

Pontificia Universidad Católica de Valparaíso  
Facultad de Ingeniería  
Escuela de Ingeniería Informática

# **SISTEMA DE RECONOCIMIENTO GESTUAL DE LENGUA DE SEÑAS CHILENA MEDIANTE CÁMARA DIGITAL**

**CARLOS GUILLERMO GONZÁLEZ RIVEROS**  
**FRANCISCO JAVIER YIMES INOSTROZA**

Profesor Guía: **Claudio Cubillos Figueroa**  
Profesor Co-Referente: **Iván Mercado Bermúdez**

Carrera: **Ingeniería de Ejecución Informática**

Octubre 2015

## Resumen

Los gestos expresados con las manos son frecuentemente utilizados debido a su gran nivel de expresividad. Para el reconocimiento de gestos con las manos es necesario percibir posiciones tanto en ángulos como en rotaciones. Para ello, es necesario especificar los puntos de inicio y fin tanto en espacio como en tiempo del patrón gestual y enseguida, segmentar el gesto relevante e interpretar la información. Sin embargo, es posible encontrarse con ambigüedad en la segmentación, una constante variabilidad en el espacio-tiempo y esto sumado a la complejidad morfológica que posee la mano, se convierte en un interesante problema a resolver, sin embargo, la tecnología y las investigaciones sobre algoritmos para perfeccionar el reconocimiento gestual ha ido en aumento, de tal manera que la misma tecnología se ha ido desarrollando a la par para seguir perfeccionando el reconocimiento gestual.

Estas técnicas de procesamiento de imágenes que nos permiten analizar fotogramas y/o imágenes estáticas reales, para adquirir información importante y resolver problemas específicos son parte del estudio de la visión artificial o en su rama más general parte de la inteligencia artificial.

*Palabras-claves: Gestos, Patrón, Segmentar, Visión Artificial, Inteligencia Artificial, Algoritmos, Procesamiento de Imágenes, Reconocimiento Gestual, Espacio-Tiempo y Tecnología.*

## Abstract

Expressed gestures with hands are often used because of its high level of expression. For the recognition of hand gestures it is necessary to perceive both positions and rotation angles. To do this, you must specify the start and end points in both space and time and then gestural pattern, segment the gesture and interpreting relevant information. However, it is possible to find ambiguity in segmentation, constant variability in space-time and this added to the morphological complexity that has the hand, becomes an interesting problem to solve, however, technology and research algorithms to refine the gesture recognition has increased, so that the same technology has been developed in tandem to further improve the gesture recognition.

These imaging techniques that allow us to analyze frames and / or actual static images, to acquire important information and resolve specific problems are part of the study of artificial vision or in its most general branch of artificial intelligence.

*Keywords: Gestures, Pattern, Segment, Computer Vision, Artificial Intelligence, Algorithms, Image Processing, Gesture Recognition, Space-Time, and Technology.*

# Índice

<b>Glosario de Términos.....</b>	<b>iv</b>
<b>Lista de Abreviaturas o Siglas .....</b>	<b>v</b>
<b>Lista de Figuras .....</b>	<b>vi</b>
<b>Lista de Tablas .....</b>	<b>vii</b>
<b>Introducción .....</b>	<b>1</b>
<b>1 Planteamiento del Tema .....</b>	<b>2</b>
1.1 Problemática y Motivación .....	2
1.2 Definición de Objetivos .....	2
1.2.1 Objetivo general .....	2
1.2.2 Objetivos específicos.....	2
1.3 Planificación del Proyecto .....	3
1.4 Estructura del Documento .....	5
1.5 Metodología de Desarrollo .....	5
1.5.1 Metodología Rapid Application Development .....	5
<b>2 Estado del Arte .....</b>	<b>7</b>
2.1 Reconocimiento Gestual .....	7
2.1.1 Clasificación.....	7
2.1.2 Aplicaciones .....	8
2.1.3 Herramientas y/o Técnicas .....	8
2.1.3.1 Modelos Ocultos de Markov (HMM) .....	8
2.1.3.2 Aproximación por máquinas de Estados Finitos (FSM) .....	9
2.2 Modelado de Gestos.....	9
2.2.1 Taxonomía (Clasificación) Gestual.....	9
2.3 Análisis de Gestos.....	10
2.3.1 Detección de Características .....	10
2.3.2 Estimación de Parámetros .....	10
2.4 Sistemas de Reconocimiento Gestual .....	11
2.4.1 LeapMotion .....	11
2.4.2 Wised .....	12
2.4.3 Kinect .....	12
2.4.4 Openarch .....	13
2.5 Alternativas de Desarrollo Gestual .....	13
2.5.1 AForge.NET .....	14

2.5.2	OpenCV.....	14
2.5.3	Motion.py .....	14
2.5.4	SoftCollection VideoMotionDetection .....	14
2.5.5	VisionLab VCL.....	14
<b>3</b>	<b>Marco Teórico .....</b>	<b>15</b>
3.1	Lengua de Señas .....	15
3.2	Imágenes Digitales.....	16
3.2.1	Tipo de Imágenes Digitales.....	16
3.2.1.1	Imágenes de Mapa Bits .....	16
3.2.1.2	Imágenes Vectoriales .....	16
3.2.2	Colores .....	17
3.2.2.1	Espacio RGB .....	17
3.2.2.2	Espacio HSI.....	18
3.3	Preprocesamiento de Imágenes.....	19
3.3.1	Manipulación de Contraste.....	19
3.3.2	Eliminación de Ruido.....	19
3.3.3	Realce de Bordes .....	20
3.3.3.1	Método de Desplazamiento y Sustracción .....	21
3.4	Extracción de Características .....	21
3.4.1	Detección de Bordes.....	21
3.4.2	Texturas .....	22
3.4.3	Detección de Movimiento .....	23
3.5	Segmentación de Imágenes.....	23
<b>4</b>	<b>Conclusión.....</b>	<b>24</b>
<b>5</b>	<b>Referencias Bibliográficas .....</b>	<b>25</b>

## Glosario de Términos

Clustering: Algoritmo de agrupamiento.

Lengua: Se llama "lengua" al conjunto o sistema de formas o signos orales y escritos que sirven para la comunicación entre las personas de una misma comunidad lingüística.

Lenguaje: El lenguaje es el medio de comunicación de los seres humanos, quienes utilizamos signos orales y escritos, sonidos y gestos que poseen un significado que les hemos atribuido.

Fotograma: Secuencia de imágenes captadas por cámaras de video digital.

E-Learning: educación y capacitación a través de Internet. Este tipo de enseñanza online permite la interacción del usuario con el material mediante la utilización de diversas herramientas informáticas.

Segmentar: En fotografía, es el proceso de dividir una imagen digital en varias partes (grupos de píxeles) u objetos.

Píxel: Unidad básica de una imagen digitalizada en pantalla a base de puntos de color o en escala de grises.

Biometría: Aplicación de métodos estadísticos y cálculo en el estudio de los fenómenos biológicos.

Tracking: Seguimiento de movimiento, rastreo.

Estocástico: Que está sometido al azar y que es objeto de análisis estadístico.

2D: "2D" es sinónimo de "dos dimensiones". Las formas 2D incluyen cuadrados y círculos.

3D: "3D" significa tridimensional. Las formas 3D tienen profundidad, e incluyen cubos y esferas.

Monocromático: Que está compuesto de vibraciones de la misma frecuencia o bien un solo color.

Punto o Grano: Es el elemento más simple y a la vez más complejo de la imagen. Es el elemento más utilizado en el plano del diseño. Morfológica y compositivamente, el punto no es la representación geométrica de ese concepto sino que tiene una dimensión relativa y variable pudiendo adoptar infinitas formas.

Luminancia: Resultado que se obtiene al dividir la intensidad luminosa de una superficie entre su área aparente para un observador alejado de ella.

## **Lista de Abreviaturas o Siglas**

HCI: Human Computer Interaction.

HMM: Hidden Markov Models.

FSM: Finite State Machine.

MHI: Motion History Image.

D.O.S: Domestic Operation System.

RGB: Red Green Blue.

HSI: Hue Saturation Intensity.

USB: Universal Serial Bus.

## Lista de Figuras

Figura 1.1 Metodología RAD. [Elab. Propia] .....	5
Figura 1.2 Metodología Clásica de Desarrollo de Soft. [Elab. Propia].....	6
Figura 2.1 Ejemplo de Reconocimiento Gestual.....	7
Figura 2.2 Proceso de Reconocimiento Gestual [Elab. Propia] .....	8
Figura 2.3 Taxonomía de gestos de las manos para la HCI. [Elab. Propia].....	10
Figura 2.4 Dispositivo de Reconocimiento LeapMotion. [21] .....	11
Figura 2.5 Variación de onda wifi en la detección de gestos Wised. [23] .....	12
Figura 2.6 Reconocimiento de gestos a través de la Kinect. [Elab. Propia] .....	13
Figura 2.7 Representación gráfica de Openarch en Acción. [Elab. Propia] .....	13
Figura 3.1 Alfabeto de la Lengua de Señas Chilena. [10] .....	15
Figura 3.2 Dos rejillas de 5x5 píxeles. [5] .....	16
Figura 3.3 Representación tridimensional del espacio RGB. ....	18
Figura 3.4 HSI: (a) Tono, (b) Saturación, (c) Intensidad. [6] .....	18
Figura 3.5 (a) Imagen Original, (b) Aumento de Contraste. [2] .....	19
Figura 3.6 Diferencia entre imagen sin ruido y con ruido. [9].....	20
Figura 3.7 Realce de bordes, (a) imagen original, (b) imagen con filtro gaussiano para eliminar ruido, (c) realce de bordes sin filtro gaussiano, (d) realce de bordes con filtro gaussiano. ....	20
Figura 3.8 (a) Imagen Real, (b) Imagen Binarizada. [13] .....	21
Figura 3.9 Propiedades de textura: (a) Suavidad, (b) Rugosidad y (c) Regularidad.....	22
Figura 3.10 Independencia del tono o el color en la textura de una imagen digital.....	22

## **Lista de Tablas**

Tabla 1.1 Carta Gantt Primer Avance Proyecto 1.....	4
-----------------------------------------------------	---



# Introducción

La comunicación gestual ha sido, desde que se tiene conocimiento, el más importante medio de comunicación, no sólo para las personas, sino que para muchas especies animales. Cuando dos personas que hablan distintos idiomas se tratan de comunicar, utilizan gestos para facilitar la comunicación; desde gestos faciales hasta corporales, podemos entender lo que quiere expresar otra persona sin necesidad del uso de palabras; un simple movimiento de la boca, ojos, cejas, manos o cualquier extremidad, representa algo en específico.

Es de esta manera que las personas con dificultad auditiva logran comunicarse, estandarizando significados gestuales de manera amplia y así logrando expresar y compartir tanta información como la comunicación oral. Esta forma de comunicación gestual se denomina Lengua de Señas.

Científicos e ingenieros han trabajado y realizado grandes avances en lo que respecta a reconocer gestos de manera computacional, con distintas finalidades como facilitar el acceso a ciertos trabajos, dar comodidad a usuarios, ayudar a la inclusión de personas con ciertas discapacidades, etc.

En este proyecto pretendemos generar un avance en los estudios relacionados con estas tecnologías y generar un sistema que permita el reconocimiento de la Lengua de Señas, de esta forma poder ayudar y generar inclusión no solo a personas no oyentes, sino que también a aquellos que no pueden usar instrumentos como el mouse o el teclado pero pueden expresarse a través de gestos.

Para lograr esto, analizaremos cómo es el funcionamiento y procesamiento de una imagen digital, y cómo funcionan las tecnologías más avanzadas e importantes relacionadas con el reconocimiento.

# **1 Planteamiento del Tema**

## **1.1 Problemática y Motivación**

Las personas con discapacidad auditiva utilizan la lengua de señas para comunicarse, lengua que solo ellos y un grupo reducido adicional de personas la dominan. Por esta razón es que hasta el día de hoy estas personas no pueden comunicarse con oyentes, y sumándole además que son un porcentaje menor en la población, podemos entender que estas personas se sienten excluidas.

La única forma de poder interactuar con este grupo de personas no oyentes, que no es menor, es aprendiendo la lengua de señas a través de cursos presenciales y/o cursos E-Learning.

La motivación que nos lleva a la realización de este proyecto es, principalmente, el hecho de poder ayudar a personas que tienen problemas para comunicarse y se ven un tanto excluidos por la sociedad.

Tomando en cuenta que la lengua de señas se basa en la utilización de gestos como comunicación, y que las tecnologías de reconocimiento de gestos como los pertenecientes a la lengua de señas se encuentran aún en investigación y desarrollo un tanto inmaduro, encontramos una nueva motivación que es contribuir y constituir una base sólida de conocimientos para el posterior desarrollo de tecnologías que permitan utilizar el reconocimiento gestual en diferentes ámbitos de progreso.

A pesar del inmaduro desarrollo tecnológico respecto al reconocimiento gestual, algunos avances que se han logrado hoy en día han dado mucho de qué hablar, como por ejemplo la llegada de la “Kinect”, la cual ha dado un importante empuje al reconocimiento gestual. Además se han construido importantes librerías de softwares tanto libres como privadas que han ayudado bastante a este tipo de reconocimiento.

Mediante este proyecto se pretende aportar una contribución a los avances en este campo creando y desarrollando una nueva forma de aplicar el reconocimiento gestual a través una cámara digital e implementar un pequeño curso introductorio de la lengua de señas lo más didáctico y entretenido posible, todo esto por medio de dispositivos móviles smartphone o teléfonos inteligentes que permitan que una persona con capacidad limitada y/o una persona de su entorno se comuniquen entre sí, sin la necesidad de saber la lengua de señas.

## **1.2 Definición de Objetivos**

### **1.2.1 Objetivo general**

Desarrollar un sistema de reconocimiento de señas, implementando algoritmos de procesamiento y segmentación de imágenes, pertenecientes a la lengua de señas chilena, a través de una cámara digital.

### **1.2.2 Objetivos específicos**

Los objetivos específicos del actual proyecto son:

- Investigar sobre sistemas y algoritmos de reconocimiento gestual ya existentes.
- Indagar sobre metodologías de enseñanza de lengua de señas.
- Determinar la factibilidad de desarrollo del proyecto de reconocimiento.

- Clasificar y determinar los algoritmos necesarios para el procesamiento de imágenes.
- Especificar y establecer la (s) librería (s) de desarrollo para el sistema de reconocimiento gestual.
- Indagar sobre los sistemas ya existentes de reconocimiento gestual.

### **1.3 Planificación del Proyecto**

El proceso de gestión de un proyecto de software comienza con un conjunto de actividades que, globalmente, se denominan planificación del proyecto. "No podemos pedir exactitud a la fase de planificación, es solo una idea de cómo van a transcurrir las cosas. Hay que planificar el trabajo, los recursos humanos y la tecnología. Planificar es estimar."

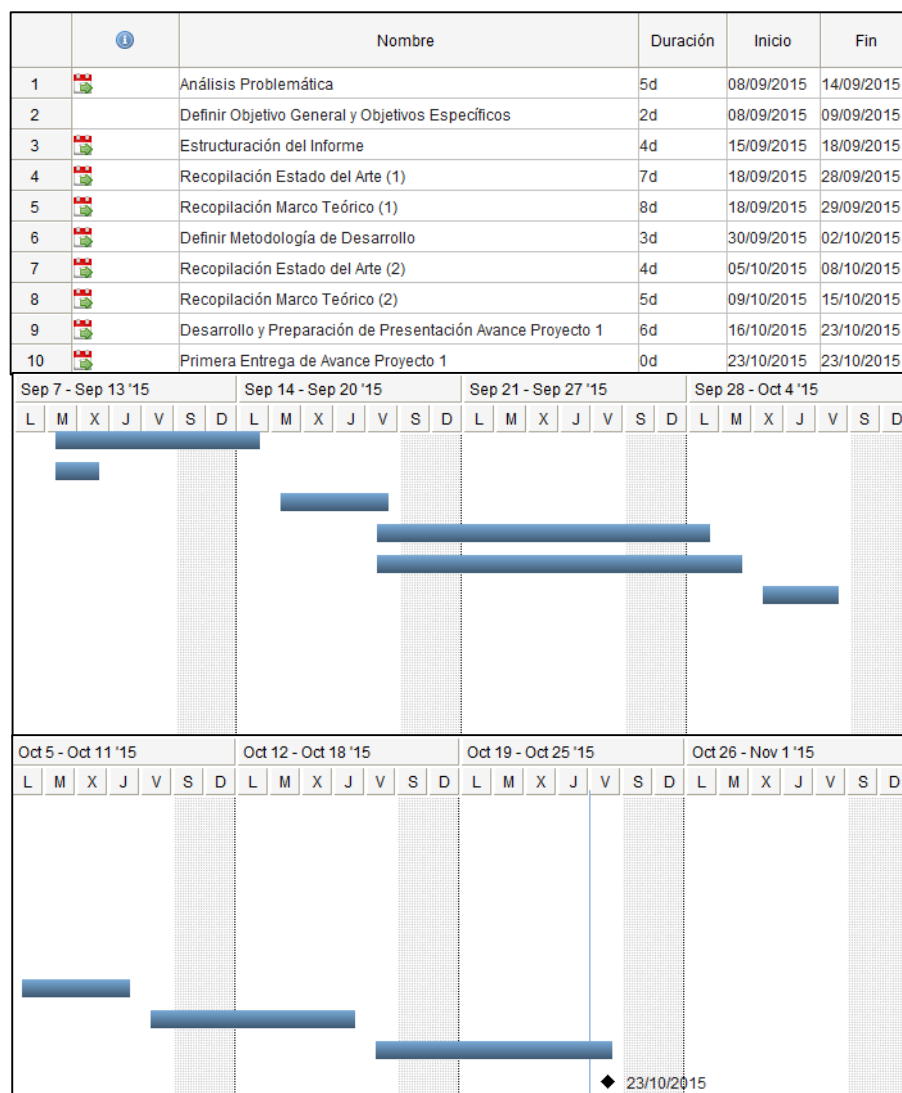
Es por eso que hemos desarrollado una carta Gantt para la correcta elaboración de este primer avance de proyecto, en donde se definieron a grandes rasgos las tareas que se realizaron, no se entró en detalles de la investigación, se tomaron los puntos más relevantes e importantes a desarrollarse a lo largo de este primer avance durante mes y medio de trabajo.

El presente proyecto se basa en dos partes:

1. Proyecto 1: En proyecto 1 existen dos partes de evaluación una presentación de avance y final donde en el avance participan el Profesor guía y el correferente, en la presentación final existe además un examinador invitado.
2. Proyecto 2: En el proyecto 2 existen tres etapas de evaluación avance, presentación del software y presentación final. Todas forman parte de una sola nota.

A continuación se presenta en la tabla 1.1 el plan de trabajo del presente primer avance de proyecto en donde se ve como primera etapa el análisis del problema en cuestión y la estructuración del documento entregable y posteriormente el desarrollo de este.

Tabla 1.1 Carta Gantt Primer Avance Proyecto 1.



## 1.4 Estructura del Documento

El presente documento se divide en 4 capítulos, los cuales serán descritos de manera breve y precisa a continuación:

Capítulo 1, Planteamiento del tema: En este capítulo, se describirá la motivación, se establecerán y definirán objetivos, junto con la metodología de desarrollo, la planificación del proyecto y finalmente estudios de factibilidad.

Capítulo 2, Estado del arte: En este capítulo, se verán las tecnologías que ya se han desarrollado sobre reconocimiento gestual, se definirán el concepto de reconocimiento gestual y las alternativas que tendremos para el desarrollo de nuestro proyecto para su posterior implementación.

Capítulo 3, Marco teórico: En este capítulo, se podrán ver en primera instancia la definición de la lengua de señas chilenas, y se verán los principales aspectos del reconocimiento de imágenes digitales.

Capítulo 4, Conclusiones: En este capítulo, se presentaran las conclusiones a las que se ha llegado tras el marco de investigación de reconocimiento gestual, en donde se concluye también cual herramienta nos facilitara la ayuda necesaria para la realización de nuestro proyecto.

## 1.5 Metodología de Desarrollo

El desarrollo de software no es una tarea fácil. Prueba de ello es que existen numerosas propuestas metodológicas que inciden en distintas dimensiones del proceso de desarrollo. Por una parte tenemos aquellas propuestas más tradicionales que se centran especialmente en el control del proceso, estableciendo rigurosamente las actividades involucradas, los artefactos que se deben producir, y las herramientas y notaciones que se usarán. Estas propuestas han demostrado ser efectivas y necesarias en un gran número de proyectos, pero también han presentado problemas en muchos otros.

Una posible mejora es incluir en los procesos de desarrollo más actividades, más artefactos y más restricciones, basándose en los puntos débiles detectados. Sin embargo, el resultado final sería un proceso de desarrollo más complejo que puede incluso limitar la propia habilidad del equipo para llevar a cabo el proyecto.

### 1.5.1 Metodología Rapid Application Development

El desarrollo rápido de aplicaciones (RAD) es una metodología de desarrollo de software, que implica el desarrollo iterativo y la construcción de prototipos. El desarrollo rápido de aplicaciones es un término originalmente utilizado para describir un proceso de desarrollo de software introducido por James Martin en 1991.

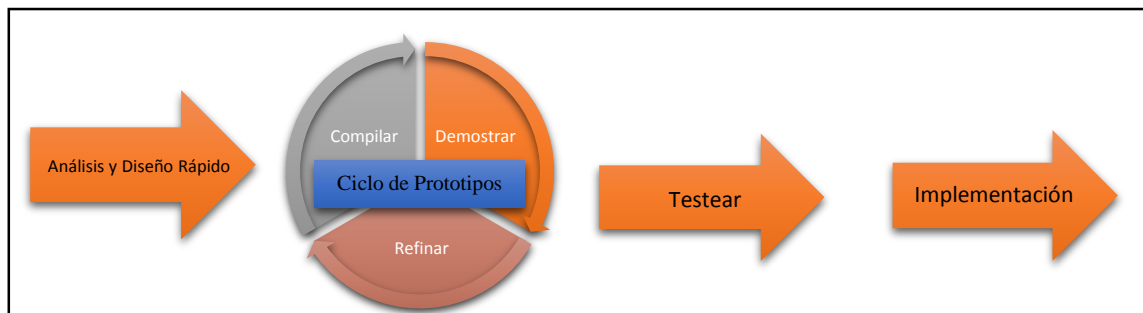


Figura 1.1 Metodología RAD. [Elab. Propia]

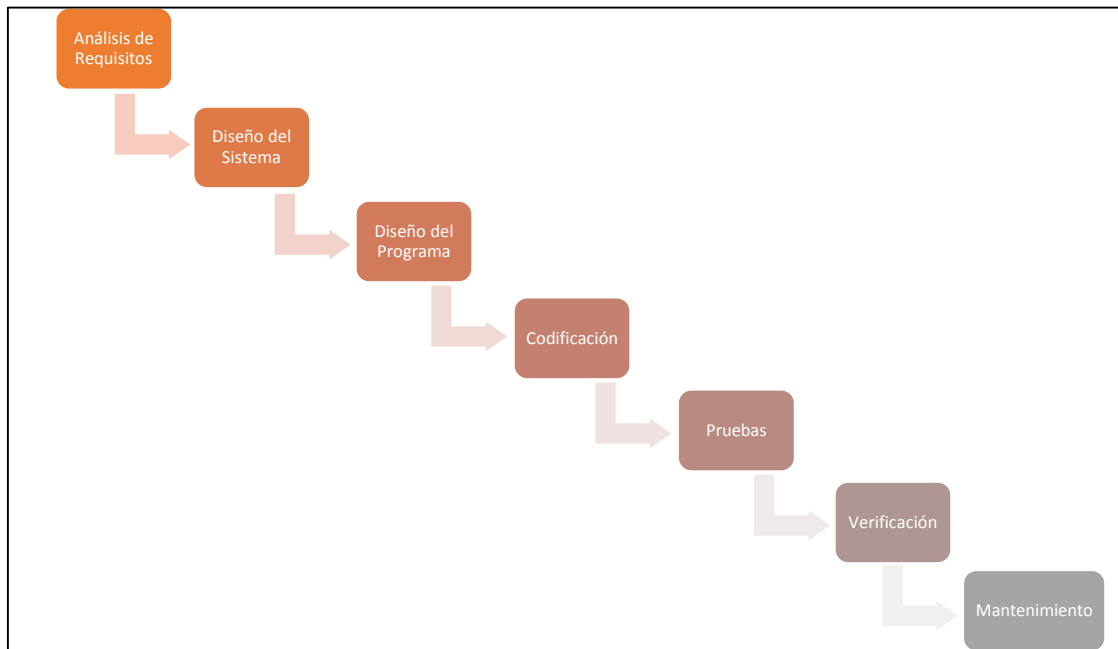


Figura 1.2 Metodología Clásica de Desarrollo de Soft. [Elab. Propia]

Si bien tenemos la intención de desarrollar un sistema, pero el sistema queremos llevarlo a cabo en un dispositivo móvil, es por eso que hemos decidido trabajar con esta metodología y no con otra, ya que para que el reconocimiento de gestos se pueda lograr pasaremos por varias pruebas de imagen, diseño de la aplicación entre muchos otros factores que nos harán ir modificando el proyecto a medida que se logran ciertas metas propuestas a la hora de desarrollar.

Elegir un enfoque clásico o algún otro, nos tomaría mucho tiempo, ya que se deben seguir las normas que dicta la metodología y como nuestro proyecto se basa en prueba y error, lo mejor es ir realizando prototipos para ir viendo alguna funcionalidad en tiempo real.

## 2 Estado del Arte

### 2.1 Reconocimiento Gestual

El gesto en sí es un movimiento corporal con significado, que puede ser realizado con los dedos, las manos, los brazos, la cabeza, con todo el tronco; o incluso con una variación de la expresión de la cara o mediante el movimiento de ojos. Por sí mismo constituye una parte importante de la comunicación natural cotidiana, en muchas ocasiones como refuerzo al lenguaje hablado y, cuando éste no es posible, como la primera alternativa de comunicación. [8]

El reconocimiento gestual es un tema de la visión por computador o bien de la visión artificial, área de la inteligencia artificial, además de la tecnología del lenguaje que tiene como objetivo reconocer e interpretar gestos y/o patrones humanos a través de algoritmos matemáticos aplicados a la imagen que se reconoce mediante ciertas técnicas de visión, por medio de dispositivos de entrada como cámaras, webcams, o dispositivos móviles.

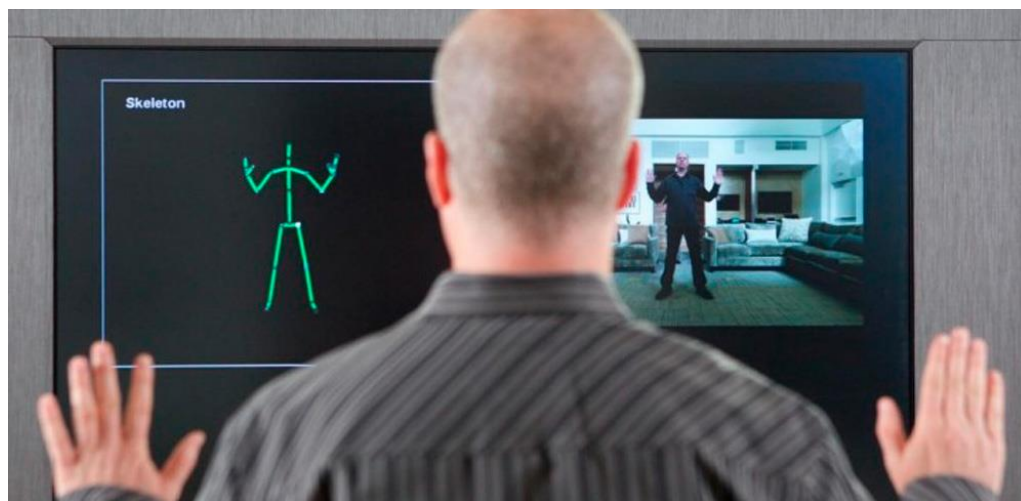


Figura 2.1 Ejemplo de Reconocimiento Gestual.

Como se puede ver en la Figura 2.1, no solo se trata de reconocer manos o cara, sino toda actividad corporal que produzca un cambio significativo a la imagen, pero como nuestro proyecto está centrado en el reconocimiento del alfabeto de lengua de señas chilena, éste está enfocado en el rescate de la forma y movimiento de manos.

#### 2.1.1 Clasificación

Los gestos se pueden clasificar como gestos estáticos y dinámicos:

- Estáticos: El usuario asume cierta postura, o configuración de su cuerpo o de la cara, fija.
- Dinámicos: En los que la información esta codificada en el movimiento, a través de la evolución temporal de las posturas, o bien una mezcla de ambos, como, por ejemplo la lengua de señas. [8]

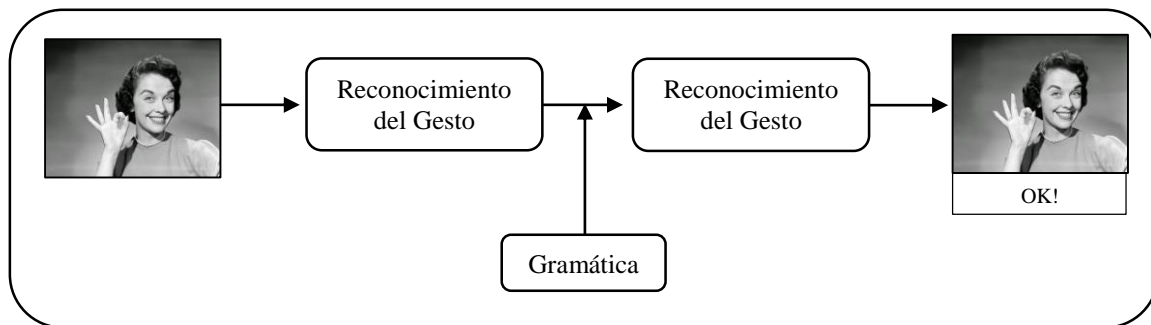


Figura 2.2 Proceso de Reconocimiento Gestual [Elab. Propia]

## 2.1.2 Aplicaciones

Las aplicaciones posibles son numerosas y sólo están limitadas por la imaginación. Éstas podrían influir, no sólo específicamente en la industria de la informática, sino también cambiando la forma en la que las personas se comunican, se transportan o son atendidas por sistemas automáticos, pues dichos sistemas serían más conscientes de la situación física de sus usuarios. Esencialmente el reconocimiento de gestos puede ser extremadamente útil para:

- Reconocimiento de la lengua de señas.
- Navegar o realizar manipulaciones en entornos de realidad virtual o realidad aumentada.
- Computación afectiva. Las expresiones faciales podrían ser usadas por el ordenador para interpretar estados emocionales del usuario.
- Controladores virtuales, manejo de acciones como el apagar la luz en un espacio realizando señas que se reconozcan a través de señales de onda.
- Nuevas tecnologías de ocio y juegos de tipo inmersivo. El estado de los espectadores, posición o postura podría ser captado e integrado en la acción.

## 2.1.3 Herramientas y/o Técnicas

Las técnicas de reconocimiento gestual han experimentado un gran auge en el último tiempo, esto principalmente por la gran cantidad de aplicaciones que está a demostrado tener y lo importante que ha sido para la inclusión de personas en situación de discapacidad, las nuevas modalidades de HCI, lenguaje de señas, navegación en entornos virtuales, reconocimiento facial, biometría aplicado en seguridad, seguimiento o tracking de movimientos, entre muchas otras.

### 2.1.3.1 Modelos Ocultos de Markov (HMM)

Los modelos ocultos de Markov, o HMM, por su sigla en inglés (Hidden Markov Models), son un tipo de modelo estadístico ampliamente utilizado en el reconocimiento del habla y que también es aplicado en el reconocimiento del movimiento humano. Los HMM son empleados para modelar la transición de estados entre un conjunto de modelos dinámicos y, además, tienen la capacidad de modelar dinámicas de alto y bajo nivel [21].

Los HMM son un proceso doblemente estocástico, que consta de una red probabilística con estados ocultos y observables; este proceso está gobernado por una cadena de Markov con un numero finito de estados y un conjunto de funciones aleatorias, cada una asociada a un estado [22]. En los HMM, los estados ocultos corresponden a diferentes fases en la ejecución de una



acción. Los estados de transición están condicionados por un estado previo y no por un historial de estados; esta es la suposición markoviana. A esto se agrega que las observaciones están condicionadas únicamente por el estado actual, por lo cual las siguientes observaciones se consideran independientes [23].

### **2.1.3.2 Aproximación por máquinas de Estados Finitos (FSM)**

Las máquinas de estado finito, o FSM, por su sigla en inglés (Finite State Machine), fueron de las primeras técnicas a ser utilizadas en el reconocimiento de gestos de las manos dado que este tipo de problemas tienen que ver con el procesamiento de secuencias [24]. Utilizando las FSM, los gestos pueden ser modelados como estados en una configuración espacio-temporal. El gesto es reconocido como una serie de trayectorias provenientes de un sensor, constituyendo un ensamble que es representado por medio de puntos en un espacio bidimensional. El reconocimiento de gestos se realiza con una FSM que fue entrenada fuera de línea usando la mayor cantidad posible de ejemplos de cada gesto como datos de entrenamiento [22].

En los problemas de reconocimiento de gestos, en lo que estos son definidos como el cambio de una postura de una mano a otra postura, también han sido usadas las FSM en las que los estados corresponden a las formas de las clases y se consideran todas las posibles transiciones entre estados [25]. Entre las estrategias basadas en estados para aprendizaje y reconocimiento, los gestos son definidos como una secuencia ordenada de estados usando agrupamiento espacial (spatial clustering) y alineamiento temporal [22].

## **2.2 Modelado de Gestos**

Los gestos de la mano no pueden definirse fácilmente y, por lo general, estas definiciones están relacionadas principalmente con los movimientos del cuerpo utilizados para la comunicación. Sin embargo, los significados de los gestos pueden depender de la información espacial, del contexto en el que se desarrollan, de las transiciones que pudieran darse y de la información simbólica y afectiva que transmiten [22].

Entonces, el análisis de gestos y su posterior reconocimiento pueden considerarse como un problema en el que es importante construir un modelo gestual sobre un conjunto de parámetros espaciales y temporales [26]. Teniendo esto en cuenta, existen diferentes formas de determinar el modelo gestual [27] basándose en el tipo de gesto que se debe utilizar, así como diferentes tipos de modelos que pueden ser usados aisladamente o en conjunto.

### **2.2.1 Taxonomía (Clasificación) Gestual**

Es posible definir una taxonomía gestual de manos que se adapte al contexto de la interacción humano-computador (HCI), y que puede ser de utilidad para establecer el modelo de gestos. Pavlovic et al. [26] propone una taxonomía (ver figura 2.3) en la que los movimientos pueden clasificarse en dos grupos: los no intencionales (aquellos que no transmiten ninguna información significativa), y los que constituyen gestos (que a su vez tienen dos modalidades, los comunicativos y los manipulativos).

Los gestos manipulativos son usados para interactuar con objetos en el ambiente; los gestos comunicativos, por lo general, van acompañados de la voz. Los gestos comunicativos pueden clasificarse en actos o símbolos; estos últimos tienen un rol lingüístico y, a su vez, pueden clasificarse en acciones referenciales (referirse a conceptos o a un objeto por el movimiento que realiza) o en modeladores, los cuales son gestos que pueden cambiar el

significado de la comunicación (por lo general verbal) y controlan el modo de interpretación de la misma [28]. Los actos son gestos relacionados directamente con la interpretación propia del movimiento, y se clasifican en miméticos (pueden considerarse pantomimas o mímicas sobre el uso de objetos y acciones con los objetos) y deícticos (relacionados con movimientos para señalar objetos, ubicaciones o direcciones) [29].

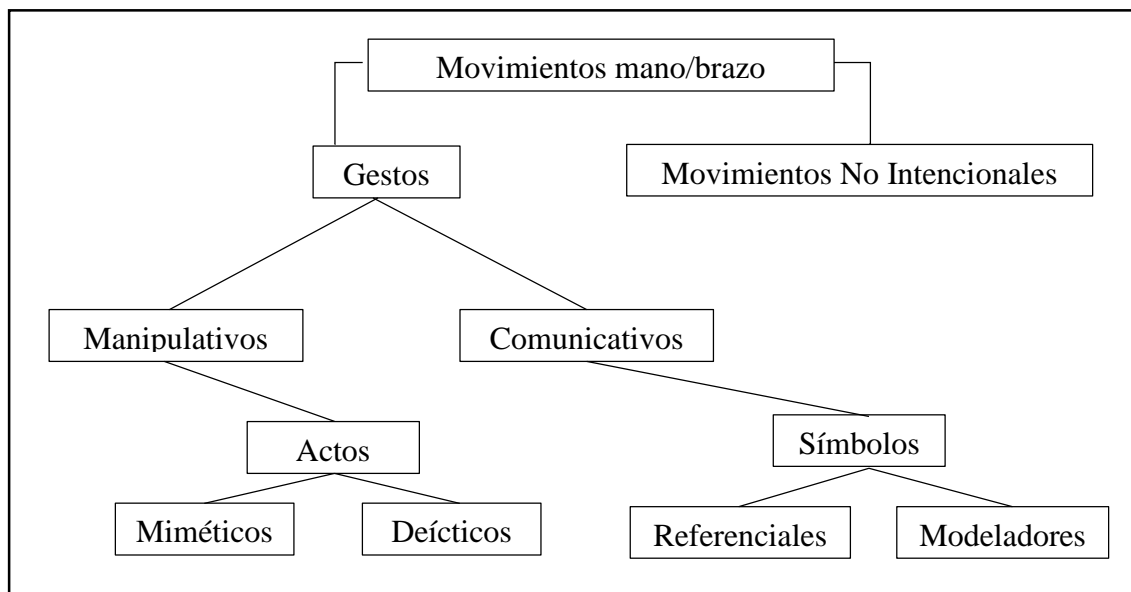


Figura 2.3 Taxonomía de gestos de las manos para la HCI. [Elab. Propia]

## 2.3 Análisis de Gestos

El objetivo de este paso es estimar los parámetros del modelo usando mediciones de las imágenes de video. En este paso se usan dos tareas: la detección de características, en donde se busca extraer características relevantes de la imagen, y la estimación de parámetros, usando las características para calcular los parámetros del modelo [26]. La importancia de este paso es inmensa, teniendo en cuenta que un error en la extracción de características implica la obtención de una mala clasificación.

### 2.3.1 Detección de Características

En la detección de características el objetivo es encontrar y señalar los aspectos claves de la imagen que serán utilizados en la estimación de los parámetros del modelo gestual escogido. Aquí se dan dos procesos: la localización de la persona que realiza los gestos y, una vez completado este proceso, se puede proceder a la detección de las características deseadas [26].

Con la localización se busca detectar a la persona que ejecuta los gestos, diferenciándola del resto de la imagen (o también llamado ambiente). Para ello, se utilizan diferentes modelos y métodos de rastreo [30], en los que por lo general se ejecutan dos pasos: la segmentación de la figura-fondo y la correspondencia temporal.

### 2.3.2 Estimación de Parámetros

La estimación de parámetros es la última etapa de la fase de análisis de gestos, depende de dos factores: las características seleccionadas, y los parámetros requeridos por el modelo de

reconocimiento [26]. Los modelos basados en la apariencia generalmente se usan para describir o identificar acciones gestuales. La estimación de parámetros de estos modelos usa una descripción compacta de la imagen o de la secuencia de imágenes. Existen varias formas de definir parámetros para estos modelos; un primer método es seleccionar un conjunto de fotogramas o frames importantes visualmente como los parámetros del modelo (por ejemplo los frames que contengan el objeto por reconocer) [26]. Otro método es utilizar la acumulación de información espacio-temporal en una secuencia de imágenes y transformarla en una sola imagen 2D; este método se denomina motion history image (MHI) [31]. Esa imagen puede ser fácilmente parametrizada utilizando técnicas de descripción de imágenes en 2D como la descripción de momentos geométricos o la descomposición por valores propios [26]. La mayor ventaja de los enfoques basados en la apariencia es la simplicidad en la etapa de estimación de parámetros; sin embargo, esta simplicidad puede hacer que en algunos casos se pierda precisión en la información espacial de las imágenes [26].

## 2.4 Sistemas de Reconocimiento Gestual

### 2.4.1 LeapMotion

Fundado en el 2010 por Michael Buckwald, David Holz, en San Francisco, California, EE.UU.

El controlador LeapMotion es un pequeño dispositivo USB que está diseñado para ser colocado en un escritorio físico, mirando hacia arriba. Utiliza dos cámaras de infrarrojos monocromática y tres infrarrojos LEDs, el dispositivo observa un área aproximadamente semiesférica, a una distancia de aproximadamente 1 metro. Los LEDs generan patrones con luz Infrarroja y las cámaras generan cerca de 300 fotogramas por segundo, que se envían a través de un cable USB al ordenador central, donde es analizada por el software del controlador de LeapMotion usando "Matemáticas Complejas". Motion Controller resalta un área de observación más pequeña y de mayor resolución a diferencia del producto Kinect, que es más adecuado para todo el cuerpo, seguimiento en un espacio del tamaño de una sala de estar. Algunas de las tareas a realizar son: navegar por un sitio web, el uso de gestos de pellizcar para hacer zoom en los mapas, dibujos de alta precisión, y la manipulación de visualizaciones de datos 3D complejas. [21]



Figura 2.4 Dispositivo de Reconocimiento LeapMotion. [21]

### 2.4.2 Wisee

Wisee es una novedosa interfaz de interacción que aprovecha las señales inalámbricas en curso en el medio ambiente, para permitir la detección y el reconocimiento de gestos humanos. Dado que las señales inalámbricas no requieren de líneas de visión y pueden atravesar las paredes. Wisee permite que en toda la casa sea controlada sin una gran cantidad de fuentes inalámbricas por ejemplo un router wifi y un par de dispositivos móviles en alguna habitación. Wisee es el primer sistema que permite reconocer gestos en la línea de visión, sin línea de visión y a través de la pared de distintos escenarios, no así otros sistemas de reconocimiento de gestos como Kinect, LeapMotion.

Wisee no requiere ninguna infraestructura de cámaras ni de instrumentación de dispositivos para los usuarios. Hasta el momento Wisee detecta 9 gestos con un promedio del 94% de precisión. [23]

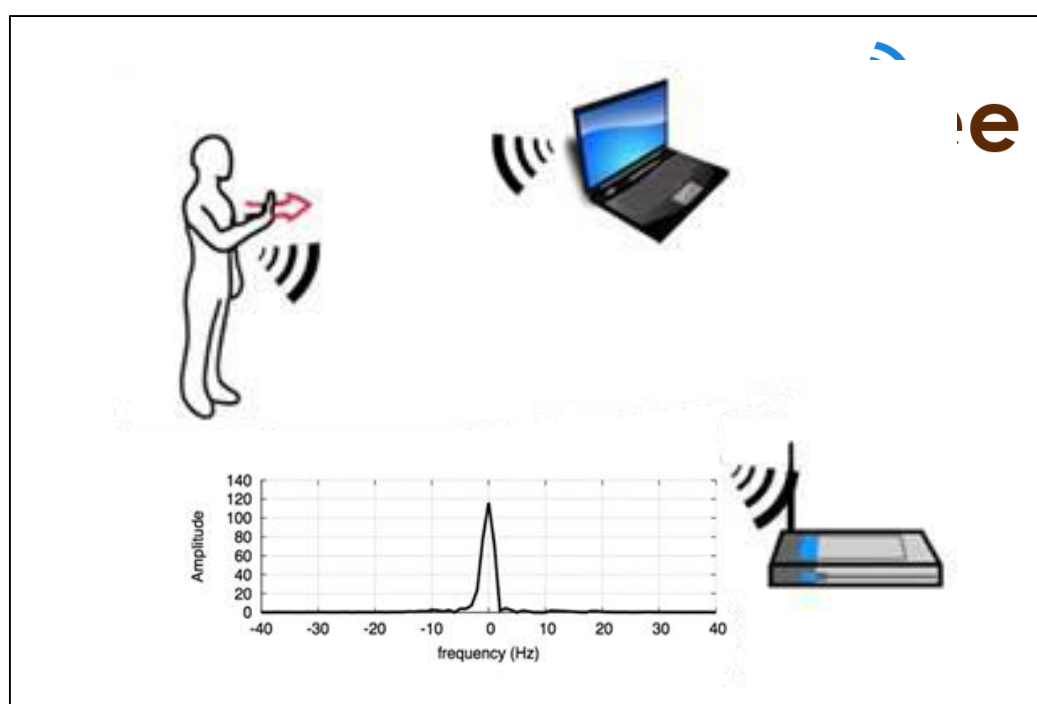


Figura 2.5 Variación de onda wifi en la detección de gestos Wisee. [23]

### 2.4.3 Kinect

Kinect (originalmente conocido por el nombre en clave Project Natal), es un controlador de juego libre y entretenimiento creado por Alex Kipman, desarrollado por Microsoft para la videoconsola Xbox 360, y desde junio del 2011 para PC a través de Windows 7 y Windows 8. Kinect permite a los usuarios controlar e interactuar con la consola sin necesidad de tener contacto físico con un controlador de videojuegos tradicional, mediante una interfaz natural de usuario que reconoce gestos, comandos de voz, objetos e imágenes. El dispositivo tiene como objetivo primordial aumentar el uso de la Xbox 360, más allá de la base de jugadores que posee en la actualidad. Kinect fue lanzado en Norteamérica el 4 de noviembre de 2010. [24]

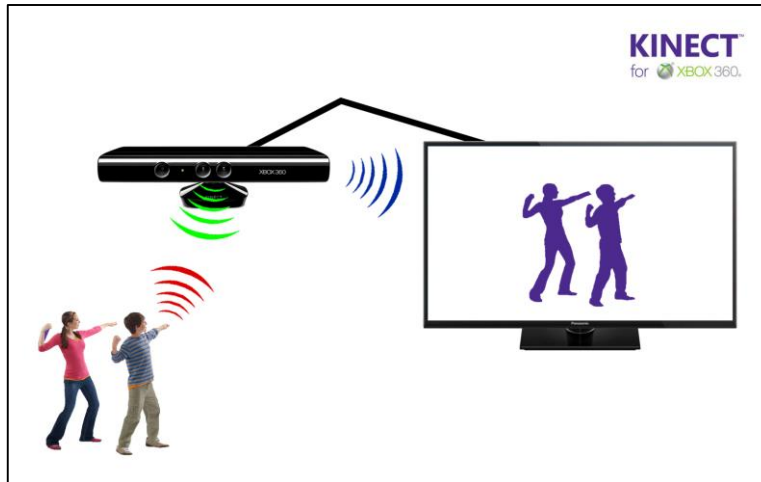


Figura 2.6 Reconocimiento de gestos a través de la Kinect. [Elab. Propia]

#### 2.4.4 Openarch

Openarch es un prototipo real de una vivienda inteligente. La primera vivienda diseñada desde el inicio para incorporar una capa digital que conecta la casa y sus elementos a Internet. Sus habitantes participan y se incorporan a una nueva vida conectada y digital. Es flexible y gracias a su capacidad de transformación se adapta a cualquier condición que requiera el usuario.

La capa digital a la que a partir de ahora llamaremos D.O.S (domestic operating system) incluye una serie de elementos que permiten a los usuarios estar conectados con cualquier persona o espacio, controlar los elementos de la casa mediante el movimiento del cuerpo, realizar conferencias desde la casa, conocer el consumo eléctrico en cada instante, activar cualquier electrodoméstico desde el trabajo, compartir en video y en directo las recetas de cocina con el resto del mundo, crear tu propio plató de TV en el salón, etc. [22]

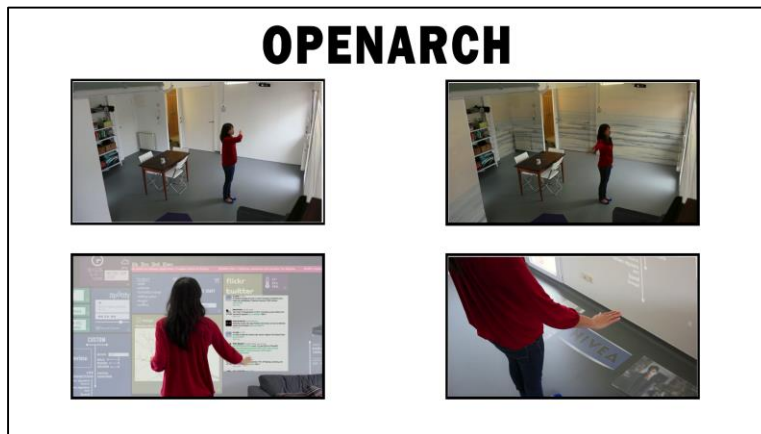


Figura 2.7 Representación gráfica de Openarch en Acción. [Elab. Propia]

### 2.5 Alternativas de Desarrollo Gestual

Se ha realizado una investigación sobre las librerías existentes que pudieran facilitar en todo lo posible la tarea de reconocimiento de movimiento gestual. Se ha realizado una búsqueda

generalizada, considerando cualquier plataforma, valorando las ventajas e inconvenientes de cada una de ellas.

De los resultados obtenidos podríamos destacar dos: AForge.Net y OpenCV, ya que son las más completas que existen.

### **2.5.1 AForge.NET**

AForge.NET Framework es un marco de C # diseñado para desarrolladores e investigadores en los campos de la Visión por Computador e Inteligencia Artificial, procesamiento de imágenes, visión artificial, redes neuronales, algoritmos genéticos, aprendizaje automático, etc. [20]

Tiene la ventaja de ser fácil de usar además de ser una librería de código abierto o libre. El único inconveniente es que no es tan eficiente en el reconocimiento gestual.

### **2.5.2 OpenCV**

OpenCV (Open Source Computer Vision Library) es una fuente abierta de visión por computador y la biblioteca de software de aprendizaje automático. OpenCV fue construido para proporcionar una infraestructura común para aplicaciones de visión por computador y acelerar el uso de la percepción de la máquina en los productos comerciales.

La biblioteca cuenta con más de 2500 algoritmos optimizados, que incluye un amplio conjunto de clásico y algoritmos de visión por computador y aprendizaje automático.

Estos algoritmos se pueden utilizar para detectar y reconocer las caras, identificar objetos, clasificar las acciones humanas en videos, movimientos de cámara pista, objetos de la pista en movimiento, extraer modelos 3D de objetos, producen nubes de puntos 3D de cámaras estéreo, unir imágenes juntos para producir una alta resolución imagen de toda una escena, encontrar imágenes similares de una base de datos de imágenes, eliminar los ojos rojos de las imágenes tomadas con flash, seguir los movimientos de los ojos, reconocer el paisaje y establecer marcadores para cubrirás de realidad aumentada, etc.

Cuenta con interfaces de C ++, C, Python, Java y MATLAB y es compatible con Windows, Linux, Android y Mac OS. OpenCV se inclina principalmente hacia las aplicaciones de visión en tiempo real, además de ser rápida, eficiente y código abierto. [25]

### **2.5.3 Motion.py**

Motion.py es una librería escrita en lenguaje python que permite indicar los píxeles en los que se ha producido cambios en una imagen con respecto a otra. Es fácil de utilizar, también es de código libre o abierto, pero solo trabaja con imágenes y no video.

### **2.5.4 SoftCollection VideoMotionDetection**

Esta biblioteca creada en lenguaje .NET que permite identificar uno o más objetos en movimiento en tiempo real, tiene el inconveniente de ser comercial y de código privado, además no está especializada en el reconocimiento de gestos.

### **2.5.5 VisionLab VCL**

Librería que permite capturar y grabar video, además de detectar objetos, crea aplicaciones fácil y rápidamente, es de licencia comercial.

## 3 Marco Teórico

### 3.1 Lengua de Señas

La lengua de señas es la lengua natural de las personas sordas, se basa en movimientos y expresiones a través de ojos, rostro, boca, cuerpo y principalmente manos, siendo esta última nuestra principal base a estudiar.

Las personas sordas nacen con la capacidad biológica del lenguaje intacta. Sin embargo, su limitación auditiva impide que puedan apropiarse del idioma o lengua oral que se habla a su alrededor.

Al igual que los idiomas orales, existen diferentes lenguas de señas. Esto debido a que como ha pasado con los demás idiomas, las lenguas de señas han surgido de acuerdo a los procesos de educación o de influencia de unos países sobre otros al igual que a las características propias de cada región, así por ejemplo existe una lengua de señas colombiana, una lengua de señas argentina, una lengua de señas española, una lengua de señas americana, que no son iguales. Lo mismo que sucede con los idiomas orales, (español, francés, inglés de estados unidos, inglés de Inglaterra, portugués, etc.) las lenguas de señas también presenta diferencias. [4]

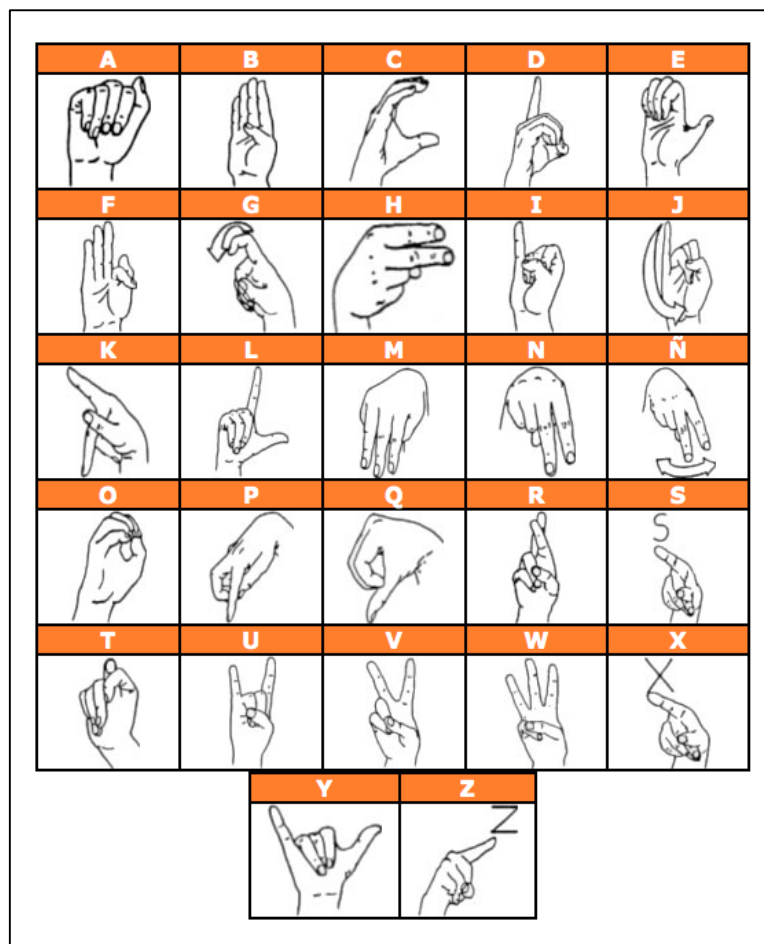


Figura 3.1 Alfabeto de la Lengua de Señas Chilena. [10]

## 3.2 Imágenes Digitales

La imagen digital está formada por un conjunto definido de puntos llamados píxeles. La imagen tradicional, sobre material posible, también está formada por puntos o granos; la gran diferencia es que en la fotografía tradicional, los granos son irregulares y están situados de forma aleatoria, mientras que en la digital forman una matriz con filas y columnas.

### 3.2.1 Tipo de Imágenes Digitales

Por la forma de manejar los datos en un archivo de imagen, se puede hablar de dos modos principales para manipular la información que integra una imagen digital. Estos modos son las imágenes de mapa de bits y las imágenes vectoriales. Dado que cada uno se adapta mejor a un tipo de imagen, antes de conocer los diferentes espacios, debe conocerse el funcionamiento tanto de imágenes vectoriales como de imágenes de mapa de bits.

#### 3.2.1.1 Imágenes de Mapa Bits

Las imágenes de mapa de bits están formadas por una rejilla de celdas. A cada una de estas celdas, que se denominan píxeles, se le asigna un valor de color y luminancia propios. Por esto, cuando vemos todo el conjunto de celdas, tenemos la ilusión de una imagen de tono continuo.

El pixel es una unidad de información, no una unidad de medida, ya que no se corresponde con un tamaño concreto. Un píxel puede ser muy pequeño (0.1 mm) o muy grande (1 cm). Cuando creamos una imagen de mapa de bits se genera una rejilla específica de píxeles; por esto, al modificar su tamaño, transformamos, a su vez, la distribución y coloración de los píxeles, por lo que los objetos, dentro de la imagen, suelen deformarse. Así los objetos pierden o ganan algunos de los píxeles que los definen, como se muestra en la figura 3.2.

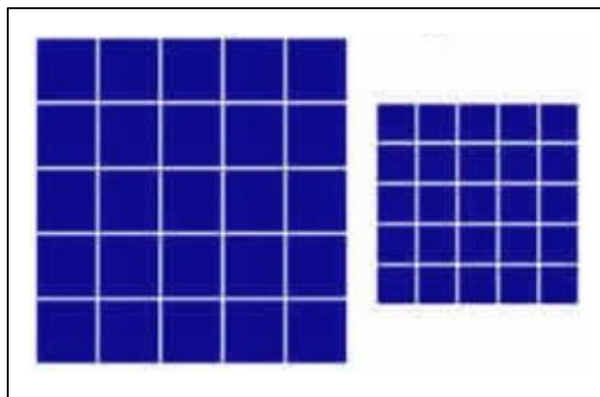


Figura 3.2 Dos rejillas de 5x5 píxeles. [5]

Gracias a esta característica, que siempre hay que tener en cuenta, las imágenes de mapa de bits se crean con un tamaño determinado y pierden calidad si se modifican sus dimensiones.

#### 3.2.1.2 Imágenes Vectoriales

Los llamados gráficos orientados a objetos son las imágenes vectoriales. Su tamaño es mucho más reducido, en comparación con los mapas de bits, porque el modo como organizan la información de una imagen es más simple que en aquellos. Dicha simplicidad radica en



generar los objetos que conforman una imagen a través de trazos geométricos determinados por cálculos y fórmulas matemáticas. De manera tal que los gráficos vectoriales se visualizan a partir de las coordenadas de una línea guardadas como referencia, mismas que forman los objetos a partir de la definición matemática de los puntos y líneas rectas o curvas.

Mientras que en un mapa de bits los objetos se definen píxel por píxel, por lo que no pueden manipularse individualmente, caso contrario sucede con las imágenes vectoriales, con ellas cada objeto, dentro de una imagen, puede modificarse sin que se alteren los demás. En un gráfico de vectores se delimitan por la posición de los puntos inicial y final y por la trayectoria de la línea que los une. Como la información se guarda matemáticamente, esto hace posible que las imágenes vectoriales sean independientes de la resolución del monitor, ya que, al contrario de los mapas de bits, no dependen de los píxeles. Por lo tanto, tienen la mínima o máxima resolución que permita el formato en que se almacenen. Asimismo, al aumentar o reducir la resolución de un gráfico vectorial, tampoco se pierde definición en la imagen, porque la computadora solo tiene que redefinir las coordenadas y vectores que la imagen contenga para redimensionar los objetos. [5]

### **3.2.2 Colores**

Aunque el color es una de las características más importantes que definen a los objetos, hasta hace poco tiempo se le prestaba poca atención debido al coste computacional y a la memoria necesaria para procesar imágenes de color.

Para obtener una imagen hay que transformar primero los parámetros cromáticos en eléctricos y representar los colores, lo cual puede hacerse de distintas maneras, dando lugar a los distintos espacios de color.

Un espacio de color es un método por el que se puede especificar, crear o visualizar cualquier color. Dependiendo del tipo de sensor y aplicación se han desarrollado diversos espacios de colores.

#### **3.2.2.1 Espacio RGB**

El espacio RGB se basa en la combinación de tres señales de luminancia cromática distinta: rojo, verde y azul (Red, Green, Blue). La manera más intuitiva de conseguir un color concreto es determinar la cantidad de color rojo, verde y azul que se necesita combinar, para ello se realiza la suma aritmética de las componentes:  $X = R + G + B$ . [3]

Si tomamos en cuenta que cada una de las componentes tiene variaciones de tonalidad que van desde el 0 al 255 (28), entonces el cubo posee  $(28)^3 = 16,777,216$  colores representados vectorialmente; de esta manera podemos distinguir los extremos como el rojo (255,0,0), verde (0,255,0), azul (0,0,255), negro (0,0,0) y blanco (255,255,255).

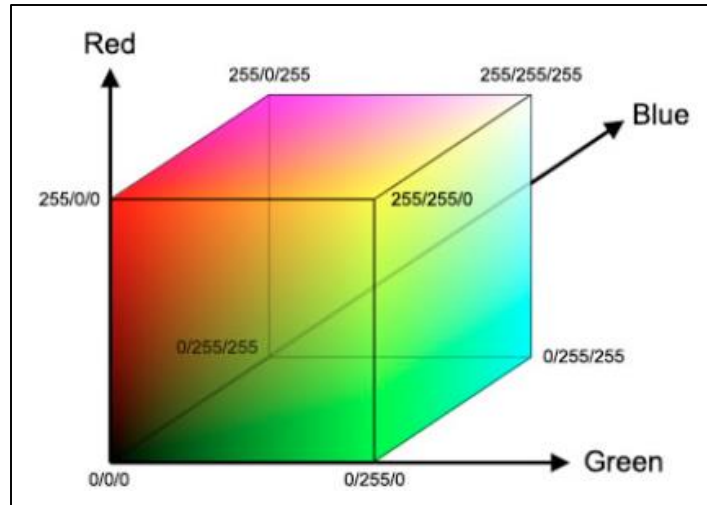


Figura 3.3 Representación tridimensional del espacio RGB.

### 3.2.2.2 Espacio HSI

El espacio de color HSI se basa en el modo de percibir los color que tenemos los humanos. Dicho sistema caracteriza el color en términos de tinte (Hue), saturación (Saturation) y brillo (Intensity).

El tono está asociado con la longitud de onda dominante en una mezcla de ondas luminosas. Así el tono representa el color dominante tal y como lo percibimos; cuando decimos que un objeto es rojo, verde o café, estamos indicando su tono en la Figura 3.4 (a) Tono.

La saturación se refiere a la cantidad de luz blanca mezclada con el color dominante. La saturación es un atributo que nos diferencia un color intenso de uno pálido. Cada uno de los colores primarios tiene su mayor valor de saturación antes de ser mezclados con otros. Así el azul cielos es muy claro (menos saturado), mientras que el azul marino es más opaco (más saturado). Otro ejemplo es el color rosa (rojo y blanco) que está menos saturado; mientras que el color rojo está totalmente saturado en la Figura 3.4 (b) Saturación.

La intensidad representa la iluminación percibida. Da la sensación de que algún objeto refleja más o menos luz. Este atributo lo podemos ver claramente en un televisor en blanco y negro, en la Figura 3.4 (c) Intensidad. [6]

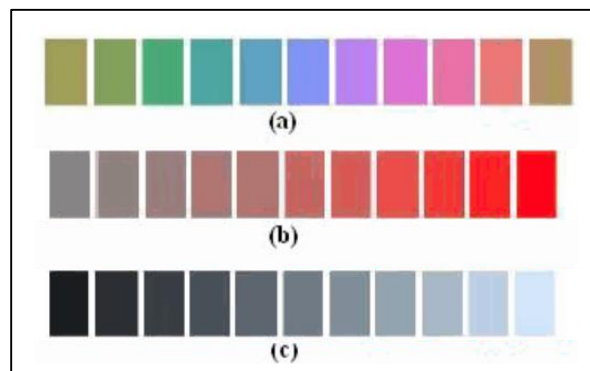


Figura 3.4 HSI: (a) Tono, (b) Saturación, (c) Intensidad. [6]

### 3.3 Preprocesamiento de Imágenes

El preprocesamiento de imágenes comprende aquellos algoritmos cuya finalidad es conseguir una mejora en la apariencia de la imagen original. Esta mejora en la apariencia consiste en resaltar determinadas características de la imagen o en eliminar aquello que las oculta. No se trata de mejorar la imagen de manera que quede lo más ideal posible sino que pueda ser analizada de la forma más simple.

#### 3.3.1 Manipulación de Contraste

El contraste muestra las variaciones locales del brillo, incrementa el cambio de luminosidad entre las zonas más oscuras o más claras de una fotografía, simulando a su vez, un mejor enfoque y claridad de imagen [1]. Su manipulación busca mejorar unas zonas en perjuicio de otras, y así distanciar más los píxeles con valores bajos o a la inversa. En una imagen en escala de grises, consiste en aumentar el rango dinámico de los niveles de gris de la imagen [2].

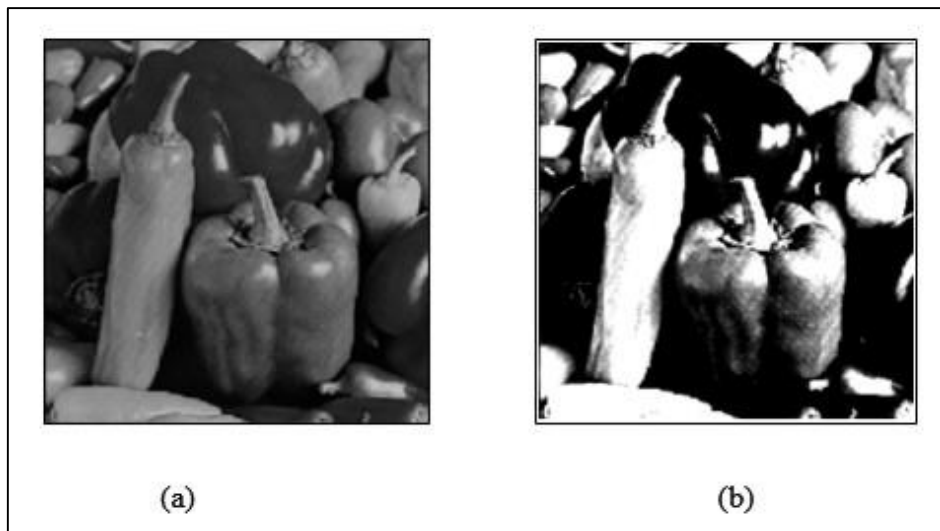


Figura 3.5 (a) Imagen Original, (b) Aumento de Contraste. [2]

#### 3.3.2 Eliminación de Ruido

Todas las imágenes tienen una cierta cantidad de ruido, valores distorsionados, bien debidos al sensor de la cámara o al medio de transmisión de la señal. El ruido se manifiesta generalmente en píxeles aislados que toman un valor gris deferente al de sus vecinos. El ruido puede clasificarse en cuatro tipos:

- Gaussiano: Produce pequeñas variaciones en la imagen. Es debido, por ejemplo, a las diferentes ganancias en el sensor, ruido en los digitalizadores, perturbadores en la transmisión, etc.
- Impulsional: El valor del píxel no tiene relación con el valor ideal sino con el valor del ruido que toma valores muy altos o bajos. Se caracteriza entonces porque el píxel toma un valor máximo, causado por una saturación del sensor, o mínimo si se ha perdido la señal.
- Frecuencial: La imagen obtenida es la suma entre imagen ideal y otra señal, la interferencia, caracterizada por ser una senoide de frecuencia determinada.

- Multiplicativo: La imagen obtenida es la multiplicación de dos señales.



Figura 3.6 Diferencia entre imagen sin ruido y con ruido. [9]

### 3.3.3 Realce de Bordes

El realce de bordes transforma una imagen de manera que exhibe sólo el detalle de bordes o fronteras. Los bordes aparecen como las líneas de contorno de los objetos dentro de la imagen (Figura 3.7 (c) y Figura 3.7 (d)). Estos contornos pueden utilizarse en posteriores operaciones de análisis de imágenes para el reconocimiento de objetos o rasgos. [7]

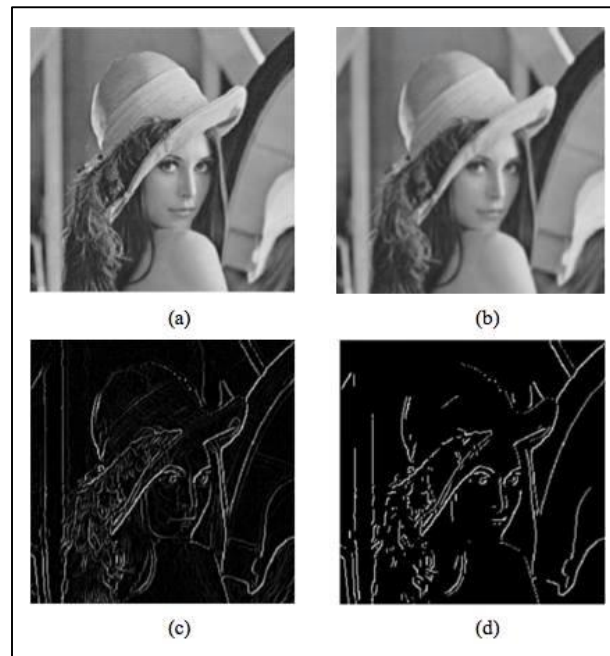


Figura 3.7 Realce de bordes, (a) imagen original, (b) imagen con filtro gaussiano para eliminar ruido, (c) realce de bordes sin filtro gaussiano, (d) realce de bordes con filtro gaussiano.

El realce de bordes tiene un efecto opuesto a la eliminación de ruido, ya que de lo que se trata es de resaltar aquellos píxeles que representan un valor de gris distinto al de sus vecinos. Por ello, si la imagen es ruidosa, el efecto de ruido se multiplicará; por lo que antes de resaltar los bordes, habrá que eliminar el ruido como se muestra en la Figura 3.7.

Los realces de bordes son implementados a través de filtros espaciales. Uno de los más usados es el método de desplazamiento y sustracción.

### 3.3.3.1 Método de Desplazamiento y Sustracción

Permite realzar información de bordes verticales u horizontales, desplazando una imagen a la izquierda en un píxel y restando esta nueva imagen de la original, se resaltarán los bordes verticales. Esto se debe a que el valor de brillo de cada píxel de entrada es restado de su vecino horizontal, dando un valor de la diferencia de brillo entre ambos, conocida como pendiente en el contexto de una imagen.

Si dos píxeles adyacentes tienen brillos muy diferentes (existencia de borde), la diferencia resultará grande. Por el contrario, si dos píxeles adyacentes tienen valores similares de brillo (no existencia de borde), la diferencia de brillos resultante será pequeña. El efecto es una imagen donde aparecen los contornos direccionales. Análogamente, el realce de borde horizontal se implementa desplazando la imagen un píxel hacia arriba y restando [7].

## 3.4 Extracción de Características

Una vez que la imagen ha sido procesada, se puede proceder a detectar los objetos presentes en ella; para ello se busca en la imagen cuáles son las características que definen al objeto, como los bordes, líneas, esquinas, texturas e incluso movimiento.

### 3.4.1 Detección de Bordes

Una de las características más útiles que se encuentran en una imagen, la constituyen los bordes, ya que se utilizan para definir los límites entre si y el fondo de la imagen. [3]

Los bordes de una imagen digital se pueden definir como transiciones entre dos regiones de niveles de gris significativamente distintos. Suministran una valiosa información sobre las fronteras de los objetos y puede ser utilizada para segmentar la imagen, reconocer objetos, etc. [11]

Ya hemos mencionado anteriormente que una de las técnicas para la el realce de bordes es eliminando ruido, con el fin de destacar sólo los bordes importantes.

Otra técnica para la ideal detección de bordes es binarizar la imagen, la binarización consiste en representar los píxeles, ya sea en color o escala de grises, en solo un bit (0 o 1), ya sea completamente negro o completamente blanco; de esta manera los bordes quedarán completamente marcados.

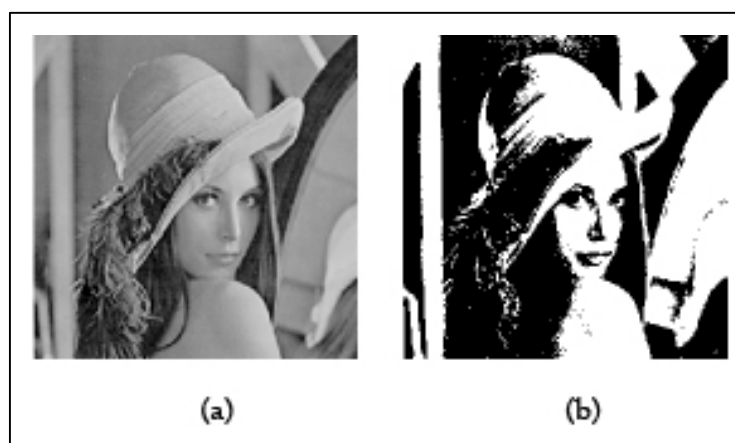


Figura 3.8 (a) Imagen Real, (b) Imagen Binarizada. [13]

### 3.4.2 Texturas

La textura es una característica importante utilizada en segmentación, identificación de objetos o regiones de interés en una imagen y obtención de forma. El uso de la textura para identificar una imagen proviene de la habilidad innata de los humanos para reconocer diferencias texturales. Por medio de la visión y el tacto, el ser humano es capaz de distinguir en forma intuitiva diversos tipos de textura.

La textura, por consiguiente, es una característica de difícil definición siendo la más extendida la siguiente: “Una textura está definida por la uniformidad, densidad, grosor, rugosidad, regularidad, intensidad y direccionalidad de medidas discretas del tono en los píxeles y de sus relaciones espaciales”. En la siguiente figura se muestran las diferencias entre una textura suave (a) y otra rugosa (b), y entre una textura regular (c) de otra que no lo es (b).

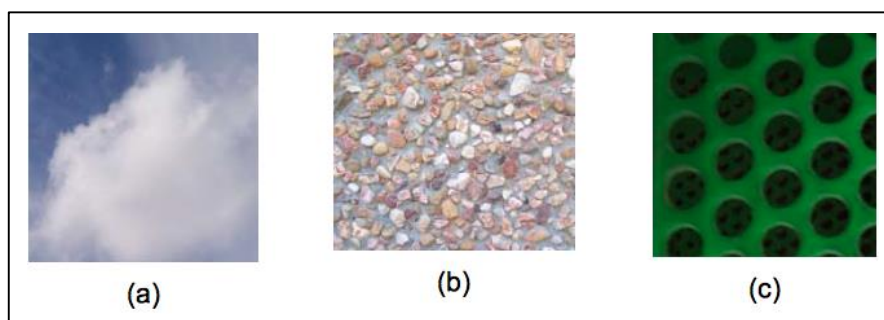


Figura 3.9 Propiedades de textura: (a) Suavidad, (b) Rugosidad y (c) Regularidad

En una imagen digital, la textura depende de tres conceptos:

- La frecuencia de cambio de los tonos en los píxeles.
- La dirección o direcciones de cambio.
- El contraste entre un píxel y sus vecinos.

Sin embargo, es independiente del tono o color de la imagen: en el ejemplo de la figura 3.10 se muestran dos imágenes de una misma textura. Aunque una sea en tonos claros y la otra en tonos oscuros, se observa que se trata del mismo tipo de textura [12].



Figura 3.10 Independencia del tono o el color en la textura de una imagen digital

### 3.4.3 Detección de Movimiento

Otro tipo de característica es la debida el movimiento. Puede ayudar a la segmentación ya que los puntos del mismo objeto presentan unas velocidades semejantes. Los distintos tipos de movimiento que existen son.

- Movimiento de la cámara.
- Movimiento de los objetos.
- Cambios en la iluminación.
- Cambios en la estructura, forma o tamaño del objeto.

El método más sencillo es aquel que se basa en las diferencias de imágenes. Aunque estos algoritmos son menos precisos, son mucho más simples y por lo tanto pueden procesar la imagen en poco tiempo, por lo que su utilidad es mayor. Los pasos a la hora de analizar el movimiento son tres:

- Determinar en qué zonas de la imagen existe algún objeto que esté moviéndose. En esta primera etapa no es importante la exactitud, sólo detectar las futuras zonas de interés en las que se realizará un análisis más detallado.
- En la segunda etapa, la atención sólo está en aquellas zonas donde se haya detectado actividad y se extrae información que pueda ser usada para el reconocimiento de objetos, análisis del movimiento, etc.
- La última etapa tiene el reconocimiento previo del problema en concreto. Así por ejemplo la información obtenida de un objeto pequeño que se mueve lentamente cerca de la cámara de otra igual forma, pero mayor que se mueve rápidamente pero lejos de ella es la misma [3].

## 3.5 Segmentación de Imágenes

La segmentación de imágenes divide la imagen en sus partes constituyentes hasta un nivel de subdivisión en el que se aíslan las regiones u objetos de interés. Los algoritmos de segmentación se basan en una de estas dos propiedades básicas de los valores del nivel de gris: discontinuidad o similitud entre los niveles de gris de píxeles vecinos [14].

- Similitud: Cada uno de los píxeles de un elemento tiene valores parecidos para alguna propiedad.
- Discontinuidad: Los objetos destacan del entorno y tienen, por tanto, unos bordes definidos.
- Conectividad: Los píxeles pertenecientes al mismo objeto tienen que ser contiguos, es decir, deben estar agrupados.

## **4 Conclusión**

Las tecnologías de reconocimiento de señas son bastante complejas, unas utilizan hardware especializado como detectores de ondas específicas, otras, utilizando cámaras normales, contienen algoritmos complejos y sofisticados; y no solo sistemas, sino incluso librerías bastante completas pero sencillas de usar, y muchas de ellas gratuitas. Estas tecnologías, algoritmos y librerías nos dan una clara ventaja a la hora de crear nuestro propio sistema de reconocimiento.

Los conceptos y formas de procesamiento de la imagen digital nos han dado los conocimientos necesarios para el posterior desarrollo. Información que nos ha dado a entender que las probabilidades de éxito son mayores a las pensadas en un principio. De esta forma nos acercamos cada vez más a nuestra meta que es el desarrollo del sistema de reconocimiento de señas, y así en un futuro poder ser un aporte a la investigación, ingeniería e inclusión.



## 5 Referencias Bibliográficas

- [1] “Contraste” (<http://www.digitalfotored.com/imagendigital/contraste.htm>)
- [2] “Realce y Restauración de imagen” ([http://www.lpi.tel.uva.es/muitic/pim/docus/Realce\\_y\\_restauracion.pdf](http://www.lpi.tel.uva.es/muitic/pim/docus/Realce_y_restauracion.pdf))
- [3] Visión por Computador
- [4] “Cómo funciona la lengua de señas” (<http://www.batanga.com/curiosidades/4910/como-funciona-la-lengua-de-senas>)
- [5] “Formatos de Imagen Digital” ([http://www.revista.unam.mx/vol.6/num5/art50/may\\_art50.pdf](http://www.revista.unam.mx/vol.6/num5/art50/may_art50.pdf))
- [6] “Espacio HSI” (<http://www.scielo.org.mx/pdf/rmfe/v54n2/v54n2a11.pdf>)
- [7] “Realce de imágenes: filtrado espacial” (<http://www.aet.org.es/revistas/revista17/AET17-04.pdf>)
- [8] “Reconocimiento gestual mediante técnicas avanzadas de visión por computador” (<ftp://tesis.bbt.ull.es/ccppytec/cp346.pdf>)
- [9] “Ruido en la imagen digital” (<http://www.guiaspracticas.com/camaras-de-fotos/ruido-en-la-imagen-digital>)
- [10] “Alfabeto Lengua de Señas Chilenas” (<http://manosquehablan.com.ar/alfabetos/chile/>)
- [11] “Detección de bordes en una imagen” ([http://www4.ujaen.es/~satorres/practicas/practica3\\_vc.pdf](http://www4.ujaen.es/~satorres/practicas/practica3_vc.pdf))
- [12] “Concepto de Textura” (<http://bibing.us.es/proyectos/abreproy/11494/fichero/PROYECTO%252FCapitulo+3.pdf>)
- [13] “Binarizar Imágenes” ([http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S169233242012000200014](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S169233242012000200014))
- [14] “Segmentación de imágenes” (<http://alojamientos.us.es/gtocom/pid/tema4.pdf>)
- [15] “Descripción AForget.NET” (<http://www.aforgenet.com/framework/docs/>)
- [16] “Descripción LeapMotion” ([https://en.wikipedia.org/wiki/Leap\\_Motion](https://en.wikipedia.org/wiki/Leap_Motion))
- [17] “Descripción Openarch” (<http://www.openarch.cc/es>)
- [18] “Descripción Wisee” (<http://wisee.cs.washington.edu/>)
- [19] “Descripción Kinect” (<https://dev.windows.com/en-us/kinect/tools>)
- [20] “Descripción OpenCV” (<http://opencv.org/>)
- [21] “Hand modeling, analysis and recognition”, Ying Wu and T S Huang, 2001.
- [22] “Gesture Recognition: A Survey, Systems, Man, and Cybernetics, Part C: Applications and Reviews”, S. Mitra and T. Acharya, 2007.
- [23] “Vision-based human motion analysis: An overview, Computer Vision and Image Understanding” Ronald Poppe, 2007.

- [24] “A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition, Computer Vision and Image Understanding”, Agnes Just and Sebastien Marcel, 2009.
- [25] “A Visual System for Hand Gesture Recognition in Human-Computer Interaction, Image Analysis”, Matti-Antero Okkonen, Vili Kellokumpu, Matti Pietikainen, and Janne Heikkila, 2007.
- [26] “Visual interpretation of hand gestures for Human-computer interaction”, V. I. Pavlovic, R. Sharma, and T. S. Huang, 1997.
- [27] “Gesture Recognition Technology”, Pallavi Halarnkar, Sahil Shah, Harsh Shah, HARDIK Shah, and Jay Shah 2012.
- [28] “A feature-based approach to continuous-gesture analysis”, Alan Daniel Wexelblat, 1994.
- [29] “A taxonomy of gestures in human computer interactions”, Maria Karam and M. C. Schraefel, 2005.
- [30] “A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding”, Thomas B Moeslund, Adrian Hilton, and Volker Kruger, 2006.
- [31] “A real time hand gesture recognition system using motion history image”, Chen-Chiung Hsieh, Dung-Hua Liou, and D Lee, 2010.