

Benemérita Universidad Autónoma de Puebla



Facultad de Ciencias de la Computación

Tesis presentada para obtener el grado de Maestría en
Ciencias de la Computación

Título: Imitación de expresiones faciales humanas en un robot humanoide.

Presenta: Marcos Ariel Leiva Vasconcellos

Director de Tesis: Dr. David Eduardo Pinto Avendaño

Agosto de 2023

Resumen

La robótica en los últimos años ha tenido un incremento en cuanto a la creación de robots. El uso de robots en la vida cotidiana está aumentando y con ello la necesidad de que su comunicación con los humanos se realice de manera efectiva. En este caso los robots humanoides necesitan comunicarse y expresarse de manera similar a como lo hacen los humanos para ser aceptados entre estos. La comunicación a través del habla se ha incrementado con el uso de grandes modelos del lenguaje, sin embargo, la comunicación no verbal, realizada por las expresiones y gestos no está aún tan desarrollada. Varias investigaciones se han dedicado a que el robot realice las expresiones faciales tradicionales, pero existen muy pocas investigaciones encaminadas a que el robot realice una expresión facial no tradicional.

El Laboratorio de Ingeniería del Lenguaje y de Conocimiento adquirió un robot humanoide de la compañía Hanson Robotics el cual permite, gracias a sus 32 grados de libertad en el rostro y una piel artificial llamada Frubber, realizar expresiones faciales. Este robot solo trae implementado las expresiones faciales tradicionales, pero cuenta con la tecnología necesaria para implementar cualquier expresión que se desee.

Este trabajo propone una metodología para la imitación de expresiones faciales de un humano en el robot humanoide Arthur. Para esto se obtuvieron los puntos de referencia faciales utilizando la librería MediaPipe Face Mesh, la cual detecta 468 puntos en el rostro. Con estos puntos se determina el movimiento que realizan y son mapeados a los servomotores ubicados en el rostro del humanoide. Esto permite que el robot pueda realizar el movimiento observado a través de una cámara.

Para la evaluación de los resultados se realiza una evaluación subjetiva de 2 expertos los cuales evalúan del 1 al 5 cuánto se parece la expresión realizada por el humano y la realizada por el robot. Se evalúan 10 personas y se utilizan las expresiones planteadas en la rueda de Plutchik para evaluar las mismas expresiones en las personas. Se presentan los resultados evidencian una imitación satisfactoria de las expresiones faciales humanas en el robot, aunque se deben realizar mejoras en algunas expresiones donde la distancia del movimiento es pequeña.

Abstract

In recent years, robotics has experienced an increase in the creation of robots. The use of robots in everyday life is growing, along with the need for their communication with humans to be carried out effectively. In the case of humanoid robots, they need to communicate and express themselves similarly to humans to be accepted among them. Speech communication has seen an increase with the use of large language models; however, non-verbal communication through expressions and gestures is not yet as well developed. Several research efforts have been dedicated to having robots perform traditional facial expressions, but there are very few studies focused on having robots perform non-traditional facial expressions.

The Language and Knowledge Engineering Laboratory acquired a humanoid robot from Hanson Robotics, which, thanks to its 32 degrees of freedom in the face and an artificial skin called Frubber, can perform facial expressions. This robot is currently equipped with only traditional facial expressions, but it has the necessary technology to implement any desired expression.

This work proposes a methodology for mimicking human facial expressions on the humanoid robot Arthur. To achieve this, facial landmark points were obtained using the MediaPipe Face Mesh library, which detects 468 points on the face. These points determine the movements they perform and are mapped to the servomotors located on the humanoid's face. This allows the robot to replicate the observed movement captured through a camera.

To assess the results, a subjective evaluation was carried out by two experts, who rated on a scale from 1 to 5 how closely the expression performed by the human resembled the one performed by the robot. Ten individuals were evaluated, and the expressions proposed in the Plutchik's wheel were used to assess the same expressions in humans. The presented results indicate a satisfactory imitation of human facial expressions in the robot, although improvements are needed in some expressions where the movement's range is small.

Índice General

CAPÍTULO 1: INTRODUCCIÓN	1
1.1 Motivación	1
1.2 Problemática.....	4
1.3 Pregunta de investigación.....	5
1.4 Objetivos	5
1.4.1 Objetivo general	5
1.4.2 Objetivos específicos.....	5
1.5 Trabajos relacionados.....	5
1.6 Organización del documento.....	8
CAPÍTULO 2: MARCO TEÓRICO	9
2.1 Movimientos faciales	9
2.2 Robótica	9
2.2.1 Robot.....	10
2.2.2 Robot humanoide	10
2.3 Visión por computadora	11
2.4 Aprendizaje automático.....	11
2.4.1 Aprendizaje supervisado	12
2.4.2 Aprendizaje profundo.....	14
2.5 Redes Neuronales Convolucionales	14
2.6 Algoritmos para la detección de puntos de interés en el rostro	15
2.6.1 DLib-ml.....	15
2.6.2 <i>MediaPipe Facemesh</i>	17
2.7 OpenCV.....	18
2.8 ROS (Robot Operating System).....	18
2.9 Python	19
2.10 Evaluación de la propuesta.....	20
2.10.1 Rueda de emociones de Plutchik.....	20
2.10.2 Evaluación subjetiva	20
2.11 Conclusiones parciales	21
Capítulo 3: Metodología para la imitación de expresiones faciales humanas en un robot humanoide	22
3.1 Caracterización de los servomotores.....	22
3.2 Captación de imagen	24

3.3 Detección de puntos faciales	24
3.4 Cálculo de movimiento facial	25
3.4.1 Distancia interpupilar	25
3.4.2 Movimiento de los puntos de referencia	26
3.5 Movimiento de los servomotores	27
3.6 Conclusiones parciales	28
Capítulo 4. Resultados	30
4.1 Evaluación con Expertos	30
4.2 Resultado de la evaluación	30
4.3 Conclusiones parciales	31
Capítulo 5. Conclusiones	32
Referencias	33

Índice de Tablas

Figura 1. Robot humanoide creado por Da Vinci.....	1
Figura 2. Robot Kismet.....	2
Figura 3. Robot Arthur.....	3
Figura 4. Movimiento de cada servomotor en la piel.....	3
Figura 5. Distribución de los servomotores en la cabeza.....	4
Figura 6. RASA con las expresiones antes mencionadas.....	5
Figura 7. Robot ZipSa.....	6
Figura 8. Eva 2.0.....	7
Figura 9. Golem X-1.....	7
Figura 10. Músculos del rostro.....	9
Figura 11. Expresiones faciales comunes.....	9
Figura 12. Ejemplo de uso del aprendizaje supervisado.....	13
Figura 13. Ejemplo de red neuronal artificial.....	14
Figura 14. Ejemplo del aprendizaje profundo.....	14
Figura 15. Ejemplo de una Red Neuronal Convolutiva.....	15
Figura 16. Puntos de referencia facial.....	16
Figura 17. Puntos clave con DLib.....	16
Figura 18. Puntos de referencia de la librería MediaPipe Facemesh.....	17
Figura 19. ROS.....	18
Figura 20. Esquema de comunicación publicador-suscriptor en ROS.....	19
Figura 21. Rueda de emociones de Plutchik.....	20
Figura 22. Sistema para Imitar expresiones faciales en un robot humanoide.....	22
Figura 23. Cámaras integradas en Arthur.....	24
Figura 24. Puntos faciales de interés.....	24
Figura 25. Distancia interpupilar.....	25
Figura 26. Puntos para determinar el centro de la pupila.....	25
Figura 27. Distancia entre dos puntos.....	26
Figura 28. Cálculo de movimiento por puntos.....	26
Figura 29. Tabla de control RX-28.....	27
Figura 30. Archivo YAML.....	28

Índice de Tablas

Tabla 1. Aportaciones y limitaciones de las investigaciones consultadas	8
Tabla 2. Servomotores utilizados en el robot humanoide.....	22
Tabla 3. Resultados de evaluación del Experto 1	30
Tabla 4. Resultados de evaluación del Experto 2.....	30

CAPÍTULO 1: INTRODUCCIÓN

Este capítulo contiene un preámbulo general del trabajo presentado. Contiene la motivación, descripción de la problemática, plasma la pregunta de investigación, plantea el objetivo general y específicos, realiza un estudio del estado del arte y plantea la estructura que tendrá la tesis.

1.1 Motivación

En los últimos años, los avances en el campo de la robótica han dado lugar a la creación de robots cada vez más sofisticados y capaces de realizar una amplia gama de tareas. Sin embargo, uno de los aspectos clave que aún se encuentra en desarrollo es la capacidad de estos robots para comunicarse de manera efectiva con los seres humanos. La comunicación no verbal, en particular las expresiones faciales, desempeña un papel crucial en nuestras interacciones sociales, ya que nos permite transmitir y comprender emociones y estados de ánimo.

Los robots humanoides, o androides, son robots con un diseño similar a los humanos capaces de imitar las funciones del ser humano. Para que la sociedad los acepte los humanoides deben tener características humanas, así como posibilidad de comunicarse, moverse e interactuar con los objetos que lo rodean. El reto fundamental de la robótica es construir robots que tengan características humanas tanto por la piel, la silueta y las características fundamentales, por ejemplo, ojos, nariz, boca, etc. Estas características junto con el conocimiento de su entorno y la interacción con otras personas o humanoides harían al androide.

En la actualidad se ha incrementado el desarrollo de robots humanoides y a pesar de que los avances tecnológicos han permitido que estos tengan un comportamiento cada vez más parecido al de los humanos presentan algunas limitaciones sobre todo en la empatía al interactuar con los humanos.

El primer robot humanoide fue diseñado por Da Vinci en 1495, fue un autómatas humanoide y tenía un aspecto de caballero, ver Figura 1, pues tenía una armadura medieval. Este robot realizaba algunos movimientos similares a los humanos, estaba equipado con engranajes, ruedas, poleas y cables. Esto permitía que se pudiera controlar el movimiento de los hombros, manos, caderas y piernas haciendo que el robot pudiera sentarse, ponerse de pie, abrir y cerrar la boca, mover la cabeza, etc.[31]



Figura 1. Robot humanoide creado por Da Vinci

A partir del siglo 20 hubo un incremento en la creación de robots humanoides, en 1937 *Westinghouse Electric Corporation* creó “Elektro”, un androide capaz de caminar, distinguir entre distintos colores y pronunciar algunas palabras. Con la evolución tecnológica y las investigaciones en las ciencias computacionales se ha ido dotando a los robots humanoides de características “humanas”, tanto en la apariencia como en la comunicación verbal y no verbal.

Una de las principales características fundamentales para que el robot sea aceptado como humano es la posibilidad de mostrar mediante expresiones faciales la empatía en una conversación, siendo la conversación no verbal una parte importante en la comunicación interpersonal[20]. Por este motivo muchas creaciones robóticas se han enfocado en el desarrollo de las cabezas robóticas que muestren expresiones faciales o interactúen con el medio de una forma empática.

En la década del 90 el laboratorio de inteligencia artificial del MIT (Instituto Tecnológico de Massachusetts) crea Kismet, ver Figura 2. Aunque Kismet no es el primer robot expresivo se considera como uno de los influyentes en el campo de la robótica expresiva. Tenía una cara con características faciales animadas, movimiento cabeza y emisiones de sonidos que le permitían comunicarse y expresar expresiones básicas como felicidad, tristeza o sorpresa.[47]

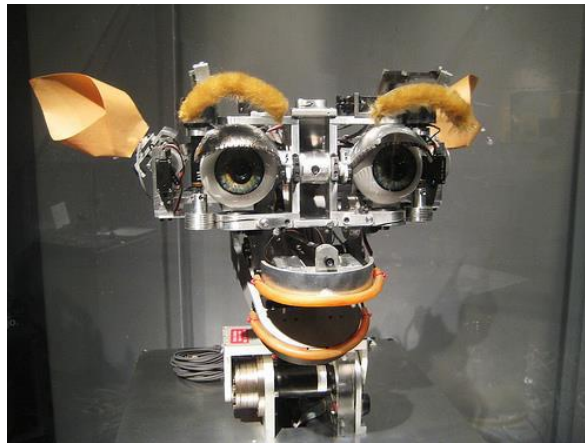


Figura 2. Robot Kismet

El laboratorio de Ingeniería del Lenguaje y el Conocimiento adquirió el robot humanoide Arthur de la compañía Hanson Robotics, este robot fue creado en 2014 y es solamente la cabeza, la cual cuenta con 32 grados de libertad activados por servomotores y cuenta con una piel artificial llamada Frubber la cual da una sensación de una piel real[13]. Los 32 grados de libertad se obtienen utilizando servomotores reversibles, al ser reversible permite simular la acción de dos grupos de músculos controlados individualmente. Este robot es capaz de realizar expresiones faciales, escuchar y hablar mediante micrófonos y bocinas, lo que permiten realizar una interacción robot-humano de manera natural, sin embargo, en la actualidad dichas expresiones faciales están predeterminadas y solo se activan cuando se presionan los botones correspondientes, limitando su interacción con el humano a la persona que lo esté operando.

En la Figura 3 se muestra al robot humanoide Arthur, el torso y extremidades son de acrílico y solo sirven para dar forma al robot, por lo que toda la investigación se basa en la cara. En la Figura 4 se observan los movimientos permitidos por cada uno de los servomotores en la piel artificial y la Figura 5 muestra la ubicación de cada uno de los motores.



Figura 3. Robot Arthur

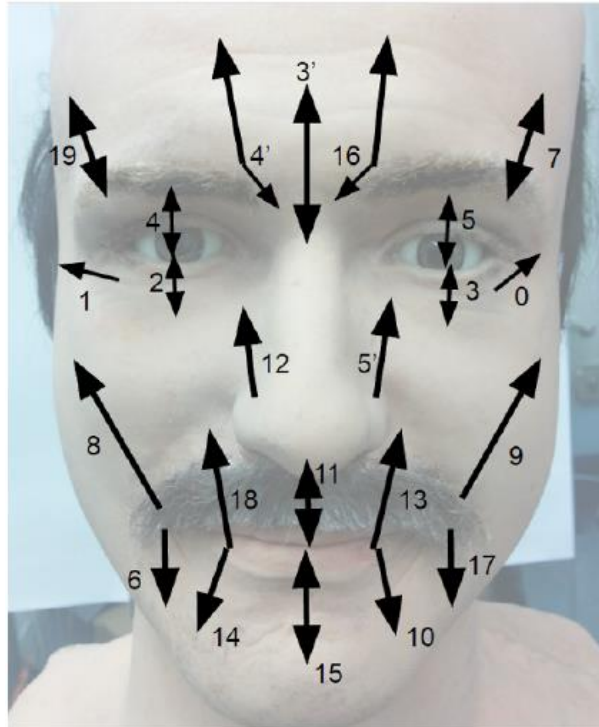


Figura 4. Movimiento de cada servomotor en la piel

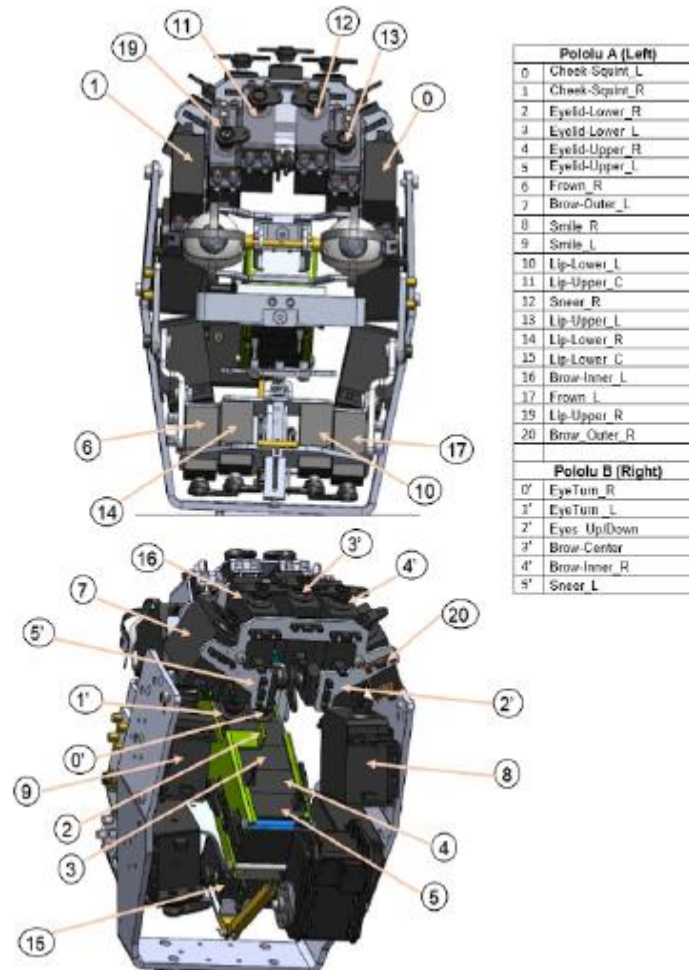


Figura 5. Distribución de los servomotores en la cabeza

Arthur tiene dos cámaras, una en el torso y una en el ojo derecho, las cuales pueden utilizarse en el procesamiento digital de video en tiempo real para interactuar directamente con la persona que esté al frente. La implementación del robot es utilizando ROS como plataforma de desarrollo, por lo que pueden implementarse nodos para realizar el procesamiento de las imágenes captada por la cámara y poder interactuar con la persona que esté frente al robot. Teniendo los grados de libertad de movimiento que presenta la cabeza del androide y la posibilidad de programación puede realizarse una interacción autónoma entre la persona y el robot.

1.2 Problemática

Actualmente el robot tiene predeterminado expresiones faciales tales como Contento, Triste, Miedo, Malvado, Sorpresa, Disgustado, Asustado; los cuales se activan cuando el operador presiona cada una de las expresiones. Teniendo la posibilidad tecnológica de realizar las expresiones faciales de manera automática se determina como problema a resolver: deficiencias en materia de reproducción mecatrónica de expresiones faciales en un robot humanoide, buscando que dichos procesos se encuentren automatizados bajo la identificación automática y en tiempo real de movimientos faciales en rostros humanos.

1.3 Pregunta de investigación

El trabajo presentado busca responder la siguiente pregunta de investigación.

¿Es posible reproducir las expresiones faciales en un robot humanoide en tiempo real, mediante una cámara realizando un mapeado de dichas expresiones con el conjunto de servomotores en el robot?

1.4 Objetivos

A continuación, se presentan el objetivo general y los objetivos específicos.

1.4.1 Objetivo general

Identificar de manera automática y en tiempo real los movimientos musculares en rostros de seres humanos y mapearlos en un robot humanoide utilizando aprendizaje automático, con la finalidad de imitar las expresiones faciales.

1.4.2 Objetivos específicos

Los objetivos específicos de este trabajo son los siguientes:

1. Caracterizar los servomotores actuales del robot humanoide Arthur para determinar el correcto funcionamiento de estos.
2. Identificar los movimientos del rostro de las personas mediante el uso de una cámara.
3. Mapear las expresiones faciales del ser humano a los servomotores del robot para que realice las expresiones captadas por el humano.

1.5 Trabajos relacionados

Los autores en [12] proponen un método para imitar las expresiones faciales utilizando redes neuronales convolucionales, el sistema se entró con la base de casos CK+ el cual es muy popular para el entrenamiento de estos sistemas. Esta base de casos consta de secuencias de videos de 123 personas con 8 expresiones faciales, en el caso del artículo sólo toman en cuenta las expresiones: Contento, Triste, Sorpresa, Enfadado y Neutral. El robot utilizado es RASA, tal como se muestra en la Figura 6.

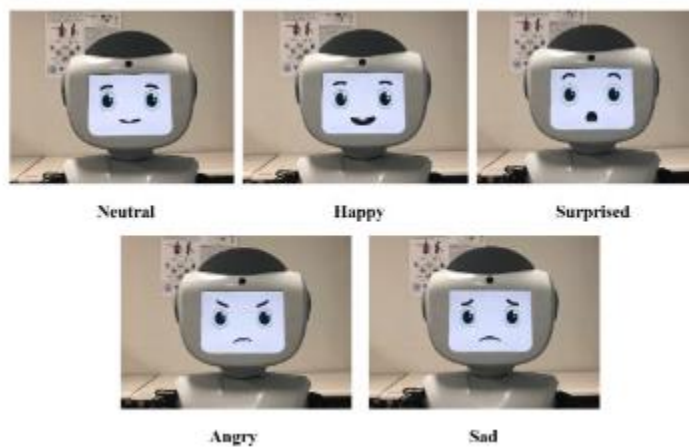


Figura 6. RASA con las expresiones antes mencionadas

Esta investigación no toma en cuenta las expresiones que pueda realizar una persona que se salga de las entrenadas con anterioridad. Además, debido a las limitaciones de la computadora interna en el robot el procesamiento no puede realizarse en tiempo real.

En [23] proponen un *framework* para mostrar las expresiones faciales utilizando redes neuronales convolucionales y FER-2013 para el entrenamiento. Este sistema solo detecta: Enfado, Contento, Sorpresa, Triste y Neutral y se implementó en el robot ZipSa, en la Figura 7 se muestra este robot.



Figura 7. Robot ZipSa

Esta investigación solo contempla 5 expresiones faciales y resulta lento en el procesamiento de dichas expresiones.

La investigación de [6] propone un sistema que aprende desde la observación visual y utilizan aprendizaje auto supervisado para su entrenamiento utilizando puntos en el rostro humano para realizar la mímica. Para la obtención de los puntos faciales utilizan OpenPose[4]. Los puntos son luego ubicados en el robot Eva 2.0 la cual se muestra en la Figura 8.



Figura 8. Eva 2.0

Los autores en [44] utilizan el robot GolemX-1, el cual es un robot minimalista tal como se muestra en la Figura 9. El robot tiene predeterminado 7 expresiones para mostrar a las personas y que estas la identifiquen.



Figura 9. Golem X-1

Luego de analizadas muchas de las investigaciones acerca de las expresiones faciales en los robots humanoides podemos resumir que en todos los casos se utilizan Redes Neuronales Convolucionales o alguna versión de ellas. Las Redes Neuronales Convolucionales (CNN por sus siglas en inglés) se definirán en el capítulo dos de la presente tesis. La mayoría de las investigaciones se centran en las emociones principales (Contento, Triste, Sorprendido, Disgustado, etc.). De los artículos consultados solo uno [6] aplica algoritmos para detectar y realizar la expresión facial en el robot, sin limitarse a las expresiones básicas. La Tabla 1 muestra un resumen de cada investigación consultada teniendo en cuenta su aportación y limitaciones.

Tabla 1. Aportaciones y limitaciones de las investigaciones consultadas

Título del artículo	Año	Aportaciones	Limitaciones
Human-Robot interaction based on Facial Expression Imitation.	2019	<ul style="list-style-type: none"> • Método para imitar expresiones faciales • Utilizan Redes Neuronales Convolucionales • Base de Casos CK+ • Aplicado al robot RASA 	<ul style="list-style-type: none"> • Detecta solo expresiones Contento, Triste, Sorprendido, Enfadado y Neutral. • El robot utilizado expresa las expresiones a través de una pantalla.
Robotics facial expression of anger in collaborative human– robot interaction	2019	<ul style="list-style-type: none"> • Método para imitar las expresiones faciales. • Utilizan Redes Nacionales Convolucionales. • Aplicado al Robot GolemX-1. 	<ul style="list-style-type: none"> • Detecta solo las expresiones Enfadado, Contento, Miedo, Triste, Sorprendido, Triste y Neutral. • El robot utilizado es minimalista por lo que su imitación se limita a puntos específicos.
Robot Facial Expression Framework for Enhancing Empathy in Human-Robot Interaction	2021	<ul style="list-style-type: none"> • <i>Framework</i> para mostrar Redes Neuronales Convolucionales. • Base de Casos FER-2013 • Aplicado al Robot ZipSa 	<ul style="list-style-type: none"> • Detecta solo las expresiones Enfadado, Contento, Sorprendido, Triste y Neutral. • El robot utilizado expresa las expresiones a través de una pantalla.
Driving Animatronic Robotic Face with Learned Models	2021	<ul style="list-style-type: none"> • Sistema que aprende desde la observación visual. • Aprendizaje auto supervisado. • Obtiene puntos faciales utilizando OpenPose. • Aplicado a Eva 2.0 	<ul style="list-style-type: none"> • OpenPose tiene buen rendimiento, pero existen mejores. • Robot propietario.

1.6 Organización del documento

El contenido del siguiente documento está organizado de la siguiente forma:

- Capítulo dos contiene el marco teórico con la definición de cada una de las herramientas utilizadas para la realización de este trabajo.
- Capítulo tres presenta la metodología presentada para la imitación de las expresiones faciales de una persona en un robot humanoide.
- Capítulo cuatro contiene los resultados de los experimentos realizados con la metodología presentada en el capítulo tres.
- Capítulo cinco presenta las conclusiones del trabajo realizado y el trabajo futuro.

CAPÍTULO 2: MARCO TEÓRICO

En este capítulo se definen conceptos fundamentales en el campo de la robótica, la visión por computadora, así como herramientas y librerías a utilizar en el desarrollo de la tesis. Así como las características anatómicas que interactúan en los movimientos faciales.

2.1 Movimientos faciales

El rostro es una de las partes más complejas en el cuerpo humano, está formado 268 músculos que permiten crear combinaciones complejas de movimientos para realizar determinadas expresiones faciales específicas[2, 27], ver Figura 10. Varios estudios se han enfocado en el entendimiento de las expresiones a través del rostro. El trabajo de Ekman y Friesen en 1978 se centró en los conceptos básicos de la expresión facial en perspectiva de agrupación muscular sistemática en unidades de acción.

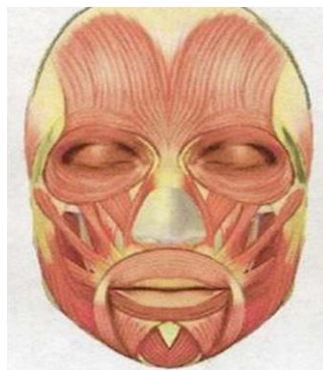


Figura 10. Músculos del rostro

Los movimientos faciales se refieren a cambios que ocurren en los músculos de la cara, los cuales resultan en diferentes expresiones faciales. Estos movimientos son una forma importante de comunicación no verbal y pueden transmitir una amplia gama de emociones, intenciones o reacciones. Las expresiones faciales más comunes son: sorpresa, felicidad, disgusto, ira, tristeza, miedo alguna de las cuales se muestran en la Figura 11[37].

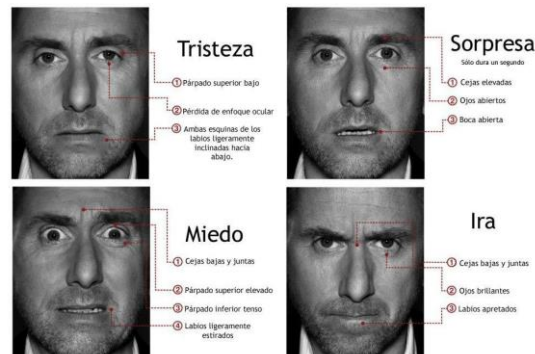


Figura 11. Expresiones faciales comunes

2.2 Robótica

La robótica es una ciencia multidisciplinaria donde interviene la mecánica, la electrónica, la ciencia de la computación y la física para construir máquinas automatizadas conocido como robots para aplicaciones industriales, científicas, domésticas y comerciales[7].

2.2.1 Robot

La palabra Robot se popularizó en los años 20 gracias a la obra teatral R.U.R. (*Rossum's Universal Robots*) creada por el checo Karel Čapek, en la obra se hacía referencia a los ayudantes como “*robot*”, que en checo significa servidumbre.

Según la RAE un robot es una máquina programable que es capaz de manipular objetos y realizar diversas operaciones[1]. Los robots pueden realizar tareas de manera autónoma o semiautónoma y pueden ser capaces de llevar a cabo acciones programadas o pueden aprender y adaptarse a su entorno utilizando algoritmos de inteligencia artificial. Estas máquinas están formadas por componentes mecánicos, eléctricos y electrónicos que permiten interactuar con su entorno y cumplir las tareas para las que se diseñó.

Los robots están compuestos principalmente por[3]:

- **Mecánica:** Es el órgano terminal, donde se ejecutarán las acciones de los demás componentes.
- **Actuadores:** Mecanismos que permiten realizar movimientos y acciones físicas. Motores, cilindros neumáticos o dispositivos similares.
- **Sensores:** Los sensores permiten recibir información del entorno para la toma de decisiones. Cámaras, sensores ultrasónicos, térmicos, etc.
- **Controladores:** Procesa la información de los sensores y toma decisiones basadas en esa información.
- **Alimentación:** Fuente de energía para el funcionamiento del robot. Batería, Electricidad, combustible, etc.
- **Programación:** Para realizar tareas específicas, los robots deben ser programados previamente. La programación define cómo el robot debe responder a diversas situaciones y llevar a cabo las acciones requeridas.

Los robots pueden clasificarse según su grado de autonomía, desde robots completamente autónomos hasta robots que requieren una supervisión humana constante. Pueden variar en cuanto a su apariencia y funcionalidad, desde robots industriales utilizados en líneas de producción hasta robots asistenciales o educativos.

2.2.2 Robot humanoide

Según [14] los robots humanoides son robots que tienen características similares a los humanos, deben poder trabajar en entornos humanos y ser capaces de interactuar con herramientas creadas para los humanos. Estos robots están diseñados para interactuar con los humanos de una manera más natural, a través de comandos de voz, gestos, expresiones faciales o respuestas emocionales simuladas.

Los androides se utilizan en diversas aplicaciones, incluyendo:

1. **Investigación y desarrollo:** Se emplean como plataformas de investigación para estudiar la biomecánica humana, la inteligencia artificial y la interacción humano-robot.
2. **Asistencia y cuidado:** Se están desarrollando robots humanoides para asistir a personas mayores o con discapacidades, brindando apoyo en actividades cotidianas y proporcionando compañía.

3. Educación: Los androides educativos son utilizadas en los diferentes niveles de educación para enseñar los conceptos impartidos de una manera intuitiva, además de ayuda a las tareas complementarias.
4. Entretenimiento: Se utilizan estos robots como actores en películas y espectáculos, o como atracciones en parques temáticos.

Los robots humanoides representan un avance en la robótica y la inteligencia artificial, aunque existen desafíos técnicos y éticos en su desarrollo y uso. A medida que esta tecnología mejore, es probable que se vea una mayor presencia en actividades cotidianas.

2.3 Visión por computadora

La visión por computadoras es un área de la inteligencia artificial y el procesamiento de imágenes que se encarga en desarrollar algoritmos para que las computadoras “vean” e interpreten el mundo al igual que lo hacen los humanos. Su objetivo principal es procesar y analizar la información visual y extraer conocimientos para la toma de decisiones [26].

Entre las aplicaciones fundamentales de los algoritmos de visión por computadoras están [26]:

- Reconocimiento de objetos: Identificar y clasificar objetos de acuerdo con su forma, tamaño, color, etc. Ej.: reconocer coche, gato o árbol.
- Reconocimiento de caracteres: Leer caracteres impresos o manuscritos.
- Reconstrucción 3D: Crear modelos tridimensionales del mundo real a partir de imágenes estereoscópicas.
- Reconocimiento facial: Identificar y reconocer rostros.
- Análisis de expresiones faciales: Determinar las expresiones realizadas por los humanos en una imagen o video.
- Detección de anomalías: Identificar patrones o comportamientos inusuales.
- Realidad aumentada: Combinar el mundo real con elementos virtuales para enriquecer la experiencia visual de los usuarios.

La visión por computadoras tiene aplicaciones en varios campos, como la industria, medicina, robótica, seguridad, conducción, entretenimiento y más. Su evolución impulsa su uso en diversos aspectos de la tecnología y la vida cotidiana.

2.4 Aprendizaje automático

El aprendizaje automático (*machine learning*) es un área de la inteligencia artificial en la cual se desarrollan algoritmos que permiten aprender de datos para aumentar su rendimiento en tareas específicas sin ser programado explícitamente. Estos algoritmos se encargan de encontrar patrones y tomar decisiones en base a ejemplos proporcionados durante su entrenamiento. El aprendizaje automático crea algoritmos de aprendizaje que construyen modelos a partir de los datos[30].

El aprendizaje automático está compuesto por varios tipos de aprendizajes, como:

- No supervisado: El modelo se entrena con datos que no están etiquetados. El objetivo es encontrar patrones ocultos agrupando datos similares o reduciendo la dimensión de los datos.
- Por refuerzo: El modelo recibe recompensas o castigos en la interacción con un entorno. El objetivo es que el modelo aprenda a tomar decisiones maximizando las recompensas.

- Supervisado: El modelo aprende con datos etiquetados. El objetivo es aprender de acuerdo con los datos de entrada.
- Semisupervisado: es una combinación del aprendizaje supervisado y no supervisado donde se utiliza un conjunto de datos parcialmente etiquetado.
- Activo: El modelo selecciona estratégicamente qué datos debe etiquetar para mejorar su rendimiento.
- Profundo: Utiliza algoritmos de redes neuronales artificiales con múltiples capas ocultas para aprender representaciones complejas de los datos.

El aprendizaje automático es una tecnología muy importante y que ha transformado muchas áreas en nuestras vidas.

2.4.1 Aprendizaje supervisado

El aprendizaje supervisado es una categoría dentro del aprendizaje automático, la característica principal de este tipo de aprendizaje es que los datos de entrenamiento están anotados, por lo que los datos de entrada contienen las salidas esperadas. Estos datos se utilizan para entrenar un modelo que pueda hacer predicciones en nuevos datos. El objetivo principal es aprender un mapeo entre un conjunto de variables de entrada y una variable de salida y aplicar ese mapeo para predecir las salidas de datos no vistos[8].

El proceso de aprendizaje supervisado implica los siguientes pasos, la Figura 12 muestra el proceso del aprendizaje supervisado:

1. Recopilación de datos: Datos relevantes para la tarea a resolver. Pueden ser estructurados (tablas) o no estructurados (textos, imágenes o audio).
2. Preprocesamiento: Se realizan transformaciones y limpieza a los datos para que sean adecuados para el aprendizaje.
3. Selección del algoritmo: La selección del algoritmo depende de la tarea y los datos, existen varios algoritmos como modelos lineales, árboles de decisión, máquina de soporte vectorial, redes neuronales, etc.
4. Entrenamiento del modelo: El entrenamiento se ajusta al conjunto de datos para aprender de los datos.
5. Evaluación: El modelo se evalúa usando datos distintos al de entrenamiento para medir el rendimiento general.
6. Ajuste y mejora: Si la evaluación no es satisfactoria se ajustan los parámetros o se eligen diferentes algoritmos para mejorar su rendimiento.
7. Predicción y uso: Cuando el modelo tiene buen rendimiento se puede utilizar para la toma de decisiones sobre nuevos datos.

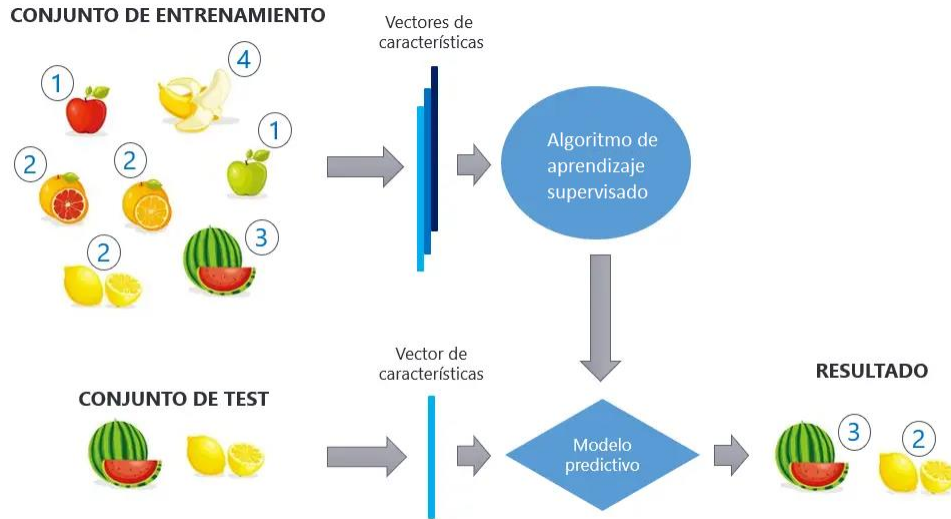


Figura 12. Ejemplo de uso del aprendizaje supervisado

El aprendizaje supervisado es uno de los enfoques más comunes del aprendizaje automático y se utilizan en la visión por computadoras, reconocimiento de voz, recomendación de productos, diagnóstico médico y otras tareas donde se utilicen los datos etiquetados. Los algoritmos más utilizados en este tipo de aprendizaje son: regresión, máquinas de soporte vectorial, K-vecinos más cercanos, árboles de decisión y redes neuronales.

2.4.1.1 Redes Neuronales Artificiales

Las redes neuronales artificiales (ANN por sus siglas en inglés) es un método del aprendizaje supervisado que se inspira en la estructura y funcionamiento del cerebro en los seres vivos, especialmente la actividad bioeléctrica de las neuronas. Estas redes están diseñadas para aprender patrones mediante el entrenamiento con datos etiquetados donde se ajustan los parámetros del modelo[28]. La Figura 13 muestra un ejemplo de una red neuronal artificial.

Las redes neuronales artificiales están compuestas por:

1. Una capa de entrada, llamada vector de entrada, que representa las variables independientes.
2. Opcionalmente una o más capas de procesamiento ocultas.
3. Conexiones ponderadas entre nodos en capas adyacentes.
4. Capa de salida que representan las variables dependientes.

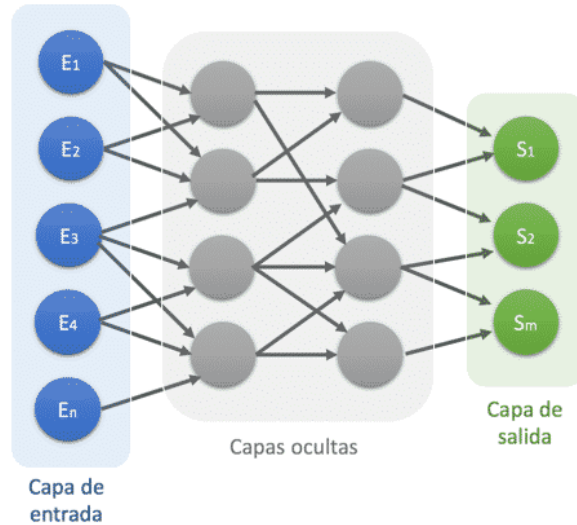


Figura 13. Ejemplo de red neuronal artificial

El entrenamiento de una red neuronal implica la actualización de los pesos y las conexiones en función a las diferencias entre las salidas de la red y la salida real en cada caso de entrenamiento. Este proceso se realiza varias veces hasta que la red pueda generalizar y clasificar bien nuevos datos.

2.4.2 Aprendizaje profundo

El aprendizaje profundo se enfoca en el uso de algoritmos de redes neuronales artificiales con múltiples capas ocultas para aprender representaciones jerárquicas. Este tipo de redes son muy poderosas para resolver problemas complejos en diversas áreas. Este tipo de redes ha demostrado ser efectivo en tareas de alta complejidad por lo que ha superado a otros métodos de aprendizaje automático en varias aplicaciones[17]. La Figura 14 muestra el funcionamiento de los métodos de aprendizaje profundo.

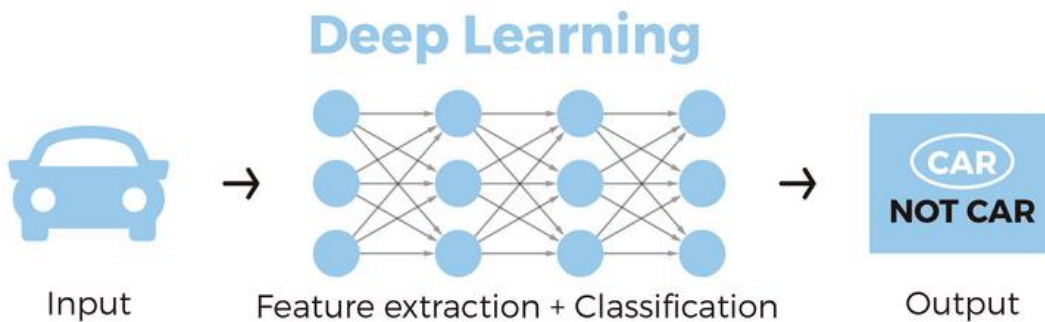


Figura 14. Ejemplo del aprendizaje profundo

2.5 Redes Neuronales Convolucionales

Las redes neuronales convolucionales (*Convolutional Neural Network CNN*) pertenecen al campo del aprendizaje profundo. Estos algoritmos son un tipo especializado de arquitectura de redes neuronales artificiales para procesar datos con estructura de cuadrículas, de allí que sea muy importante su uso para imágenes[29]. Este tipo de redes han revolucionado el campo de la visión por computadora debido a su rendimiento en la clasificación de imágenes, segmentación, detección de objetos, etc. Estas redes tienen tres tipos de capas[43], ver Figura 15:

- Capa convolucional: Es una capa fundamental en estas redes neuronales. Se aplican filtros (*kernels*) a pequeñas regiones de la imagen para extraer características específicas, como bordes, texturas, o patrones relevantes.
- Capa de agrupamiento: Esta capa reduce el tamaño espacial de las representaciones de las características, lo que disminuye la cantidad de parámetros.
- Capa totalmente conectada: Al final de la red se utilizan capas completamente conectadas para combinar las características extraídas y realizar la clasificación.

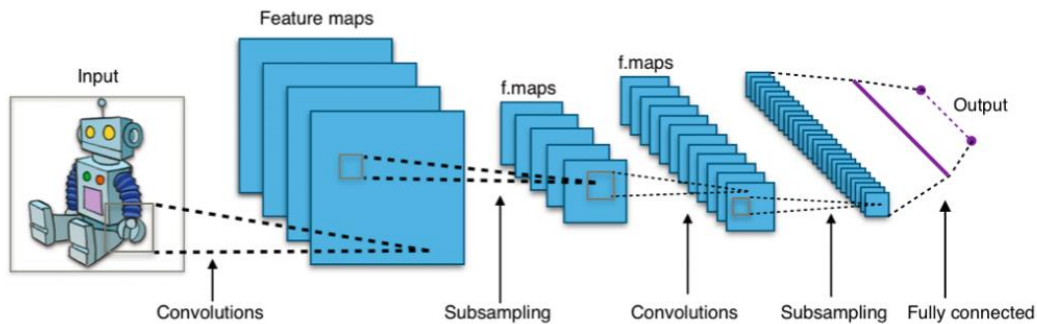


Figura 15. Ejemplo de una Red Neuronal Convolucional

Según la investigación en [18] los algoritmos con mejor rendimiento para la detección de expresiones faciales utilizan redes neuronales convolucionales para la extracción de características particulares del rostro.

2.6 Algoritmos para la detección de puntos de interés en el rostro

Para la realización de esta tesis se evaluaron dos algoritmos de aprendizaje automático para la detección en tiempo real de las expresiones faciales, estos algoritmos muestran buen rendimiento y consumen pocos recursos computacionales, lo que lo hacen ideal para su implementación en la robótica.

2.6.1 DLib-ml

DLib-ml es una biblioteca de código abierto escrita en C++, desarrollada por Davis King. Esta biblioteca está desarrollada para la investigación y para proyectos comerciales reales [16]. Está intencionada para utilizar en lenguajes como R, Python o MATLAB. Es una herramienta poderosa para detectar y localizar puntos clave en el rostro (*landmark*) en imágenes y videos. Para la detección de estos puntos utiliza el clasificador *Haar Cascade* un algoritmo que permite detectar objetos basándose en características específicas[25], también utiliza el descubrimiento de patrones de HOG[10] (*Histogram of Oriented Gradients*) y una máquina de soporte vectorial(SVM)[19] para realizar la detección de puntos clave y la alineación de rostros. Es muy popular para en la visión por computadoras debido a su precisión y eficiencia en la detección de los puntos clave.

Los puntos clave, también conocido como puntos de referencia faciales, son ubicaciones específicas en la cara, como los ojos, la nariz, los labios y las cejas, estos puntos caracterizan y describen la geometría facial[5], ver la Figura 16.

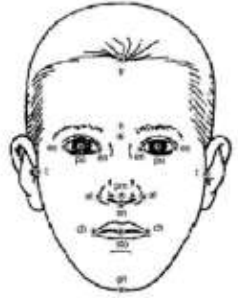


Figura 16. Puntos de referencia facial

Para la detección de los puntos clave DLib se entrenó con *iBUG 300-W face landmark dataset*[35] el cual consiste en 300 imágenes en interiores y 300 en exteriores en la naturaleza, cada imagen es anotada con 68 puntos clave utilizando una metodología semi automática[40]. La Figura 17 muestra los puntos clave utilizado en DLib.

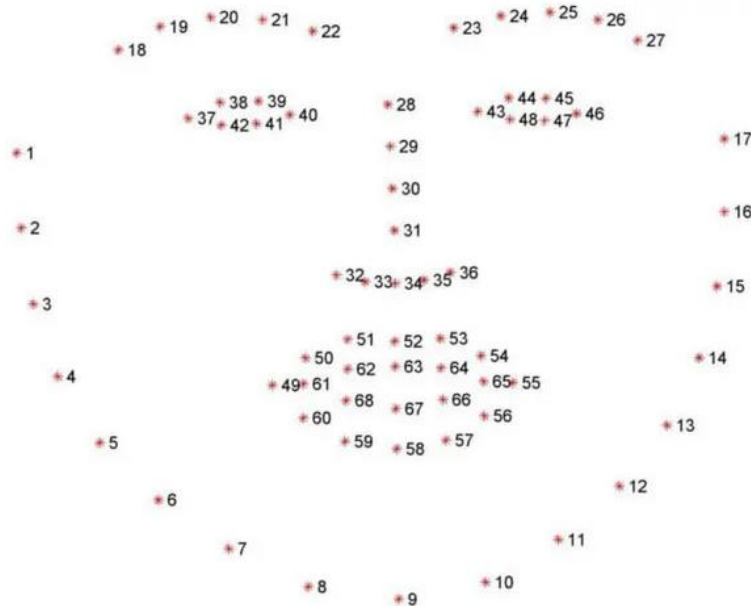


Figura 17. Puntos clave con DLib

A pesar de ser una biblioteca potente y muy utilizada, DLib tiene algunas limitaciones que impiden que pueda usarse en la detección de expresiones faciales para que el robot humanoide las imite. La principal limitación son los pocos puntos clave que utiliza para el rostro. De acuerdo con la figura 4 quedan movimientos de servomotores en el robot humanoide no cubiertos por estos puntos. Además, los puntos en el contorno de la cara en ocasiones no están bien ubicados, por lo que no pueden utilizarse para extraer la relación de aspecto del rostro. La precisión de la librería puede variar según la diversidad racial y étnica debido a los datos de entrenamiento. Otra limitación es la actualización y mantenimiento de esta, al ser de código abierto y mantenida por la comunidad las actualizaciones y mejoras pueden no estar disponibles rápidamente.

2.6.2 *MediaPipe Facemesh*

La librería *MediaPipe Facemesh* fue desarrollada por Google, de código abierto y es un modelo de visión por computadora para detectar varios puntos en rostro en tiempo real. Esta librería forma parte del *framework* MediaPipe, el cual provee una serie de librerías pre entrenadas para varias tareas de visión por computadoras, por ejemplo gestos con la mano, detección de pose del cuerpo, detección de objetos, detección de rostros, segmentación de imágenes y segmentación interactiva[38]. Esta biblioteca soporta varios lenguajes de programación, pero está optimizado para C++, Python y JavaScript, por lo que se recomienda su uso con estos lenguajes para aplicaciones de celulares, web y otras tareas de visión por computadoras[39].

Facemesh utiliza un modelo pre entrenado de Red Neuronal Convolutiva para detectar 468 puntos de referencia en el rostro, estos puntos son tridimensionales, por lo que también tiene en cuenta la profundidad entre los diferentes puntos. Incluyendo los puntos clave antes mencionados. Estos puntos pueden utilizarse para una gran variedad de aplicaciones, incluido la detección de expresiones faciales. Este modelo se entrenó utilizando 30000 fotos tomadas desde varios celulares con varias condiciones de iluminación, ambiente y sensores, estas fotos fueron tomadas de varias regiones geográficas, teniendo en cuenta todos los tonos de piel. Esto permite que este modelo permita detectar los puntos de referencia en una amplia gama de personas. La distancia media absoluta del modelo fue de 2.56%, lo que se considera una predicción de alta calidad[15]. La Figura 18 muestra los puntos de referencia de la librería *Facemesh*.

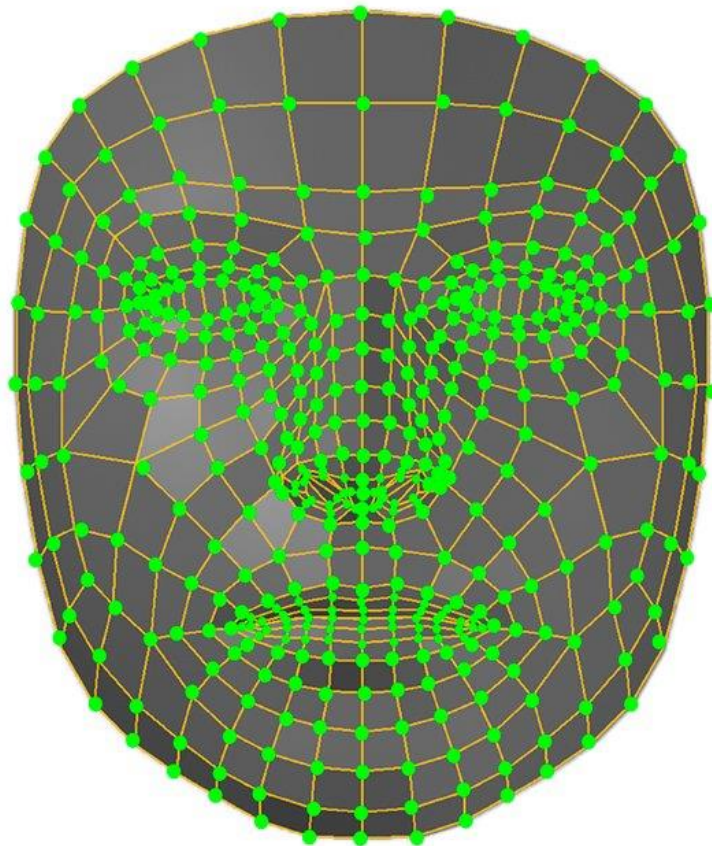


Figura 18. Puntos de referencia de la librería *MediaPipe Facemesh*

Se decidió utilizar este modelo para la detección de las expresiones del rostro por tener todos los puntos de referencia en los que se pueden mapear cada uno de los servomotores. No se tendrán en cuenta todos los puntos que nos brinda, solo se utilizarán aquellos que coincidan con las posiciones de los servomotores y otros que utilizarán como puntos para calcular la distancia desde la cámara hasta el rostro de la persona. En total se tomaron 23 puntos de referencia y 4 puntos para calcular la distancia del rostro.

2.7 OpenCV

OpenCV (*Open Computer Vision*) es una librería desarrollada por Intel en 1990. Es una librería que se utiliza para visión por computadoras y aprendizaje automático por tener algoritmos potentes relacionados con estos campos. Es la biblioteca más utilizada de visión por computadoras. Su popularidad se debe a que es libre, multiplataforma y cuenta con buena documentación. OpenCV es utilizada en varios campos, como la robótica, seguridad, realidad aumentada, edición de video e imagen entre otros. Esta librería está desarrollada en C++ pero puede utilizarse también desde Python, Java, etc.[34].

2.8 ROS (Robot Operating System)

ROS es un conjunto de librerías y herramientas que permiten construir aplicaciones para robots, es código abierto y multiplataforma, cuenta con dos versiones y cada una de ellas con diferentes distribuciones, en nuestro caso se utilizará ROS 1. En la actualidad es muy utilizado en las investigaciones, prototipado, desarrollo y prueba de robots[46], ver Figura 19[45].

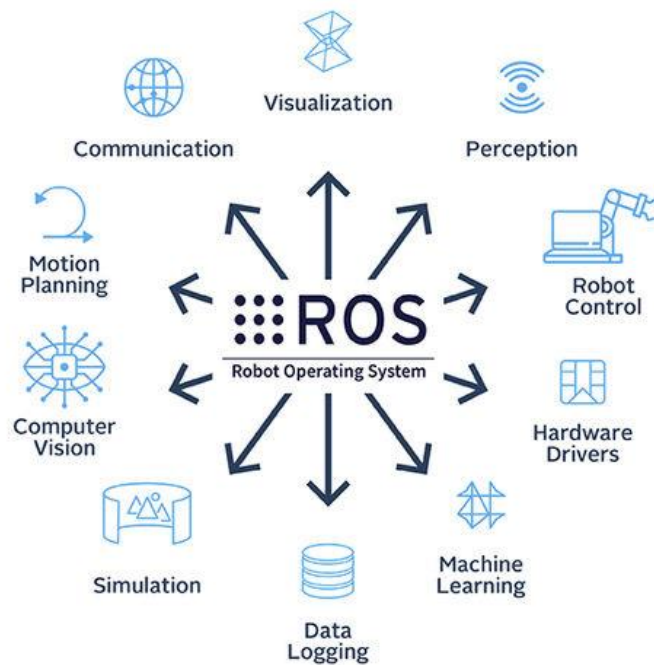


Figura 19. ROS

ROS permite la abstracción del hardware permitiendo a los desarrolladores escribir código independiente del robot. Cada proceso en ROS se denomina nodo y es donde se realiza todo el

procesamiento, entre los nodos se comunican utilizando mensajes mediante tópicos (*topics*) desde la misma máquina o desde otra de forma remota. La mensajería entre nodos se realiza mediante publicador-suscriptor, tal como se muestra en la Figura 20, en el caso de utilizar ROS 1 se necesita iniciar ROS mediante un ROS Máster, el cual se encarga de iniciar los servicios, y crear los nodos y tópicos necesarios para que exista la comunicación entre nodos. A parte de la mensajería antes mencionada ROS también permite el uso de servicios para implementar funcionalidades que no dependan de los mensajes entre nodos. La implementación de los nodos se realiza utilizando C++, Python, Java entre otros, aunque está optimizado para C++ y Python.

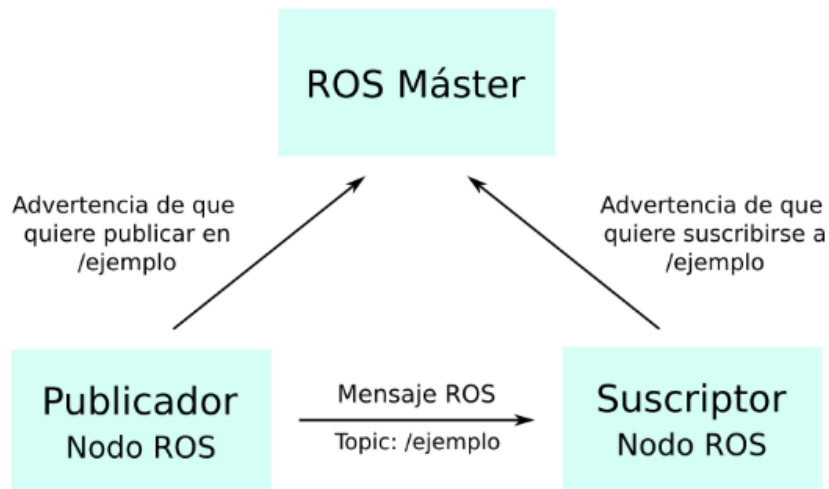


Figura 20. Esquema de comunicación publicador-suscriptor en ROS

La distribución de ROS utilizada será ROS Noetic, última versión de ROS 1 liberada en 2020. Se determinó utilizar esta versión de ROS porque los demás nodos están desarrollados en ROS 1 también y aunque existe un protocolo de comunicación entre las versiones es un poco lento, a parte algunas librerías y herramientas no están completamente implementadas en ROS 2.

2.9 Python

Python es un lenguaje de programación de alto nivel conocido por su simplicidad, fortaleza y versatilidad, de código abierto. Fue creado por Guido van Rossum y liberado para su uso en 1991. Es un lenguaje orientado a objetos, interpretado, de tipado dinámico y multiplataforma[49]. Se caracteriza por el uso de una sintaxis concisa y es obligatorio la tabulación. Es un lenguaje adecuado para programadores noveles y expertos. Es considerado el más popular entre los lenguajes de programación actuales[48].

Python es utilizado para una gran variedad de aplicaciones, incluyendo desarrollo web, computación científica, análisis de datos, visualización, aprendizaje automático, visión por computadoras, entre otros.

Se decidió utilizar Python (versión 3.8) como lenguaje de programación para esta investigación debido a que el sistema utiliza algoritmos de aprendizaje automático para detectar los puntos de referencia en el rostro y pueden utilizarse librerías de Python en este sentido, la comunicación con los servomotores se puede realizar desde este lenguaje de programación, la librería OpenCV para la

adquisición del video está implementada para utilizar desde Python y se utilizará ROS con Python para la implementación en el robot.

2.10 Evaluación de la propuesta

A continuación, se describe la herramienta que se utilizará para las expresiones faciales y la forma de evaluar la propuesta en esta tesis.

2.10.1 Rueda de emociones de Plutchik

Para la evaluación se utilizará la rueda de emociones de Plutchik, ver Figura 21[36], esta es un herramienta gráfica que representa las diferentes emociones humanas, organizadas en círculo o rueda. Fue desarrollada por Robert Plutchik en los 80. La rueda de Plutchik consta de 8 emociones básicas que se sitúan en los puntos cardinales del círculo: alegría, confianza, miedo, sorpresa, tristeza, repugnancia, ira y anticipación[24].

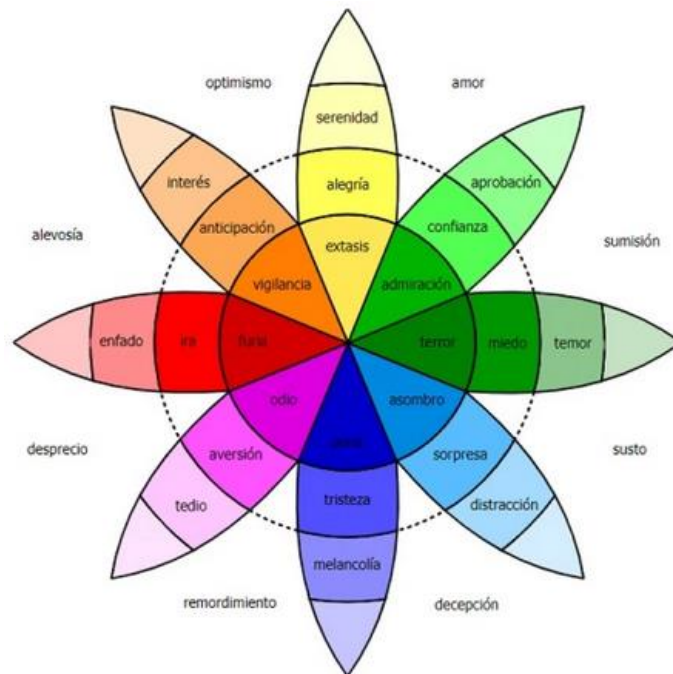


Figura 21. Rueda de emociones de Plutchik

La evaluación se realizará sobre las 8 emociones básicas y cada persona deberá realizar la expresión facial de cada una de estas expresiones.

2.10.2 Evaluación subjetiva

La evaluación de las emociones mediante la expresión facial no es preciso, pero en este caso se evaluará si la expresión realizada por el robot es semejante al del robot. Para esto se realizará una evaluación subjetiva realizada por dos expertos. La evaluación subjetiva es una evaluación influenciada por la opinión y percepción personal. Esta evaluación no se basa en mediciones objetivas, pero puede ser útil en algunos contextos[9]. En este caso se toma este tipo de evaluación porque no hay forma efectiva de medir una expresión facial ya que en ella participan de varios músculos y cada persona puede realizar expresiones faciales similares, pero no con los mismos

movimientos. La evaluación será cuantitativa y será del uno al diez siendo el uno el menos parecido entre las expresiones del robot y del humano y diez el máximo de semejanza en la expresión.

2.11 Conclusiones parciales

En este capítulo se definieron los conceptos y se definieron las herramientas a utilizar en el desarrollo de esta tesis. Se definió para la obtención de la imagen utilizar OpenCV. Para la detección del movimiento muscular en rostro se decidió utilizar *MediaPipe Facemesh* ya que ofrece mejor rendimiento en la detección en los puntos de referencia faciales. Como lenguaje de programación se decidió utilizar Python pues es un lenguaje robusto y que cuenta con las librerías necesarias para la interacción humano-máquina. Se utilizará ROS para la implementación del método en el robot y la evaluación se realizará mediante expertos utilizando la evaluación subjetiva, teniendo en cuenta la semejanza de la expresión creada por el robot en comparación con la que realizó el humano. Para que las expresiones faciales sean uniformes en cada persona se utilizará la rueda de Plutchik como base a las expresiones que realizará cada persona evaluada.

Capítulo 3: Metodología para la imitación de expresiones faciales humanas en un robot humanoide

En este capítulo se establecen los pasos utilizados para la imitación de expresiones faciales en un robot humanoide. Esta propuesta consta de 4 etapas: Caracterización de los servomotores, Captación de imagen mediante webcam, Mapeado de los movimientos adquiridos con los servomotores en el robot e Imitación de las expresiones faciales. La Figura 22 muestra un esquema general del funcionamiento de la metodología propuesta, obviando la integración en ROS. Cada una de las fases de desarrollo corresponde a una sección del capítulo.

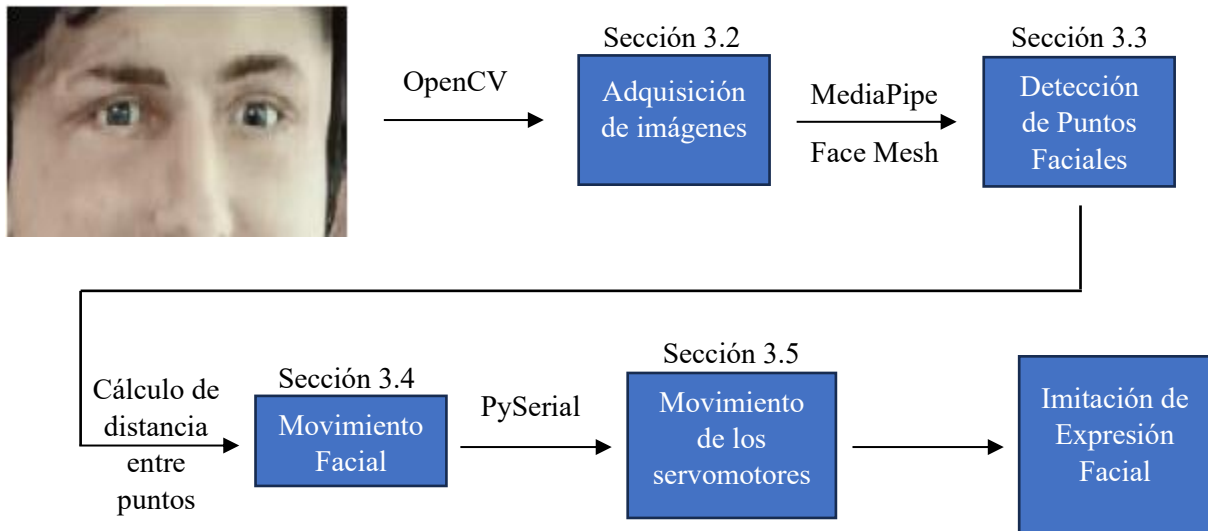



Figura 22. Sistema para Imitar expresiones faciales en un robot humanoide

3.1 Caracterización de los servomotores

En la Tabla 2 se muestran los servomotores utilizados de acuerdo al manual de usuario proporcionado con el robot[13], así como la cantidad usada de cada uno y una imagen ilustrativa. La distribución de los servomotores en el robot se muestra en la Figura 5. Los servomotores Pololu están agrupados utilizando dos Pololu Maestro Mini de 24 pines[41] una de las tarjetas tiene conectado 20 de estos servomotores y se encargan de los movimientos de la cara y la otra tarjeta tienen conectado 6 servomotores que se encargan de los movimientos de los ojos.

Tabla 2. Servomotores utilizados en el robot humanoide

Cantidad	Marca	Modelo	Imagen
26	Pololu	XPert RC Brushless CI-2401[51]	

1	Dynamixel	MX-28[32]	
1	Dynamixel	RX-28[22]	
2	Dynamixel	RX-64[33]	

Se les realizaron pruebas a todos los servomotores para revisar su funcionamiento utilizando los movimientos mostrados en la Figura 4. Después de estas pruebas se determinó que:

1. Todos los servomotores Dynamixel funcionaban de forma correcta.
2. 13 servomotores Pololu funcionaban de forma correcta.
3. 6 servomotores Pololu no funcionaban.
4. 7 servomotores Pololu funcionaban de forma parcial (realizaban movimiento, pero no de acuerdo con la figura 4 o estaba limitado el mismo).

Se realizaron intervenciones en las dos tarjetas Pololu Mini Maestro de 24 pines[41], se rectificaron todos los contactos de los servomotores para mejorar la conectividad eléctrica en la tarjeta. Luego de esta intervención se logró mejorar el estado de los servomotores afectados quedando de la siguiente manera la funcionalidad de los servomotores:

1. 22 servomotores funcionaron de manera correcta.
2. 3 servomotores no funcionan.
3. 1 servomotor funciona de manera parcial

El estado de los servomotores que no funcionaron o funcionaron de manera parcial se debe a la desconexión de los hilos atados a la piel artificial, para corregir estas dificultades sería necesario desconectar todos los servomotores y retirar la piel que recubre la cara, para realizar esto se tendría que conocer la distribución de cada hilo en la piel algo que tardaría mucho tiempo y no es objetivo de esta tesis.

3.2 Captación de imagen

Arthur tiene integrado dos cámaras, ver Figura 23, una en el espacio del pecho y otra en el ojo izquierdo. Estas cámaras son genéricas y en el manual de usuario no se especifican sus características. Para la adquisición de la imagen se utilizará la cámara que tiene en el ojo izquierdo, ya que la persona podría realizar los movimientos faciales y poder observarlos mientras los va haciendo.



Figura 23. Cámaras integradas en Arthur

Una de las tareas fundamentales para imitar las expresiones faciales es la adquisición de la imagen, para esto se utiliza OpenCV, el cual tal y como se describió en la sección 2.7 OpenCV es una librería muy común para la adquisición y procesamiento de imágenes. En este caso se utiliza OpenCV en su versión 4.6.0 y la implementación se realizará en Python versión 3.8. Se obtendrán 20 cuadros por segundo(fps) para su procesamiento, lo cual es adecuado para el procesamiento de videos estándar tal y como es el procesamiento de imágenes médicas, además es lo máximo que se puede obtener con la cámara con la que se obtendrán las imágenes por la calidad de esta. La cámara tiene una resolución FullHD(1080p) ideal para el procesamiento rápido y con calidad de los movimientos faciales.

3.3 Detección de puntos faciales

Los puntos faciales se detectarán utilizando MediaPipe Face Mesh, tal y como se mencionó en 2.6.2 *MediaPipe Facemesh*. Luego de obtenida cada imagen mediante la cámara se utilizará esto para detectar los puntos. De los 468, solo se tomarán los puntos: 9, 107, 336, 293, 63, 255, 25, 186, 410, 43, 273, 15, 11, 303, 73, 404, 180, 374, 159, 386, 145, 371, 142, ver Figura 24. Estos puntos de referencia faciales son los mismos donde están localizados los servomotores del robot humanoide, permitiendo tener en cada uno de ellos un punto asociado y permitir el movimiento respectivo.



Figura 24. Puntos faciales de interés

Se utilizan los puntos extremos del lateral izquierdo y lateral derecho de los ojos para luego calcular la relación de aspecto entre los ojos para conocer la distancia entre la cámara y la persona. Esto se realiza utilizando la distancia interpupilar, la cual se definirá en la siguiente sección. En este caso los puntos que se utilizan son 33, 133, 362, 263.

3.4 Cálculo de movimiento facial

La detección de movimiento de cada punto es una tarea fundamental para luego el robot pueda realizar el mismo movimiento mediante los servomotores por lo tanto se deben tener en cuenta varios factores para un correcto cálculo de movimiento de cada punto de referencia en el rostro.

3.4.1 Distancia interpupilar

La distancia interpupilar(ver Figura 25) es un método simple y efectivo para tener una medida estándar de la persona y la cual puede utilizarse para medir la distancia entre la persona y la cámara[11]. Esta distancia se mide entre el centro de las pupilas del ojo izquierdo y derecho. Para calcular esta distancia se utilizarán los puntos de referencia 33, 133, 362, 263 los cuales coinciden con los extremos laterales de cada ojo tal y como se muestra en la Figura 26, a partir de estos puntos se determina el centro de la pupila de cada ojo. Se utilizan estos puntos para determinar el centro de la pupila para evitar errores en caso de que la persona cierre los ojos.

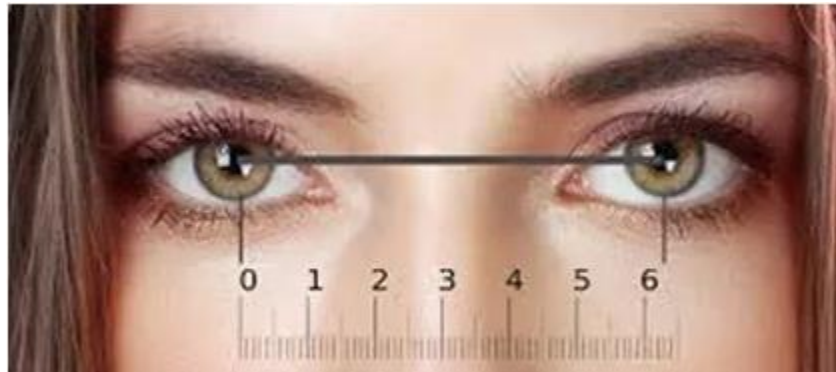


Figura 25. Distancia interpupilar



Figura 26. Puntos para determinar el centro de la pupila

La distancia se calculará utilizando la distancia euclidiana, la cual es la distancia entre dos puntos en el espacio euclídeo, se deriva del teorema de Pitágoras (ver (1)). Se utilizó esta distancia porque es una medida simple y muy utilizada en aplicaciones de visión por computadora. El resultado de esta

medida se tomará en cuenta para conocer si la persona está cerca o lejos de la cámara y así tenerlo en cuenta para detectar cuánto se realizó de movimiento.

$$d_E(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

3.4.2 Movimiento de los puntos de referencia

Para el cálculo del movimiento de los puntos de referencia se tiene en cuenta el centro de las pupilas derecha e izquierda. Se localiza el centro de ambas pupilas hallando el punto medio entre los extremos derecho e izquierdo de cada ojo tal y como se explicó en la sección anterior.

El movimiento de cada punto de referencia facial se detecta y se toman el punto inicial y final. Este desplazamiento se calcula utilizando la distancia euclidiana (ver (1)). La Figura 27 muestra el cálculo de dichas distancias entre puntos.

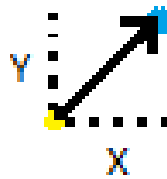


Figura 27. Distancia entre dos puntos

La Figura 28 muestra el proceso de obtención de las distancias. Inicialmente se calcula una distancia base desde el centro de la pupila hasta todos los puntos de su lado. Esta distancia base es la distancia que se toma en cuenta para determinar si el movimiento ha sido hacia arriba o hacia abajo dado que la distancia calculada es un valor positivo. Al realizar algún movimiento facial la persona se obtiene una nueva coordenada de el mismo punto y se calcula su nueva distancia, esta nueva distancia se compara con la distancia base para determinar su orientación (abajo, arriba, izquierda o derecha). La nueva distancia será la nueva comparación para determinar el nuevo movimiento del punto, teniendo en cuenta también la distancia base. Se tiene en cuenta un umbral para determinar el movimiento y así evitar micro movimientos que no se tienen en cuenta debido a la precisión del modelo.

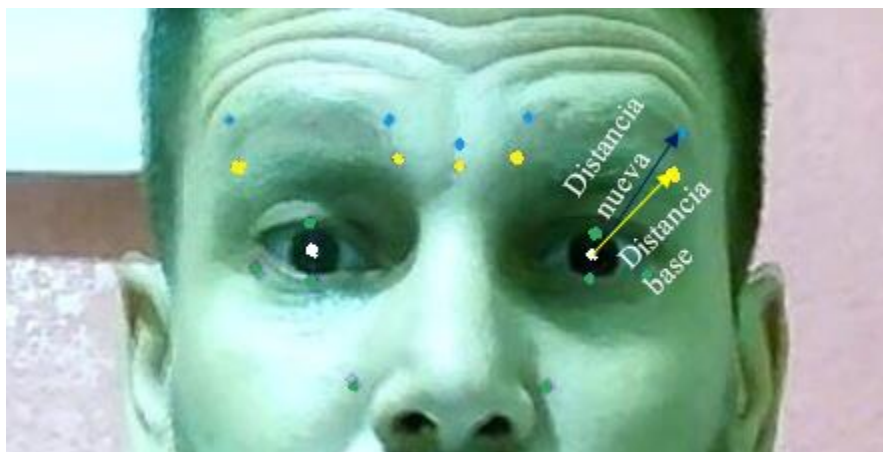


Figura 28. Cálculo de movimiento por puntos

La Figura 28 muestra un ejemplo del cálculo de estos movimientos. La persona mueve las cejas y se realiza el cálculo de distancia según el procedimiento descrito. Se tendrían dos distancias y con eso se calcula el desplazamiento que realiza cada punto al realizar una expresión facial.

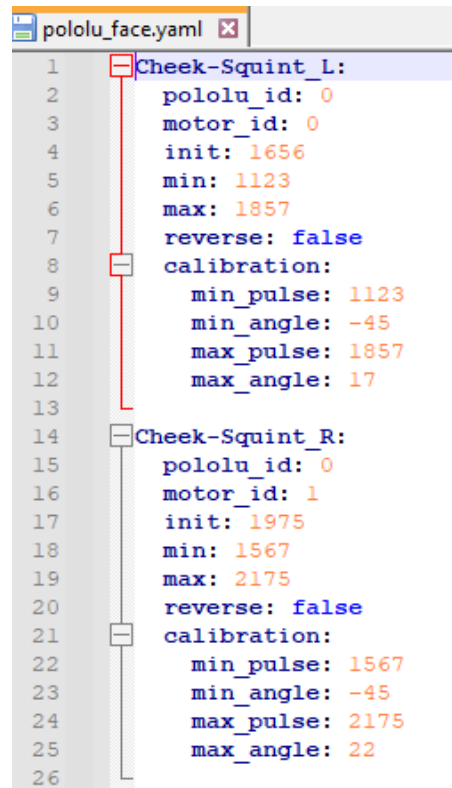
3.5 Movimiento de los servomotores

Para la comunicación con los servomotores se utiliza la librería *Maestro Servo Controller*[50] para Python con la cual se puede realizar el control de los servomotores de la marca Pololu, esta librería es una implementación de la librería PySerial[42] Python que proporciona una interfaz para la comunicación con dispositivos electrónicos que se conectan a través del puerto serie. Esta librería permite al desarrollador interactuar con estos dispositivos mediante la creación y manejo de datos. Este puerto es muy utilizado en la interacción con los servomotores. Para el control de los servomotores Dynamixel se utiliza Dynamixel SDK[21], el cual provee funciones de control para el control de estos tipos de servomotores. Para el control de los Dynamixel se realizó mediante la tabla de control las especificaciones proporcionadas por el fabricante. La Figura 29 muestra la tabla de control para el Dynamixel RX-28, el cual es el responsable de mover la mandíbula inferior.

Address	Size(Byte)	Data Name	Description	Access	Initial Value
24	1	Torque Enable	Motor Torque On/Off	RW	0
25	1	LED	Status LED On/Off	RW	0
26	1	CW Compliance Margin	CW Compliance Margin	RW	1
27	1	CCW Compliance Margin	CCW Compliance Margin	RW	1
28	1	CW Compliance Slope	CW Compliance Slope	RW	32
29	1	CCW Compliance Slope	CCW Compliance Slope	RW	32
30	2	Goal Position	Target Position	RW	-
32	2	Moving Speed	Moving Speed	RW	-
34	2	Torque Limit	Torque Limit	RW	Max Torque
36	2	Present Position	Present Position	R	-
38	2	Present Speed	Present Speed	R	-
40	2	Present Load	Present Load	R	-
42	1	Present Voltage	Present Voltage	R	-
43	1	Present Temperature	Present Temperature	R	-
44	1	Registered	If Instruction is registered	R	0

Figura 29. Tabla de control RX-28

Cada servomotor tiene 3 parámetros fundamentales que se tendrán en cuenta, estos son los valores máximos, mínimo e inicial para el correcto movimiento en el rostro del humanoide. Estos parámetros se tomaron de archivos “yaml” incluidos en el código del robot. La Figura 30 muestra un archivo “yaml” donde vienen estos parámetros. En este caso por cada servomotor, nombrado en el archivo por donde está ubicado, se tiene el identificador de motor, identificador de la tarjeta Pololu Maestro en la que se encuentra y los valores antes mencionados, también se tienen los ángulos y pulsos máximo y mínimo, pero esos no se tendrán en cuenta en este caso.



```
1 Cheek-Squint_L:
2   pololu_id: 0
3   motor_id: 0
4   init: 1656
5   min: 1123
6   max: 1857
7   reverse: false
8   calibration:
9     min_pulse: 1123
10    min_angle: -45
11    max_pulse: 1857
12    max_angle: 17
13
14 Cheek-Squint_R:
15   pololu_id: 0
16   motor_id: 1
17   init: 1975
18   min: 1567
19   max: 2175
20   reverse: false
21   calibration:
22     min_pulse: 1567
23     min_angle: -45
24     max_pulse: 2175
25     max_angle: 22
26
```

Figura 30. Archivo YAML

Una vez configurado los servomotores se procede a mover el servomotor de acuerdo con la distancia obtenida, tal y como se describe en la sección anterior. Debido a que el movimiento facial de cada persona puede variar en cuanto a distancia entre la posición inicial y final se tomará una distancia máxima para cada punto de referencia facial. Esta distancia se calculará utilizando una media entre varias personas.

3.6 Conclusiones parciales

En este capítulo se propuso una metodología para imitar el movimiento facial de un humano en un robot humanoide. Para realizar este movimiento se tomaron en cuenta los servomotores que tiene el robot. Se analizó también la librería para detectar los puntos de referencia facial de acuerdo con la localización de los servomotores seleccionándose MediaPipe Face Mesh por ser una librería robusta, rápida y con los puntos necesarios para que coincidan con los servomotores. Se calcula la distancia interpupilar para obtener la relación del rostro y así determinar la distancia entre el rostro y la cámara, además con cada punto de referencia facial se establece la distancia base que se calcula con la distancia euclidiana entre el centro de la pupila y cada punto. Una vez detectado cada movimiento se

mapea ese movimiento en el servomotor correspondiente permitiendo que se logre una expresión en el robot.

Capítulo 4. Resultados

En este capítulo se muestran los resultados de los experimentos realizados. Los experimentos son utilizando las expresiones propuestas en la rueda de Plutchik (ver sección 2.10.1 Rueda de emociones de Plutchik). La evaluación será realizada por dos expertos, de forma cuantitativa del 1 al 5.

4.1 Evaluación con Expertos

Para la evaluación se utilizan dos expertos, los cuales para estandarizar los resultados seleccionan un conjunto de expresiones faciales de la rueda de Plutchik, estas expresiones son: alegría, miedo, ira, tristeza, sorpresa y repugnancia. Se toman estas expresiones faciales porque tienen movimientos musculares típicos y pueden diferenciarse entre sí, si bien no todas las personas realizan los mismos movimientos para realizar cada una de estas expresiones la función de los expertos será determinar del 1 al 5 cuánto se parece la expresión del humano con la del robot.

La evaluación será cuantitativa y es desde 1, que representa que la expresión difiere mucho a la del humano, hasta 5, siendo una expresión similar a la que realiza la persona. Se tendrá en cuenta para la evaluación la expresión final para evitar que se evalúen partes de expresiones.

4.2 Resultado de la evaluación

Para el experimento se tomaron 10 personas al azar y cada uno de ellos realizaron las expresiones antes mencionadas y ellos expertos evalúan de acuerdo con los criterios antes mencionados. La Resultados de evaluación del Experto 1 Tabla 3 muestra la evaluación del Experto 1 y la Tabla 4 muestra la evaluación del experto 2 para cada individuo.

Tabla 3. Resultados de evaluación del Experto 1

No.	Alegría	Miedo	Ira	Tristeza	Sorpresa	Repugnancia
1	4	3	3	4	5	4
2	5	4	3	4	5	3
3	5	3	3	3	4	4
4	4	3	4	3	5	3
5	4	4	3	4	4	3
6	4	5	3	3	4	4
7	4	4	4	4	5	4
8	4	4	3	4	4	3
9	4	4	3	3	5	4
10	4	3	3	4	5	4

Tabla 4. Resultados de evaluación del Experto 2

No.	Alegría	Miedo	Ira	Tristeza	Sorpresa	Repugnancia
1	5	4	4	5	4	3
2	5	4	3	4	5	3
3	4	3	4	3	4	3
4	4	4	4	3	5	4
5	5	4	3	4	4	3
6	4	3	3	3	4	4
7	5	4	4	4	5	4

8	4	4	3	5	4	3
9	5	4	3	3	5	4
10	4	5	4	4	4	3

En la tabla 5 se muestra el promedio de calificaciones de los dos expertos por cada una de las expresiones faciales.

Experto	Alegría	Miedo	Ira	Tristeza	Sorpresa	Repugnancia
1	4.2	3.7	3.2	3.6	4.6	3.6
2	4.5	3.9	3.5	3.8	4.4	3.4

De esta forma podemos ver que las expresiones faciales donde los movimientos de los puntos de referencia faciales son más amplios tienen mejor evaluación ya que permite al sistema poder obtener una mayor distancia.

4.3 Conclusiones parciales

En este capítulo se definieron los expertos y la evaluación que realizan. Esta evaluación es cuantitativa y va de 1 a 5. Las expresiones para evaluar la propuesta se obtuvieron de la rueda de Plutchik y de acuerdo con los resultados obtenidos la metodología propuesta tuvo resultados favorables.

Capítulo 5. Conclusiones

En este trabajo se presenta una metodología para la imitación de expresiones faciales humanas en un robot humanoide, de esta propuesta se derivan las siguientes aportaciones:

1. Evaluación de dos algoritmos de aprendizaje automático para la identificación en tiempo real de expresiones faciales en seres humanos.
2. Metodología para el control de servomotores de un robot humanoide, que permitan imitar las expresiones faciales de seres humanos.

Para la selección de los puntos faciales de referencia se evaluaron dos algoritmos, DLib y MediaPipe Face Mesh, de los cuales se decidió utilizar MediaPipe Face Mesh ya que detecta más puntos faciales lo que lo hace más robusto y permite mapear cada servomotor con dichos puntos.

La metodología propuesta consta de 4 etapas: adquisición del video, obtención de los puntos de referencia, cálculo de movimientos y movimiento de los servomotores. Esta metodología se implementó en un robot humanoide utilizando ROS para la comunicación entre los nodos del robot.

La evaluación de la metodología se realizó utilizando la evaluación subjetiva mediante dos expertos, los cuales evalúan el parecido entre las expresiones faciales del robot y el humano. Se realizó la evaluación con 10 personas, las cuales tuvieron que realizar las mismas expresiones, basándose en la rueda Plutchik. De acuerdo con los resultados se considera que la imitación de las expresiones faciales humanas por el robot humanoide Arthur son favorables, principalmente en las expresiones donde las distancias en el movimiento de los puntos son mayores. Algunas expresiones faciales tales como la ira y la tristeza presentan resultados discretos por la pequeña distancia en los movimientos en la boca fundamentalmente.

Referencias

1. ASALE, R.-, RAE: robot | Diccionario de la lengua española, <https://dle.rae.es/robot>, last accessed 2023/07/16.
2. Asheber, W.T. et al.: Humanoid Head Face Mechanism with Expandable Facial Expressions. *International Journal of Advanced Robotic Systems*. 13, 1, 29 (2016). <https://doi.org/10.5772/62181>.
3. Baturone, A.O.: *Robótica: Manipuladores y Robots Móviles*. Marcombo (2005).
4. Boesch, G.: A Guide to OpenPose in 2022, <https://viso.ai/deep-learning/openpose/>, last accessed 2022/10/16.
5. Camargo Díaz, J.L. et al.: Algoritmo para la obtención de parámetros antropométricos en imágenes de rostros frontales. *Ingeniería y Desarrollo*. 29, 1, 127–152 (2011).
6. Chen, B. et al.: Smile Like You Mean It: Driving Animatronic Robotic Face with Learned Models. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 2739–2746 (2021). <https://doi.org/10.1109/ICRA48506.2021.9560797>.
7. Cortés, F.R.: *Robótica: control de robots manipuladores*. Marcombo (2020).
8. Cunningham, P. et al.: Supervised Learning. In: Cord, M. and Cunningham, P. (eds.) *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*. pp. 21–49 Springer, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-75171-7_2.
9. Delgado Agudelo, D.M. et al.: Estimación del atributo satisfacción en test de usuarios a partir del análisis de la expresión facial. *Revista Ingenierías Universidad de Medellín*. 19, 36, 13–28 (2020). <https://doi.org/10.22395/rium.v19n36a01>.
10. Déniz, O. et al.: Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*. 32, 12, 1598–1603 (2011). <https://doi.org/10.1016/j.patrec.2011.01.004>.
11. Dodon, M. et al.: Distance Measurement Method Based on Average Interpupillary Distance. In: 2021 International Conference on Electromechanical and Energy Systems (SIELMEN). pp. 345–348 (2021). <https://doi.org/10.1109/SIELMEN53755.2021.9600308>.
12. Esfandbod, A. et al.: Human-Robot Interaction based on Facial Expression Imitation. In: 2019 7th International Conference on Robotics and Mechatronics (ICRoM). pp. 69–73 (2019). <https://doi.org/10.1109/ICRoM48714.2019.9071837>.
13. Hanson, D.: *Hanson Robotics Robot User Manual*, (2015).
14. Kajita, S. et al.: *Introduction to humanoid robotics*. Springer (2014).
15. Kartynnik, Y. et al.: Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs, <http://arxiv.org/abs/1907.06724>, (2019).

16. King, D.E.: Dlib-ml: A Machine Learning Toolkit.
17. LeCun, Y. et al.: Deep learning. *Nature*. 521, 7553, 436–444 (2015). <https://doi.org/10.1038/nature14539>.
18. Li, S., Deng, W.: Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*. 13, 3, 1195–1215 (2022). <https://doi.org/10.1109/TAFFC.2020.2981446>.
19. Mammone, A. et al.: Support vector machines. *WIREs Computational Statistics*. 1, 3, 283–289 (2009). <https://doi.org/10.1002/wics.49>.
20. Mheidly, N. et al.: Effect of Face Masks on Interpersonal Communication During the COVID-19 Pandemic. *Frontiers in Public Health*. 8, (2020).
21. Name, Y.: Overview, https://emmanual.robotis.com/docs/en/software/dynamixel/dynamixel_sdk/overview/, last accessed 2023/08/17.
22. Name, Y.: ROBOTIS e-Manual, <https://emmanual.robotis.com/docs/en/dxl/rx/rx-28/>, last accessed 2023/07/30.
23. Park, U. et al.: Robot Facial Expression Framework for Enhancing Empathy in Human-Robot Interaction. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). pp. 832–838 (2021). <https://doi.org/10.1109/RO-MAN50785.2021.9515533>.
24. Plutchik, R.: *La rueda de las emociones*, (1980).
25. Soo, S.: Object detection using Haar-cascade Classifier. (2014).
26. Szeliski, R.: *Computer Vision: Algorithms and Applications*. Springer Nature (2022).
27. Tadesse, Y., Priya, S.: Graphical Facial Expression Analysis and Design Method: An Approach to Determine Humanoid Skin Deformation. *Journal of Mechanisms and Robotics*. 4, 021010, (2012). <https://doi.org/10.1115/1.4006519>.
28. Walczak, S.: Artificial Neural Networks. In: *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. pp. 40–53 IGI Global (2019). <https://doi.org/10.4018/978-1-5225-7368-5.ch004>.
29. Wu, J.: Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China. 5, 23, 495 (2017).
30. Zhou, Z.-H.: *Machine Learning*. Springer Nature (2021).
31. ‘Automa cavaliere’: el robot que diseñó Leonardo da Vinci, <https://otech.uaeh.edu.mx/noti/index.php/robt/automa-cavaliere-el-robot-que-diseno-leonardo-da-vinci/>, last accessed 2023/07/14.

32. Dynamixel MX-28R Robot Actuator from Robotis, <https://www.trossenrobotics.com/r/dynamixel-MX-28R-robot-actuator.aspx>, last accessed 2023/07/30.
33. Dynamixel MX-64R Robot Actuator from Robotis, <https://www.trossenrobotics.com/p/MX-64R-dynamixel-robot-actuator.aspx>, last accessed 2023/07/30.
34. Home, <https://opencv.org/>, last accessed 2022/10/17.
35. i·bug - resources - Facial point annotations, <https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/>, last accessed 2023/07/21.
36. La rueda de las emociones de Robert Plutchik: ¿qué es y qué muestra?, <https://psicologiaymente.com/psicologia/rueda-emociones-robert-plutchik>, last accessed 2023/07/27.
37. LENGUAJE CORPORAL, COMPORTAMIENTOS Y ESTÍMULOS - 2º Expresiones faciales y microexpresiones - Wattpad, <https://www.wattpad.com/916797826-%CA%9F%E1%B4%87%C9%B4%C9%A2%E1%B4%9C%E1%B4%80%E1%B4%8A%E1%B4%87-%E1%B4%84%E1%B4%8F%CA%80%E1%B4%98%E1%B4%8F%CA%80%E1%B4%80%CA%9F-%E1%B4%84%E1%B4%8F%E1%B4%8D%E1%B4%98%E1%B4%8F%CA%80%E1%B4%9B%E1%B4%8D%C9%AA%E1%B4%87%C9%B4%E1%B4%9B%E1%B4%8F%EA%9C%B1-%CA%8F-%E1%B4%87%EA%9C%B1%E1%B4%9B%C3%AD%E1%B4%8D%E1%B4%9C%CA%9F%E1%B4%8F%EA%9C%B1-2%C2%B0>, last accessed 2023/07/16.
38. MediaPipe | Google for Developers, <https://developers.google.com/mediapipe/solutions>, last accessed 2023/07/24.
39. MediaPipe Solutions guide | Google for Developers, <https://developers.google.com/mediapipe/solutions/guide>, last accessed 2023/07/24.
40. Papers with Code - 300W Dataset, <https://paperswithcode.com/dataset/300w>, last accessed 2023/07/21.
41. Pololu - Mini Maestro 24-Channel USB Servo Controller (Assembled), <https://www.pololu.com/product/1356>, last accessed 2023/07/30.
42. pySerial — pySerial 3.4 documentation, <https://pyserial.readthedocs.io/en/latest/pyserial.html>, last accessed 2023/08/17.
43. ¿Qué son las redes neuronales convolucionales? | IBM, <https://www.ibm.com/mx-es/topics/convolutional-neural-networks>, last accessed 2023/07/19.

44. Robotics facial expression of anger in collaborative human–robot interaction, <https://journals.sagepub.com/doi/epub/10.1177/1729881418817972>, last accessed 2022/10/14. <https://doi.org/10.1177/1729881418817972>.
45. ROS - Software de programación by automationware | DirectIndustry, <https://www.directindustry.es/prod/automationware/product-192516-2394360.html>, last accessed 2023/07/25.
46. ROS: Home, <https://www.ros.org/>, last accessed 2022/10/16.
47. Sociable machines - The framework, <http://www.ai.mit.edu/projects/sociable/kismet.html>, last accessed 2023/07/14.
48. TIOBE Index, <https://www.tiobe.com/tiobe-index/>, last accessed 2023/07/26.
49. Welcome to Python.org, <https://www.python.org/>, last accessed 2023/07/25.
50. Welcome to the Maestro servo controller library’s documentation! — Maestro servo controller library 0.1 documentation, <https://maestro-servo.readthedocs.io/en/latest/>, last accessed 2023/08/17.
51. Xpert RC Micro CI-2401 Aluminum Case “Super Torque” High Voltage Brushless Servo, <https://xpertrcshop.com/xpert-rc-micro-ci-2401-aluminum-case-super-torque-high-voltage-brushless-servo-2/>, last accessed 2023/07/30.