de emociones

- **Estudiando la SDK v 1.8**

Basado en los estudios previos realizados, el grupo de estudio había sido capaz de utilizar la Kinect v1.8 manipulando los puntos de una malla 3D, el objetivo en esta parte, es entender el uso de los códigos. Para este efecto, se revisaron los códigos previamente implementados, sin embargo, los análisis estudiados en la teoría consideran puntos tomados en un plano 2D mas no 3D, lo que hacía que ciertos criterios de evaluación arrojará resultados que varían mucho con cada emoción detectada.

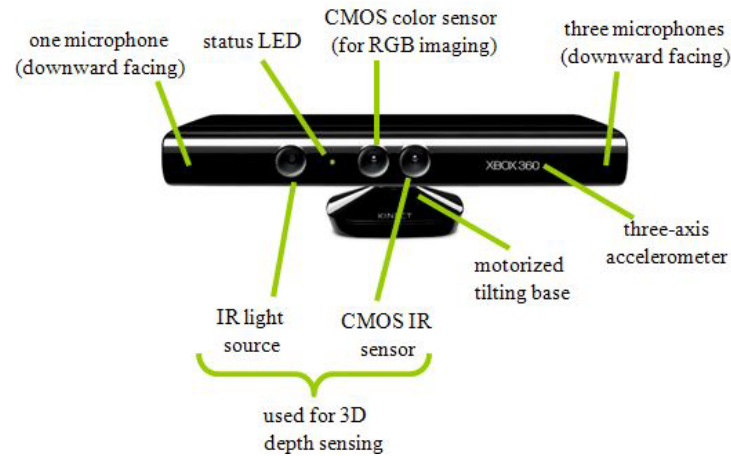


Figura 5 Kinect v1

Para solventar esta parte, se propone el uso de una malla 2D el cual arroja resultados como los que se presentan en la Figura 5, se hace referencia a un método no implementado en la Liberia 1.8, `getShapePoints` que arroja los puntos de la malla en 2D. Adicionalmente, mediante el uso de esta función se pueden determinar los puntos que nos entregan información correspondiente al momento de analizar la información y poder detectar emociones. Esto se puede observar en la Figura 6. Esta implementación se entrega para que en futuros proyectos puedan desarrollar aplicaciones con la extracción de valores de puntos en mallas 2D y obtener reconocimientos faciales más precisos gracias a una mejor detección de AU.



Figura 6 Malla 2D

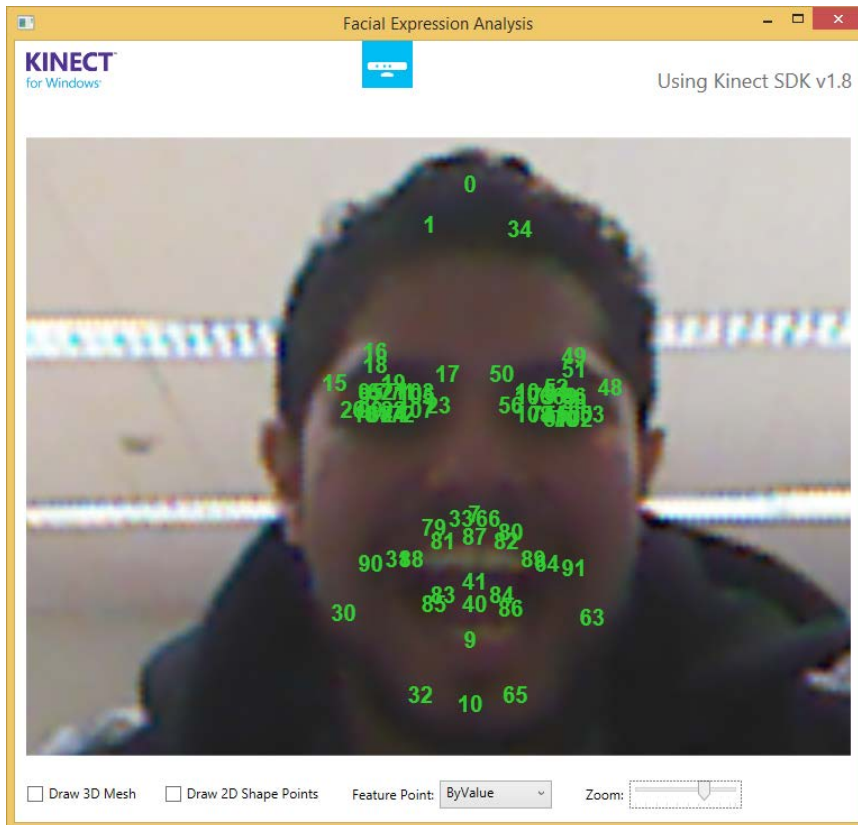


Figura 7 2D ShapePoints

- **Estudiando la SDK v2.0**

En esta versión se intenta realizar lo mismo que se planteaba con la versión 1.8, sin embargo, toda la arquitectura de lectura del sensor Kinect ha cambiado.

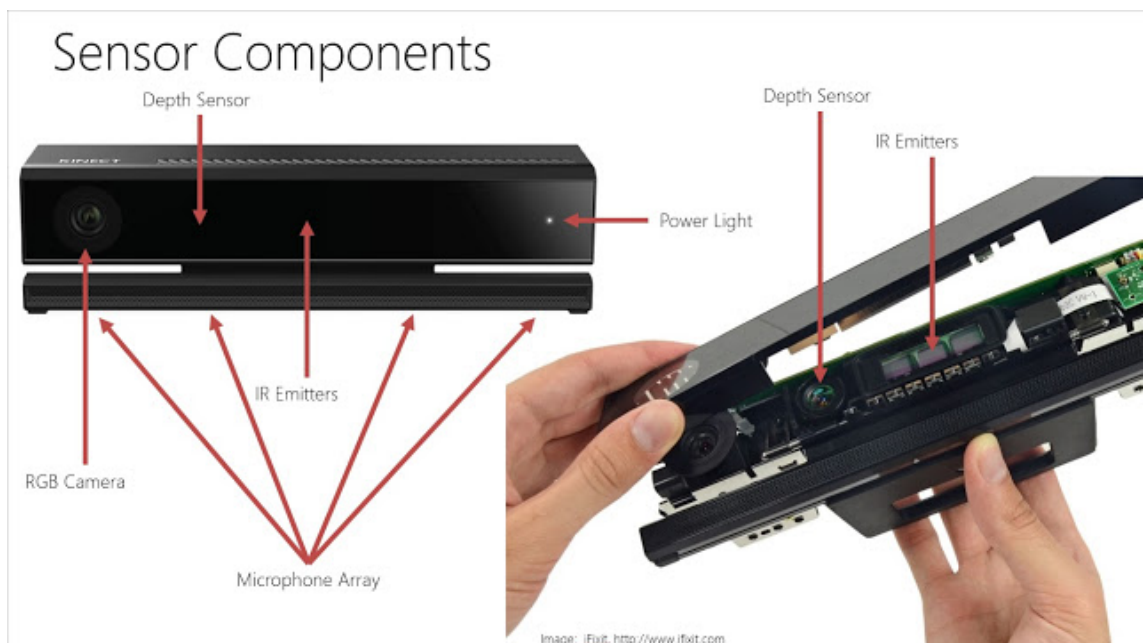


Figura 8 Kinect v2

En la versión SDK 1.8, el sensor manejaba un sensor que se podía relacionar con un evento exclusivamente. Sin embargo, en la nueva versión, se pueden tener varios lectores desde una misma fuente, es decir procesa simultáneamente muchos rostros detectados, que generan otro tipo de procesamiento de las tramas en tiempo real, lo que ha hecho que todas las librerías de la versión 2.0 sean completamente nuevas.



Figura 9 Frame processing en Kinect v1

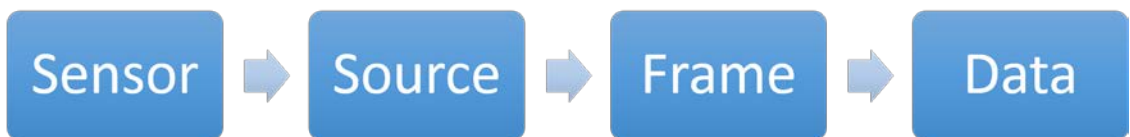


Figura 10 Frame processing en Kinect v2

Tomando en cuenta lo mencionado previamente, el paso transparente de una malla 2D a la Kinect v2 resultaría imposible, sin embargo, esta SDK provee de funcionalidades nuevas como detectar AUs. En este punto Kinect cuenta con una librería, llamada *FACE*, que puede detectar hasta 17 AUs las cuales son:

- JawOpen
- LipPucker
- JawSlideRight
- LipStretcherRight
- LipStretcherLeft
- LipCornerPullerLeft
- LipCornerPullerRight
- LipCornerDepressorLeft
- LipCornerDepressorRight
- LeftcheekPuff
- RightcheekPuff
- LefteyeClosed
- RighteyeClosed
- RighteyebrowLowerer

- LefteyebrowLowerer
- LowerlipDepressorLeft
- LowerlipDepressorRight

Los cuales nos permiten utilizar la aplicación para detectar emociones y acciones como felicidad, ojos abiertos, ojos cerrados. Los resultados los podemos observar en las Figura 11, Figura 12 y Figura 13.

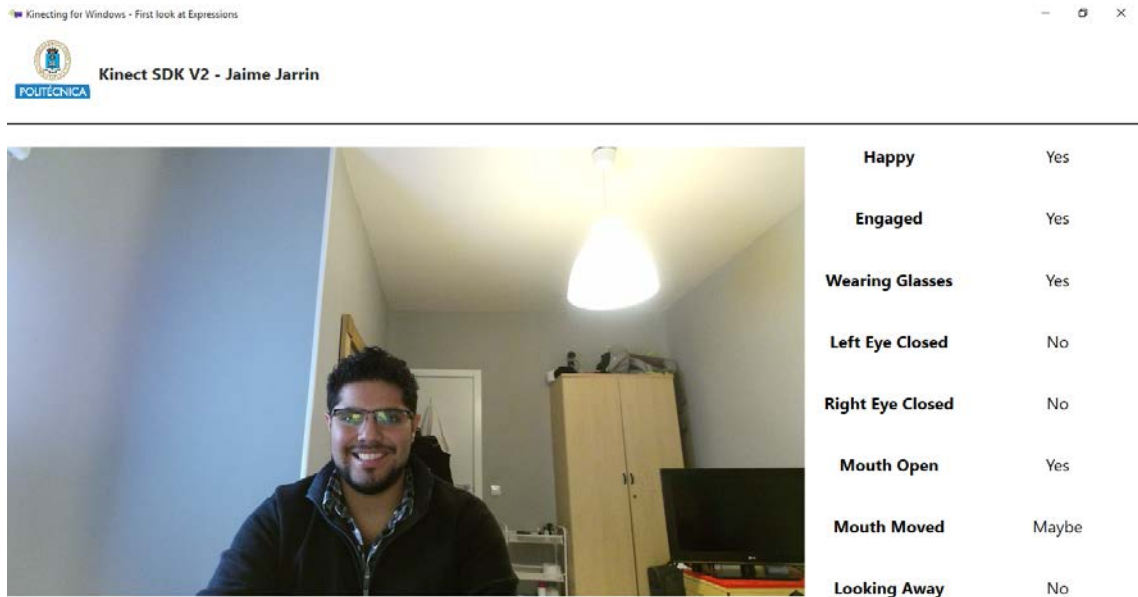


Figura 11 Feliz y con lentes

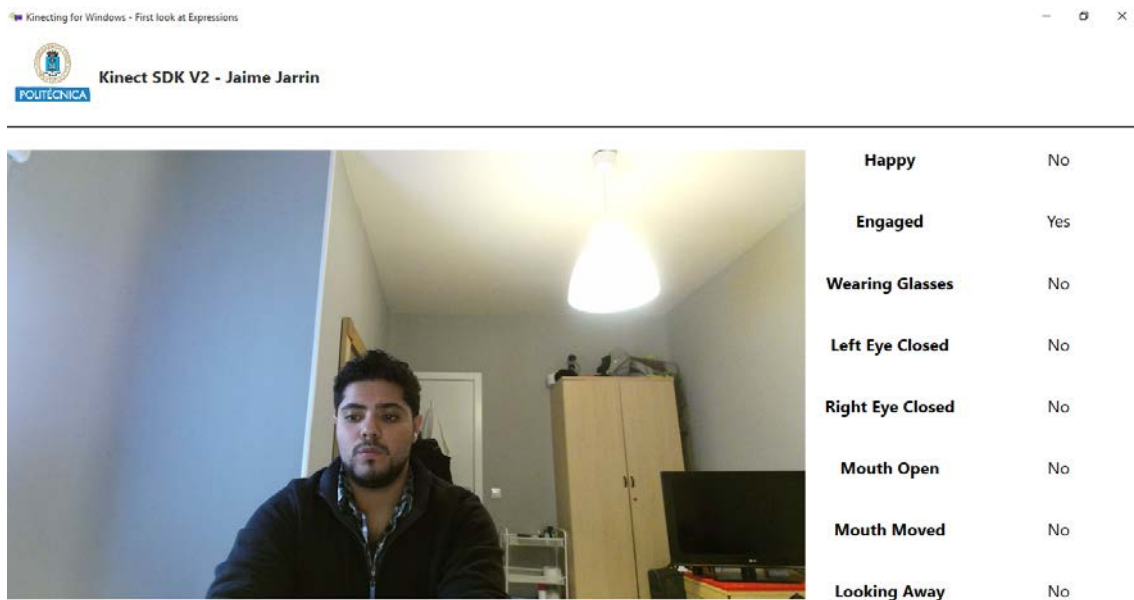
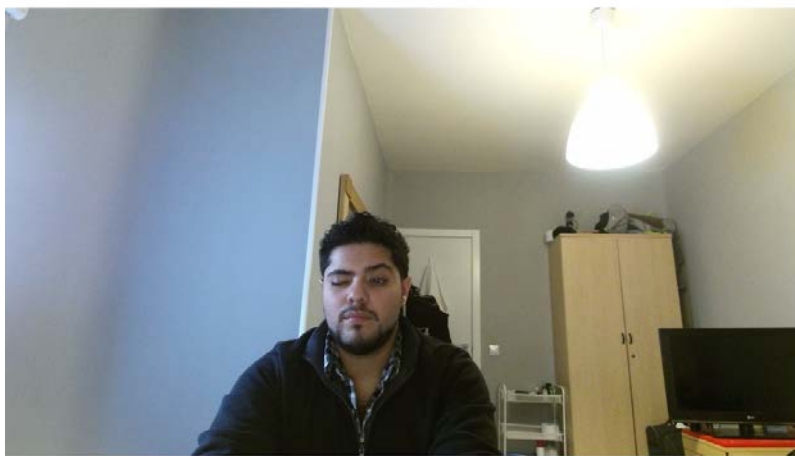


Figura 12 Cara Neutral



Happy	No
Engaged	Yes
Wearing Glasses	No
Left Eye Closed	No
Right Eye Closed	Yes
Mouth Open	Maybe
Mouth Moved	No
Looking Away	No

Figura 13 Ojo derecho cerrado

Finalmente, Kinect v2, contiene una nueva librería que procesa datos en 3D, llamada HDFace, que nos permitiría hondar aún más en la detección de AUs en torno a emociones. En la Figura 14 y Figura 15, se observa un plano 3D con un rostro representados por puntos, estos son los puntos detectado por Kinect, cada uno de ellos tiene un valor en x,y,z. Se disponen actualmente más de 1300 puntos detectados por Kinect, sin embargo, para poder procesar esta información, se deben seleccionar los puntos que son únicamente equivalentes para el proceso de detección de AUs. De esta manera se puede obtener una información más precisa de los puntos de la cara.

Los puntos a tomar en cuenta serian:

```

EyeLeft = 0
LefteyeInnercorner = 210
LefteyeOutercorner = 469
LefteyeMidtop = 241
LefteyeMidbottom = 1104
RighteyeInnercorner = 843
RighteyeOutercorner = 1117
RighteyeMidtop = 731
RighteyeMidbottom = 1090
LefteyebrowInner = 346
LefteyebrowOuter = 140
LefteyebrowCenter = 222
RighteyebrowInner = 803
RighteyebrowOuter = 758

```

RighteyebrowCenter = 849
MouthLeftcorner = 91
MouthRightcorner = 687
MouthUpperlipMidtop = 19
MouthUpperlipMidbottom = 1072
MouthLowerlipMidtop = 10
MouthLowerlipMidbottom = 8
NoseTip = 18
NoseBottom = 14
NoseBottomleft = 156
NoseBottomright = 783
NoseTop = 24
NoseTopleft = 151
NoseTopright = 772
ForeheadCenter = 28
LeftcheekCenter = 412
RightcheekCenter = 933
Leftcheekbone = 458
Rightcheekbone = 674
ChinCenter = 4
LowerjawLeftend = 1307
LowerjawRightend = 1327

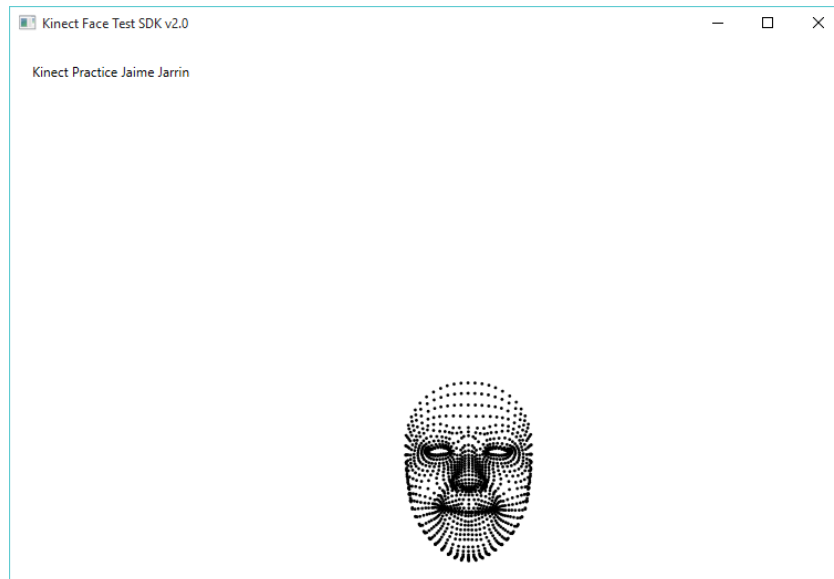


Figura 14 Feliz

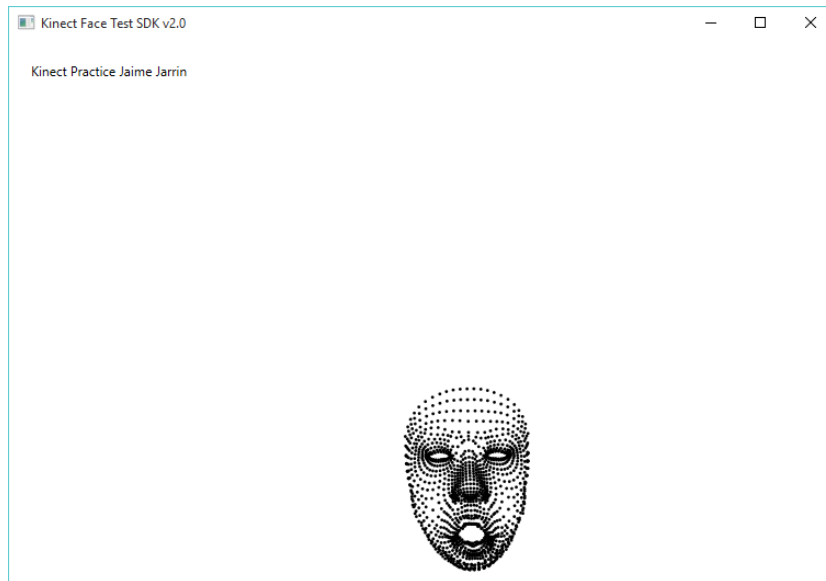


Figura 15 Boca Abierta

Al final de ambas implementaciones, se indica que el Kinect V1 tiene la capacidad de detectar un número limitado de puntos de un rostro que habilitaría a detectar emociones, sin embargo, el Kinect V2 dispone ya de librerías especializadas que entregan en un plano 2D funciones para extraer un grupo de AUs y en 3D una malla de puntos (similar a Kinect V1 en 2D) para procesar un rostro.

Después de este estudio, se plantea un trabajo futuro el cual sería agrupar los puntos determinados en Kinect V2 y generar las correspondientes agrupaciones de AU para poder detectar las emociones de una persona en tiempo real.

5. TECNOLOGÍAS Y MEDIOS TÉCNICOS UTILIZADOS

- Computador con Windows 8.1 y Windows 10.
- Entorno de desarrollo SDK v 1.8
- Entorno de desarrollo SDK v2.0
- Kinect versión 1
- Kinect versión 2
- Adaptador de Kinect versión 2 para PC
- Microsoft Visual Studio 2013/2015, para poder desarrollar el código necesario C#.
- Bases de datos científicas para obtener la información correspondiente al estado del arte de las tecnologías concretas.
- Suite de ofimática para documentar la información recopilada en esta investigación.

6. COMPETENCIAS Y HABILIDADES ADQUIRIDAS CON LAS PRÁCTICAS

Las competencias adquiridas en las prácticas realizadas son:

- Conocimiento del lenguaje de programación C# para desarrollo de aplicaciones basados en arquitecturas tipo Microsoft.
- Conocimiento de sensores de reconocimiento de video y tratamiento digital de procesado de imágenes.
- Capacidad de crear y compartir código en herramientas comunes como GitHub, LinkedIn con expertos grupos de desarrollo de nuevas tecnologías.
- Conocimientos avanzados del área de Ingeniería de Telecomunicación con capacidad para aplicarlos de una forma profesional y para innovar y desarrollar elementos y sistemas en ese área.
- Capacidad de búsqueda y selección de información, de razonamiento crítico y de elaboración y defensa de argumentos dentro del área.
- Capacidad para expresarse correctamente de forma oral y escrita y transmitir información mediante documentos y exposiciones en público.
- Capacidad de abstracción, de análisis y de síntesis y de resolución de problemas.
- Capacidad para aprender de manera autónoma nuevos conocimientos y técnicas adecuados para la concepción, el desarrollo o la explotación de sistemas y servicios de telecomunicación.
- Capacidad de utilizar la programación orientada a objetos para el desarrollo de aplicaciones que involucren procesado digital de imagen.

Las asignaturas impartidas en el programa de Master relacionadas con el trabajo que se ha llevado a cabo son:

- Procesado Señales Digitales Audio y Video

7. CONCLUSIONES

- Se ha logrado entender a rasgos iniciales el lenguaje de programación C#, el cual, al ser un lenguaje orientado a objetos está muy relacionado con instancias de Java y C++ en su estilo de programación. Por tanto, las librerías a utilizar en Kinect requieren de un manejo de objetos para cada trama (imagen) que se procese en tiempo real.
- Existen sólidos estudios previos frente al reconocimiento de emociones, sin embargo, al tratarlo de trasladar en a una implementación en tiempo real, se presentan problemas ya que los puntos detectados en el rostro no son estáticos y adicionalmente que requieren de un procesado de información que podría ralentizar la respuesta de tiempo real del sensor.
- Microsoft con el sensor Kinect v1, fue capaz de crear un dispositivo que permite a través de los sensores infrarrojo para profundidad y cámara para color, recolectar suficiente información para el tratamiento de imágenes. Sin embargo, en la primera versión dada a su calidad de cámara no se obtienen imágenes muy claras lo que a su vez no permite una detección precisa de puntos. Aun con esta limitación, es posible detectar en primer nivel puntos sobre un rostro para posteriormente detectar AUs.
- Kinect v2, al ser una versión mejorada de su predecesora dispone de sensores con mayor sensibilidad que permite un tratamiento digital de imágenes en HD con mucha más información.
- En la SDK 2.0, los sensores mejoran y las librerías de igual forma, lo que facilita en cierto modo la programación aunque limitada a las funciones desarrolladas en la librería.
- En este trabajo se investigó sobre las nuevas funcionalidades de Kinect v 2 y en base a la librería HDFACE se propone un nuevo trabajo para generar una mejor aplicación para la detección de emociones para que en posterior momento esto pueda ser mapeado aun sistema de entorno digital el cual puede ser utilizado en videojuegos.
- La labor del tutor y su función desempeñada en estas prácticas, ha sido fundamental en esta investigación. Su continuo interés y ayuda, tanto a la hora del estudio de las herramientas y conocimientos a adquirir, como a la hora de realizar las implementaciones posteriores.
- La tutora ha estado en todo momento pendiente del trabajo realizado y de los problemas que han ido surgiendo a lo largo de las prácticas. En todo momento ha prestado su ayuda, se ha mantenido pendiente de los avances y solicitando presentación de avances del proyecto en cuestión.
- Lamentablemente al ser Kinect un producto de Microsoft, la información difundida por ellas es muy poca, lo cual exige al programador a formar grupos de

investigación a través de redes sociales como GitHub o LinkedIn, lo cual, ha permitido llegar a información que ha sido hecho por desarrolladores de Microsoft para obtener explicación del uso de ciertas librerías y método.

- En general, involucrarme con una tecnología innovadora para el procesamiento digital de imágenes me ha parecido enriquecedora pues pude ser testigo de algunos conocimientos teóricos adquiridos en clase puestos en práctica en los algoritmos de procesamiento de Kinect.

8. DIARIO DE PRÁCTICAS

Semana	Actividad
15 – 22 octubre	Inicio de practicas Explicación por parte de Tutora del grupo de estudio. Estudio y aprendizaje de C# Revisión de documentación de reconocimiento de emociones.
23 – 29 octubre	Reunión con previos desarrolladores con Kinect. Análisis y entendimiento de Kinect v1
30 octubre – 5 noviembre	Desarrollo de aplicaciones con Kinect v1 Incursión en Kinect v2, instalación de SDK
6 – 12 noviembre	Presentación de avances Pruebas Kinect v2
13 – 19 noviembre	Desarrollo de código Kinect v2 en 2D
20 – 26 noviembre	Desarrollo de código Kinect v2 en 2D
26 noviembre – 2 diciembre	Desarrollo de código Kinect v2 en 2D Entendimiento de librería HDFACE para aplicaciones en 3D
3 diciembre – 12 diciembre	Generación de documentación de avances al momento.