



UNIVERSIDAD PRIVADA BOLIVIANA  
FACULTAD DE INGENIERÍA Y ARQUITECTURA  
Ingeniería de Sistemas Computacionales

## PRÁCTICA INTERNA

### Desarrollo de un Sistema Interactivo con Body Tracking

Estudiantes:  
Ricardo Fernández  
Juan Diego Garcia  
Brami Prudencio

Docente: Ingeniero Marcelo Lopez

Cochabamba, Septiembre 2020



# Contents

<b>1</b>	<b>Objetivos</b>	<b>2</b>
1.1	Objetivo General . . . . .	2
1.2	Objetivos Específicos . . . . .	2
<b>2</b>	<b>Estudio de Diagnóstico</b>	<b>3</b>
2.1	Seguimiento de la Pandemia . . . . .	3
2.2	Investigación Sobre Los Dispositivos Pioneros en el área de Body Tracking	4
2.2.1	Kinect desarrollado por Microsoft . . . . .	4
2.2.2	PlayStation Camera desarrollado por Sony Computer Entertainment . . . . .	5
2.2.3	Ámbito Clínico . . . . .	6
2.2.4	Ámbito Pedagógico . . . . .	7
2.2.5	Ámbito De Entretenimiento y Entrenamiento . . . . .	8
<b>3</b>	<b>Marco Teórico</b>	<b>12</b>
3.1	Body Tracking/Motion Capture . . . . .	12
3.1.1	Skeletal Tracking . . . . .	14
3.2	Base Teórica . . . . .	15
3.2.1	Cámara y Desambiguación . . . . .	15
3.2.2	Inteligencia Artificial (IA) . . . . .	17
<b>4</b>	<b>Estudio de alternativas</b>	<b>21</b>
4.1	Análisis de Alternativas actualmente disponibles en el mercado . . . . .	21
4.1.1	Entretenimiento . . . . .	21
4.2	Análisis de Alternativas En Herramientas Disponibles Para el Desarrollo	22
4.2.1	Trajes con sensores de movimiento . . . . .	22
4.2.2	TensorFlow . . . . .	23
4.2.3	PoseNet . . . . .	23
4.2.4	Wrnch . . . . .	24
4.2.5	OpenPose . . . . .	25
4.2.6	DeepMotion . . . . .	25
4.2.7	Unity y Kinect . . . . .	25
<b>5</b>	<b>Plan de Actividades</b>	<b>26</b>
5.0.1	Cronograma . . . . .	27

# List of Tables

2.1	Categorías y características del Sistema PEGI . . . . .	10
2.2	Categorías y características del Sistema ESRB . . . . .	11
3.1	OpenPose Output JSON Content, Which are Given X, Y and C values	16

# List of Figures

2.1	Ejemplo de un Kinect y su Uso en el Sistema Interactivo Kinect Sports	5
2.2	Ejemplo de un Kinect y su Uso en el Sistema Interactivo Kinect Sports	6
2.3	Distintos Dispositivos Específicos para Body Tracking . . . . .	6
2.4	Ejemplo de Rehabilitación empleando Body Tracking . . . . .	7
2.5	Ejemplo de Clase de Educación Física empleando Body Tracking . . . . .	8
2.6	Sistema De Clasificación PEGI . . . . .	9
2.7	Sistema De Clasificación ESRB . . . . .	10
3.1	Ejemplo de Clasificación de imagen de TensorFlow . . . . .	13
3.2	Ejemplos de Body Tracking . . . . .	14
3.3	Diversos Tipos de Seguimiento al Esqueleto seleccionados por OpenPose para el desarrollo del proyecto . . . . .	15
3.4	Ejemplo de Visión de una Cámara con sensores RGB-D . . . . .	17
3.5	Ejemplo de Neural Network . . . . .	20
4.1	Traje con Sensores de Movimiento . . . . .	23
4.2	Ejemplos de Datos en la Base de Datos COCO para la estimación de pose	24

# **Objetivos**

## **1.1 Objetivo General**

Desarrollo de un sistema interactivo con Body Tracking, empleando una cámara común para entrenamiento, seguimiento de posiciones, técnicas, baile, de uso personal en el hogar.

## **1.2 Objetivos Específicos**

1. Utilizar software existentes para el seguimiento corporal careciendo de una cámara de profundidad.
2. Implementar una función para registrar mapas propios del usuario.
3. Proveer una alternativa Open Source factible al mercado de sistemas interactivos con Body Tracking tales como Just Dance.

# **Estudio de Diagnóstico**

## **2.1 Seguimiento de la Pandemia**

A causa del COVID-19, a nivel global se han registrado aproximadamente 34 millones de personas contagiadas y 1 millón de muertes confirmadas hasta la fecha 1 de Octubre del año 2020, cuyo desenlace causa pánico y desorden público a través del mundo. La Organización Mundial de la Salud (OMS) declaró al COVID-19 como una emergencia de salud internacional, dando recomendaciones para prevenir y controlar el contagio de la misma. En Bolivia se tiene registrados 135,311 enfermos, de los cuales 7.965 han fallecido y 31.817 casos activos restantes, además de 95.529 pacientes dados de alta, dando al COVID-19 un ratio de mortalidad de 5,89% en Bolivia. Eventualmente se tomaron medidas de prevención, las cuales son fuertemente criticadas independientemente de la calidad de resultados, siendo la principal la declaración de una cuarentena que ha impactado servicios de transporte, comercio, entretenimiento, educativo y muchos más; por lo cual se han estado viendo sofocadas las posibilidades de salir libremente y realizar actividades que solían ser consideradas como cotidianas.

Esta cuarentena implicó que las ciudades del mundo impongan nuevas leyes de confinamiento, implementando advertencias, multas y cárcel para quienes incumplan las mismas, así mismo, se cancelaron eventos festivos, se cerraron escuelas y pospusieron clases, incluso negocios que no eran considerados necesarios para la primera necesidad de la población tuvieron que cerrar temporalmente durante largo tiempo hasta que las personas adoptaron costumbres de higiene personal. El deporte ha sufrido especialmente debido a la influencia del COVID-19 como nunca antes ha acontecido, por ello las personas, especialmente los deportistas precisan de mantener sus rutinas diarias de ejercicio físico y las empresas deben planear nuevos modelos de negocio en orden de ajustarse a los cambios.

Se recomienda hacer ejercicio durante al menos 30 minutos cada día o al menos 20 minutos de exigencia vigorosa. En cuanto a niños, ancianos o enfermos crónicos, consultar con un médico es recomendado. Mantenerse en casa es efectivo para evitar la expansión de esta enfermedad, pero mantener la actividad física es un punto importante.

El ejercicio en casa puede realizarse de demasiadas formas y muy variadas, basta con buscar un ambiente que respete las medidas de prevención para el COVID-19 y sea

un espacio mínimamente grande para moverse cómodamente con los brazos y piernas extendidos, aunque no sean de las mejores prácticas, es incluso suficiente un metro cuadrado para entrenar en distintos deportes o disciplinas. Se pueden exhibir múltiples ejemplos de ejercicio, por lo que es casi imposible tener una excusa, la población que no puede darse el lujo de ejercitarse, está en una situación desplorable económicamente, carece del tiempo debido a un exceso insano de trabajo, problemas de salud u otros. Se pueden realizar entrenamientos desde flexiones, sentadillas, abdominales, trotar en el mismo lugar y hacer juego de pies, así mismo se puede practicar Yoga, Tai Ji Quan, Karate, Tae Kwon Do y diversas artes marciales que poseen técnicas y prácticas tanto en ambientes grandes como pequeños y puede ser practicado en relativamente cualquier momento, solo con la presencia de voluntad y un tiempo bien administrado. Existen innumerables videos y guías de ejercicio para realizar en Internet y televisión e incluso juegos o aplicaciones para poder ejercitarse lo necesario o ir más allá.

Esto no significa que el deporte o el ejercicio deba ser obligatoriamente limitado o que se deben restringir si no se cumple con estas restricciones, sino que esta es una medida más en contra de la expansión del COVID-19 hasta que exista una vacuna o resistencia efectiva ante ella [17].

La tutoría de entrenadores a deportistas siempre ha sido a partir de órdenes y tutoriales cuya base teórica es el enfoque cognitivo (aprender procesando la información a partir de lo que vemos y adquirimos de esa experiencia). Esto significa que la experiencia personal y el entrenamiento previo o visualización de resultados de los entrenadores se verá reflejado en su metodología de enseñanza, señalizando principalmente órdenes de ejecución de movimiento, entrenamiento mental (cuyo objetivo es desarrollar la autoestima y pensamiento/visualización positiva) y la retroalimentación para corregir errores y perfeccionar la técnica[36].

## 2.2 Investigación Sobre Los Dispositivos Pioneros en el área de Body Tracking

En principio, el Body Tracking con cámaras ha ido evolucionando y existen varias herramientas y entornos que se explicarán a futuro en el Estudio de Alternativas, que permiten introducir a campos de estudio y entretenimiento a experimentar con el uso de las herramientas de Body Tracking, como el médico y educativo.

### 2.2.1 Kinect desarrollado por Microsoft

Para exponer un ejemplo revolucionario de esta nueva rama, se menciona el dispositivo Kinect, siendo uno de los dispositivos más famosos, pero desacreditado del potencial que una vez tuvo, fue desarrollado por Microsoft por 20 años, se publicó el año 2010 y su venta se descontinuó en abril del 2016, debido a la llegada de nuevas alternativas como



Figure 2.1: Ejemplo de un Kinect y su Uso en el Sistema Interactivo Kinect Sports

Fuente: [4] [23]

la realidad virtual, su trágica implementación al mercado que decantó a los usuarios y su sobre coste consumaron el hecho, a pesar de ello, desde el 2011 se lo adaptó para PC desde Windows 7 y permitió el desarrollo libre.

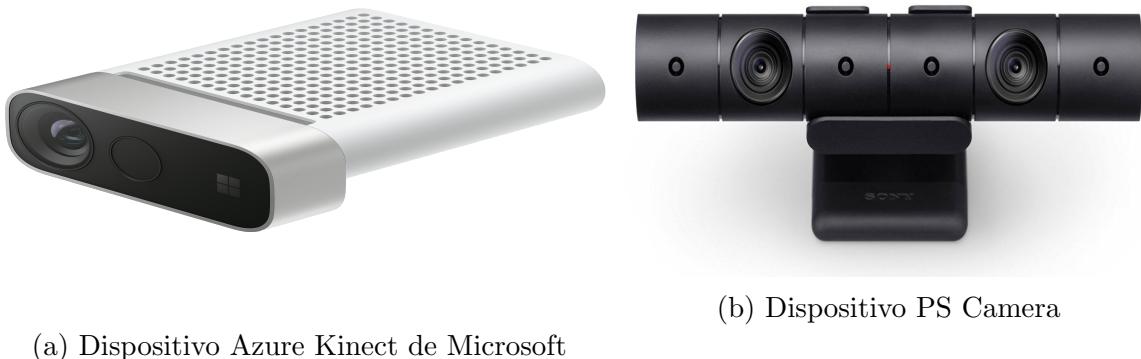
Una fortaleza del desarrollo de esta tecnología son las opciones que enriquecen las posibilidades de los usuarios en su experiencia con la tecnología, siendo el principal beneficiario la industria de los videojuegos. "Kinect ofrece una increíble cantidad de diversión a jugadores casuales, y su creativo concepto de ser libre de un control es innegablemente atractivo" denota una reseña realizada por IGN. Si bien los sistemas interactivos que emplean Body Tracking son duramente criticados tanto por críticos y usuarios, no se puede dudar de su éxito en ventas, se distinguió por ser el dispositivo electrónico de consumo más vendido en el Guiness World Records durante ese año.[16] Posteriormente, al mencionar Sistemas Interactivos con Body Tracking, se requiere de denotar la dificultad de mostrar resultados, debido a la pobreza de condiciones en muchos ambientes que los usuarios tienen, subiendo las expectativas de su desarrollo.

### Azure Kinect

Es una versión moderna y reciente por parte de Microsoft, consta de un sensor de profundidad con opciones de reducción del campo de visión (FOV) para optimización de la aplicación, posee un micrófono para mejorar la captura de sonido, una cámara RGB, un acelero metro, un giroscopio y permite la sincronía de lectura de múltiples dispositivos Kinect para el desarrollo de un proyecto. Sin embargo, este equipo no está disponible, por lo que su sola mención es suficiente.

### 2.2.2 PlayStation Camera desarrollado por Sony Computer Entertainment

El modelo comercial actual es la PlayStation Camera, la cual es un sensor de movimiento y un accesorio de cara para la PlayStation 4 y la PlayStation 5, fue desarrollado por Sony Computer Entertainment.,



(a) Dispositivo Azure Kinect de Microsoft

(b) Dispositivo PS Camera

Figure 2.2: Ejemplo de un Kinect y su Uso en el Sistema Interactivo Kinect Sports  
Fuente: [7] [11]



(a) Dispositivo PS Eye

(b) Ejemplo de Uso de PS Eye

Figure 2.3: Distintos Dispositivos Específicos para Body Tracking

Fuente: [8] [3]

Su modelo previo fue el PlayStation Eye, siendo lanzado al mercado desde el 2007, si bien fue lanzado al mercado antes que el Kinect, sus aplicaciones se distinguían en distintos campos y utilidades. Su visión estaba concentrada en una investigación de chat audio visual con quien tengan conexión, además del reconocimiento facial, de gestos, colores y proporcionar con ello una realidad mixta para socializar.

### 2.2.3 Ámbito Clínico

Realizando un enfoque principal en el papel de la rehabilitación, desde la perspectiva médica, se realizaron estudios sobre un uso alternativo a Body Tracking, distinguiendo su accesibilidad y fácil desarrollo, además que se reportan altos niveles de gozo con la interacción y ejercicio con familiares y amigos.

Se realizaron múltiples estudios, para distintas enfermedades, lesiones y situaciones medicas, se mencionaran un par de ejemplos para señalar su eficiencia. Un grupo de investigadores, estudiaron el tratamiento de deficiencia neurológica empleando consolas, tales como el Sony Playstation®2 [37] y Nintendo Wii [25] como punto de apoyo para



Figure 2.4: Ejemplo de Rehabilitación empleando Body Tracking  
Fuente: [6]

empleo en terapias, lo que niega la objetividad negativa hacia la iniciativa del empleo de sistemas interactivos de esta índole.

La detección de movimiento y la tecnología gráfica empleada en software comerciales no se limitan al movimiento de un control convencional, permiten al usuario sentir estímulos necesarios para optimizar las habilidades motoras precisas que se buscan rehabilitar, más allá de agitarse de un lado a otro. Como teoría no cumplen los requerimientos de intervención precisa de una manera sistemática, necesarios desde el punto de vista tradicional, sin embargo, su potencial puede ser útil en el camino de la rehabilitación [28].

Siendo un estudio en el ámbito, la rehabilitación basada en Simuladores Interactivos, creando desafíos de bajo costo y desarrollados de manera Amateur, para secundar la motivación de personas hospitalizadas, particularmente aquellas que necesitan recuperarse de una lesión de trauma cerebral, médula espinal o amputaciones [28]. La retroalimentación indica que en general, se enmarcaron los desafíos como divertidos y retadores, viendo con positivismo su uso posterior en el hogar de estar disponible [29].

## 2.2.4 Ámbito Pedagógico

Si bien se suspendieron las clases de este año académico a nivel primaria-secundaria en Bolivia, una obligación que consume una sustancial cantidad de tiempo es la asistencia escolar y una de las facetas importantes, la implementación de actividad física, conlleva un potencial impacto en la capacidad del estudiante de evadir una vida sedentaria reclusa y comprometerlo a actividades moderadamente intensas de práctica fuera de la labor estudiantil [19].

Se realizaron estudios sobre el potencial de los sistemas interactivos y discusiones sobre la facilidad que proporcionan en la educación y enseñanza, sin embargo, el presupuesto necesario no es necesariamente bajo para un colegio, lo cual ha privado a Bolivia de considerar estas opciones. Si este caso no impusiera un impacto negativo en



Figure 2.5: Ejemplo de Clase de Educación Física empleando Body Tracking  
Fuente: [21]

la decisión de implementarlos, podría ser considerado limpiamente como oportuno.

Los sistemas interactivos utilizan una tecnología basada en acciones de movimiento, el cual puede envolverse con un importante aspecto interactivo pedagógico, el beneficio a la inteligencia corporal-kinestésica. Se resalta el potencial que ofrece para mejorar la interacción y discusión saludable entre el alumnado y las habilidades de los profesores para manipular y presentar equipos y material multimedia. Como una herramienta educativa, tiene la capacidad de impulsar la motivación, promover el aprendizaje e interés de las actividades, crear un entorno agradable y competitivo en la clase[35].

Sin embargo, no se puede descartar las dificultades del espacio requerido, la necesidad de calibrar los dispositivos y el cambio de metodología pedagógica para incluir el sistema interactivo, por encima de todo, la persuasión a las unidades estudiantiles y a los profesores de modificar la manera tradicional de enseñar educación física, ya que los estudios aún no son claros en su totalidad de que tan tangibles son los resultados en comparación al tradicional [19].

## 2.2.5 Ámbito De Entretenimiento y Entrenamiento

Un campo popular es aquel que requiere de una forma lúdica de realizar actividad física y ejercicio, ya que, al igual que en el ámbito pedagógico y clínico, su objetivo es incitar al usuario a realizar ejercicio de una u otra forma, sin embargo, esta vez el enfoque es distinto, buscando desgastar el cuerpo de los usuarios para poder relajarse o cansarse.

Este ámbito tiene una característica importante y es su amplio margen de usuarios potenciales, los cuales, han sido mermados debido a que hasta el momento, han tenido la necesidad de adquirir dispositivos como Kinect o Play Station Camera para acceder a estos juegos interactivos.



Figure 2.6: Sistema De Clasificación PEGI

Fuente: [2]

### **Clasificación de Edades Para el Uso de Sistemas Interactivos de Entretenimiento y Entrenamiento**

Los sistemas interactivos están diseñados para distintos públicos según las temáticas que presenta el producto, en el caso de entretenimiento y entrenamiento, la población dirigida no tiene restricciones de edad, ya que títulos como Just Dance o Shape Up, ambas de la empresa Francesa Ubisoft, que clasifican sus títulos como apto para todo público en la clasificación por edades, exponiendo los sistemas de clasificación Europeo y Estadounidense, se encuentran el sistema PEGI y ESRB respectivamente.

En Latinoamericano, el año 2014, Colombia ofrece en la ley No. 1554, un nuevo sistema para clasificación de Sistemas Interactivos que se comercialicen de una u otra forma, como el alquiler o venta.

- Circulación Abierta, la clasificación para el público general. En principio de temática deportiva, educativa, informativa y fantástico.
- Circulación Restringida, clasificada para un publico mayor a 18 años. Su contenido referencia actos discriminatorios, conflicto, consumo de sustancias controladas, consumo de bebidas alcohólicas, desnudez, sexo o sexualidad, lenguaje soez, derramamiento de sangre, muerte, lesiones humanas o apuestas por dinero o propiedades.

### **Sistema PEGI**

En el sistema Pan European Game Information (PEGI), cuyo objetivo es clasificar el contenido de sistemas interactivos y software de entretenimiento.

El sistema PEGI fue desarrollado por la Federación Europea de Software Interactivo (ISFE) y se puso en practica desde el 9 de abril de 2003, empleado por 25 países. El sistema cuenta con la evaluación del Instituto Holandés para la Clasificación de Contenido Audiovisual (NICAM), como historial de trabajo, NICAM fue responsable también del sistema holandés para la clasificación de edades de películas.

Todas las categorías contienen un modo de juego en línea y de compras dentro del juego. En la figura 2.6 se encuentran las categorías del sistema PEGI, los cuales son:

Categoría	Características
Apto para todo público	Violencia en un grado moderado
Apto para mayores de 7 años	Violencia moderada, uso de armas y terror moderado.
Apto para mayores de 12 años	Violencia moderada, uso de armas, lesiones, muerte censurada, lenguaje soez, terror moderado, referencias vagas a sexo y apuestas moderadas y vigilancia.
Apto para mayores de 16 años	Violencia, uso de armas, lesiones, muerte, derramamiento de sangre, lenguaje soez, terror, temas sexuales, sustancias controladas y apuestas.
Apto para mayores de 18 años	Además de todo lo anterior, se añade la discriminación y erotismo, solo para personas con criterio o conscientes.

Table 2.1: Categorías y características del Sistema PEGI

Fuente: Elaboración propia basado en [2]



Figure 2.7: Sistema De Clasificación ESRB

Fuente: [1]

## Sistema ESRB

Con el mismo objetivo que el sistema PEGI, el sistema Entertainment Software Rating Board (ESRB), es la versión estadounidense para clasificación de contenido de sistemas interactivos, para asignar categorías que dependen del contenido y temática.

Se estableció en 1994 por la Entertainment Software Association (ESA), la cual es una asociación comercial de la industria de sistemas interactivos en Estados Unidos. Se fundó inicialmente con el nombre de Interactive Digital Software Association (IDSA), renombrado el 16 de julio de 2003. Es conocido mundialmente por ser empleado por empresas que son miembros de ESA, entre ellos Capcom, Square Enix, Ubisoft, Bandai Namco Entertainment, Nintendo.

La ESA también prepara la exposición anual de videojuegos Electronic Entertainment Expo (E3) en California, Los Ángeles.

En la figura 2.7 se encuentran las categorías del sistema ESRB, los cuales son:

Categoría	Características
Early Childhood	Contenido apto para niños, específicamente desarrollados para público infantil. Se retiro la categoría en 2018.
Everyone	Una selección de títulos diseñados para todo público, minimizando violencia fantástica y empleo de insultos leves. Aplica al deporte.
Everyone 10 and up	Permite violencia leve o fantástica, sangre animada, insultos leves. Apto para mayores de 10 años.
Teen	Contiene derramamiento de sangre, temas sugerentes, violencia moderada o humor negro, en varios casos, juegos de azar o apuestas.
Mature 17+	Contienen violencia, sangre, horror, temas sexuales e insultos. Apto para mayores de 17 años.
Adults Only 18+	Escenas prolongadas de violencia o temas sexuales, apuestas y azar con dinero real o ficticio y sangre o desnudez frecuentes. Usualmente contiene temas de controversia o exageraciones que requieren de restricciones.
Rating Pending	Se desconoce su categoría, debido a su reciente lanzamiento, usualmente depende de la intensidad de la temática presentada por el sistema interactivo.

Table 2.2: Categorías y características del Sistema ESRB

Fuente: Elaboración propia basado en [1]

# Marco Teórico

El proyecto es un sistema interactivo que busca proporcionar al usuario independientemente de sus intereses, el cumplimiento de sus expectativas, siendo un desarrollo a conciencia, pero sin tomarlo demasiado en serio, busca también romper el estereotipo que imponen los proyectos de esta modalidad, que si bien, ofrecen un producto de relativa buena calidad y mantenimiento, son o bien productos comerciales como es el caso de Just Dance o productos privatizados como son los desarrollados para la medicina.

En cambio, se tiene la expectativa de ofrecer al nivel de Open source del proyecto, que a la largaatraiga a más miembros, al igual que Linux o Apache y pueda expandir sus horizontes y calidad del producto, teniendo presente que Open Source no significa simplemente compartir el acceso al código fuente, como indica la Open Source Definition (OSD).

El desarrollo del proyecto empleara los recursos disponibles y al alcance de cualquier desarrollador, por tanto, no tomara en cuenta el manejo de cámaras de profundidad, esta aclaración es necesaria, ya que la calidad del producto final puede ser muy variable al de proyectos similares y es una característica más por la que sobresaldría este proyecto, ya que reduciría el presupuesto necesario para el consumidor.

A continuación, se debe mencionar factores importantes sobre la implementación del proyecto, como las propiedades del Body Tracking, la verdadera forma que tiene esta herramienta que tiene para ofrecer al software, por que no sobra mencionar las distinciones entre una cámara normal y una de profundidad.

## 3.1 Body Tracking/Motion Capture

El seguimiento corporal del cuerpo, normalmente conocido como Body Tracking o Motion Capture hace referencia al seguimiento del cuerpo humano a través de una cámara, existen dos acercamientos a este estudio, el enfoque de ajuste del modelo y el enfoque de aprendizaje. El enfoque de ajuste del modelo involucra ajustar el modelo formulado según imágenes previas cargadas, estimando parámetros de puntos especificados de la imagen, sin embargo, es demasiado dependiente de extremos locales y la inicialización adecuada, lo cual lo vuelve inservible en ambientes nuevos. Este modelo comparte una similitud al método Monte Carlo basados en cadenas de Markov [39].

En cambio, el enfoque de aprendizaje requiere de grandes cantidades de imágenes con notas y especificaciones del esqueleto en las imágenes, incrementando las dimensiones

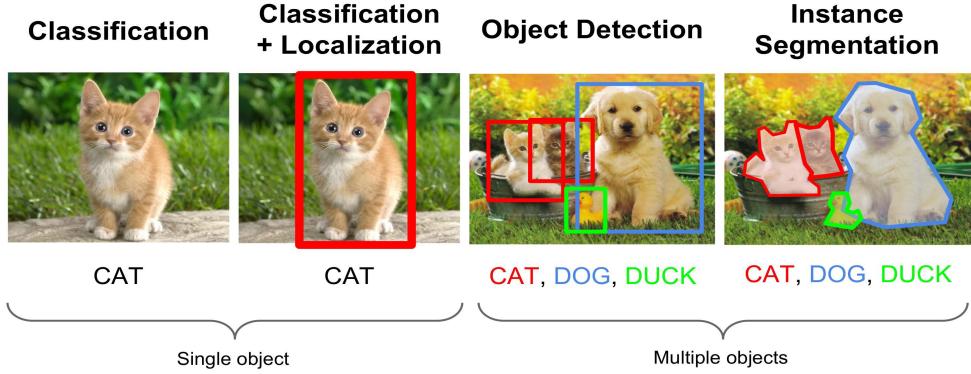


Figure 3.1: Ejemplo de Clasificación de imagen de TensorFlow  
Fuente: [38]

de espacio y memoria requerido. Un ejemplo de este tipo de reconocimiento es la herramienta PoseNet, derivado de TensorFlow, empleado para identificar objetos a partir de una base de datos propia que clasifique los elementos que se buscan identificar con silueta y nombre.

Finalmente se empleo la iteración del punto más cercano [24] el cual usa un enfoque de inicialización del esqueleto a través de fotogramas subsecuentes, clasificándolos con vértices y segmentos en un modelo 3D para este propósito.

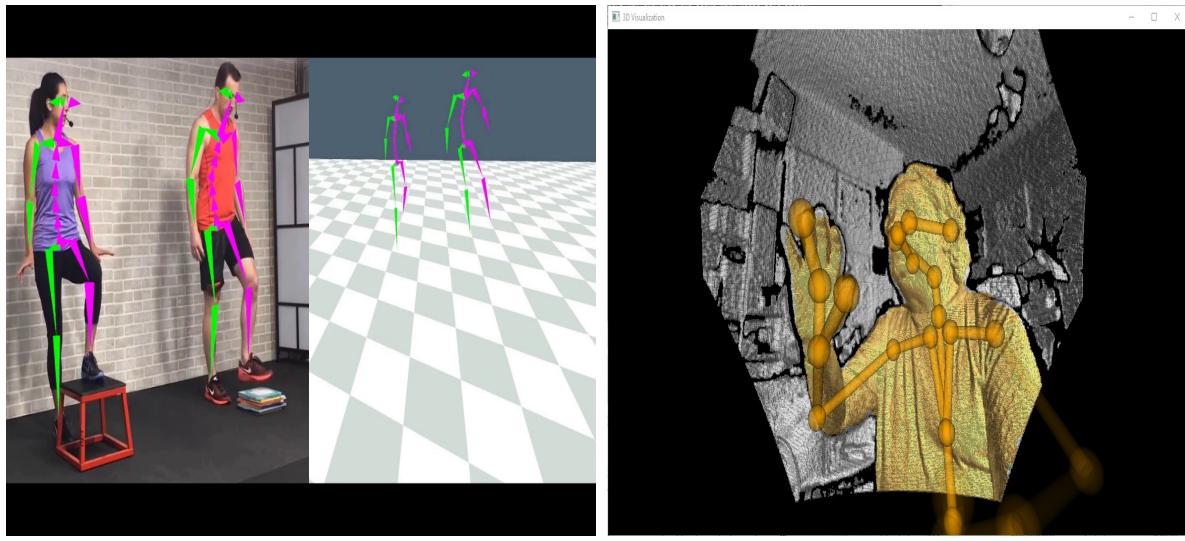
La estimación de las poses humanas representan una problemática compleja de solucionar, el cual tuvo un largo trayecto hasta salir a la luz. La complejidad se centra en las múltiples limitantes, como las mascotas, los objetos del área, las personas de alrededor, la variación del escenario, los parámetros del cuerpo (el tamaño, longitud de las extremidades, torso y otras partes del cuerpo) y la iluminación.

Empleando las herramientas de seguimiento del Esqueleto, sensores de profundidad y sensores RGB, a medida que el tiempo corre, la necesidad de herramientas como los sensores va volviéndose obsoleta con el nacer de herramientas como PoseNet y Open-Pose, que con el apoyo del Hardware mínimo necesario, son capaces de proporcionar la misma información con una calidad apenas inferior de seguimiento corporal.

## Estimación de Poses

La estimación de poses como su nombre indica, tiene la labor de estimar la pose de una persona a partir de modelos machine learning (ML) guardados en una base de datos, que proporcionan en ubicaciones específicas puntos clave que en su conjunto forman el esqueleto, además los datos sobre sus puntos se guardan como es explicado en el Análisis de Alternativas de PoseNet.

Se podría mencionar que su diferencia parte también de las palabras que lo conforman, Body Tracking/Capture Motion hace referencia a seguir el movimiento de una persona empleando Estimación de Pose, mientras que la Estimación de Poses hace referencia a marcar los puntos clave de una persona y construir un esqueleto de una persona [34].



(a) Ejemplo 1

(b) Ejemplo 2

Figure 3.2: Ejemplos de Body Tracking  
Fuente: [10] [12]

### 3.1.1 Skeletal Tracking

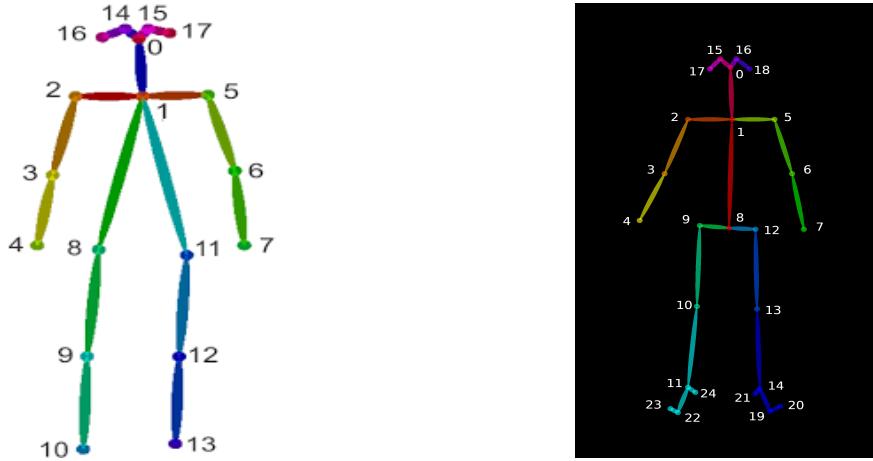
Este fue una innovación brindada por el controlador Kinect al mercado comercial, Su demanda fue elevada en la época y hasta el día de hoy sigue siendo empleado, el reconocimiento de una persona desde cualquier angulo o distancia, tomando en cuenta su figura, tamaño, color, cabello, ropa y el ambiente. Se emplea el escaneo de la imagen para reconocer puntos importantes del cuerpo que representan al cuerpo, tales como la cabeza, cuello, hombros, brazos, piernas y otros 10 a 20 puntos dependiendo la herramienta de reconocimiento que se emplee.

La herramienta seleccionada para el proyecto es OpenPose empleando este modelo de Formato de salida, para los datos del seguimiento del esqueleto es emplear la flag write\_json para guardar la información dentro un JSON, el cual contiene un objeto de esqueleto de persona dentro, que contiene un vector pose\_keypoints\_2d con puntos 2D para localizar y detectar cada punto de unión  $x_1, y_1, c_1, x_2, y_2, c_2, \dots$ . Las coordenadas x y y en el rango de  $[0,1], [-1,1]$ .

Además de la existencia de los vectores face\_keypoints\_2d, hand\_left\_keypoints\_2d y hand\_right\_keypoints\_2d, análogos a pose\_keypoints\_2d, los cuales debido a la masiva carga de memoria que requieren y la falta de ella (mínimo 4GB de Memoria dedicada estimada, contando solo con 2GB en el equipo proporcionado) serán ignorados, pero empleando pose\_keypoints\_2d.

Como dato, los vectores análogos body\_keypoints\_3d, face\_keypoints\_3d, hand\_left\_keypoints\_2d y hand\_right\_keypoints\_2d (si -3d flag se habilitase), en vez de  $x_1, y_1, c_1, x_2, y_2, c_2, \dots$ , el formato sería  $x_1, y_1, z_1, c_1, x_2, y_2, z_2, c_2, \dots$ , donde c sería 1 o 0 dependiendo si la reconstrucción 3D es exitosa.

Se empleara el modelo BODY\_25 de Caffe, para mostrar el esqueleto, que consiste en mostrar 25 puntos del esqueleto, cada uno con su conexión en los siguientes puntos



(a) Reconocimiento OpenPose Tipo COCO (b) Reconocimiento OpenPose Tipo BODY\_25

Figure 3.3: Diversos Tipos de Seguimiento al Esqueleto seleccionados por OpenPose para el desarrollo del proyecto

Fuente: OpenPose Models Image Output [15] [14]

clave:

En cuanto al resultado, este se puede guardar en formatos estándar (JSON, XML, PNG, JPG,...), existen suficientes herramientas de uso libre para leerlos, por tanto cargar los datos y cargar las imágenes, no debería representar un desafío.

## 3.2 Base Teórica

### 3.2.1 Cámara y Desambiguación

Una cámara se define como un dispositivo que permite el registro y reproducción de imágenes. A través del tiempo, se desarrollaron muchos tipos de cámara, por tanto es necesario determinar el tipo de cámara del cual se dispone para la elaboración del proyecto, siendo el selecto la cámara Web. Sin embargo, en la mayoría de proyectos relacionados al Body Tracking, se menciona el término Depth of Field (DOF), mención a una cámara de profundidad de campo.

La cámara web es un modelo pequeño de una cámara digital conectada a una computadora. Tiene la capacidad de capturar imágenes y transmitirlas a través de Internet. Un punto importante es que pueden ser empleadas para el desarrollo de aplicaciones y programas de cierta índole.

#### Depth of Field (DOF)

La profundidad de campo es el producto del deseo de producir una imagen más clara del objeto o escenario deseado. Normalmente, una cámara debe enfocar su lente para tener mayor precisión, por tanto, se adjuntaron formas para ajustar el lente. Actualmente, se desarrollaron sistemas automáticos para el ajuste del lente, los cuales incluyen la

Pos.	Punto Clave	Pos.	Punto Clave
0	Nariz	13	Rodilla Izquierda
1	Cuello	14	Tobillo Izquierdo
2	Hombro Derecho	15	Ojo Derecho
3	Codo Derecho	16	Ojo Izquierdo
4	Muñeca Derecha	17	Oreja Derecha
5	Hombro Izquierdo	18	Oreja
6	Codo Izquierdo	19	Dedo Gordo Izquierdo
7	Muñeca Izquierda	20	Dedo Menique Izquierdo
8	Cadera Central	21	Talón Izquierdo
9	Cadera Derecha	22	Dedo Gordo Derecho
10	Rodilla Derecha	23	Dedo Menique Derecho
11	Tobillo Derecho	24	Talón Derecho
12	Cadera Izquierda	25	Escenario

Table 3.1: OpenPose Output JSON Content, Which are Given X, Y and C values

Fuente:[14]

medición de la distancia al objetivo, empleando esta para ajustar el lente y obtener mayor claridad [32].

## Sensores RGB-D

Es el termino dado al conjunto de sensores RGB y Depth (de profundidad). Los sensores de RGB y profundidad son una adición a los dispositivos como Kinect, PlayStation Camera y otros.

El sensor de profundidad cumple la función de proyectar luz infrarroja y un sensor infrarrojo.

Divide su funcionalidad en dos pasos, la proyección de rayos de luz infrarroja, los cuales rebotan en el ambiente en el que se encuentra y posteriormente son detectados por el sensor de infrarrojo. Esta información es decodificada y enviada al usuario para poder emplearla como se lo estime.

RGB es un modelo de colores basado en la obtención de un color empleando una mezcla entre otros colores, todas ellas con los colores primarios, que son el rojo, verde y azul. El sensor RGB analiza el escenario que tiene en frente, determinando la cantidad de luz que requiere la producción de una imagen de calidad, optimiza y ajusta la velocidad de obturación, apertura y sensibilidad de captura de luz. Se encarga de analizar cada pixel del Frame y crear una imagen meticulosa.

En el desarrollo del proyecto, no se empleara un sensor de profundidad, ya que una cámara Web no lo tiene equipado.

La funcionalidad del sensor RGB-D es combinar la información colectada por el sensor infrarrojo y el sensor RGB, proporcionan datos de color y profundidad a cada pixel en cada Frame que analiza. Dentro de la memoria del Kinect, existen patrones de referencia, que poseen colores y profundidades ya conocidas, que facilitan la estimación

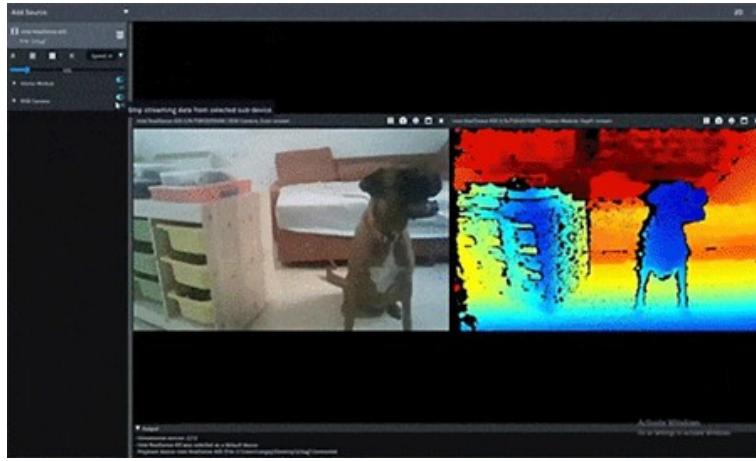


Figure 3.4: Ejemplo de Visión de una Cámara con sensores RGB-D  
Fuente: [9]

del escenario que lo necesite[31].

### 3.2.2 Inteligencia Artificial (IA)

Una faceta importante del proyecto, es la estimación de poses, elemento producto de la Inteligencia Artificial y Machine Learning, si bien, es explicado en la sección de PoseNet.

La inteligencia artificial (IA) es una rama de la ciencia de computación, que se define como ”Un programa que en un mundo arbitrario, no se las arregle peor que un humano”[20]. Esta definición se somete a muchos factores, para empezar requiere de asumir tres factores importantes. El primer factor es que cualquier dispositivo de calculo puede ser modelado por un programa. El segundo factor es que el IA es un dispositivo que permite ingresar información de afuera y dar una respuesta de acuerdo a ello. El tercer factor es asumir que la IA, al estar en contacto con el mundo este recibe información que da una respuesta interactiva con el mundo. Se asume que el mundo en que se encuentre será influenciado por la IA[20]. El mundo se considera como el ambiente con el que la IA tiene interacción, aquel que proporciona los valores que la IA procesará y aquel que reaccionará a las acciones y/o valores de salida de la IA, en un proceso de emisión y recepción continuo.

### Machine Learning (ML)

El aprendizaje de la máquina o Machine Learning es ua rama de la ciencia de computación dirigida en algoritmos Computacionales diseñada para emular la inteligencia humana, a través del aprendizaje del mundo que lo rodea.

Los modelos ML ganan popularidad debido a su capacidad de proporcionar resultados y conocimientos de un tipo específico en escenarios similares a los que ha entrenado, pero nunca antes visto. Los modelos ML provee información relevante sobre los datos relacionados a lo que se requiere, traduciendo los datos de entrada, en datos que se desean obtener. Por ejemplo, un médico, para diagnosticar a un paciente con un síntoma

tan común como fiebre, requerirá de un extenso conocimiento en enfermedades que tienen este síntoma y basándose en todo su conocimiento teórico, indagara en información que pueda relacionar con la fiebre, especificando otros problemas o falta de ellas que tenga y llegará a una conclusión sobre que enfermedad padece [33].

Una versión simplificada es decir que Machine Learning emplea algoritmos para clasificar y filtrar la información recibida, aprender de ella, guardarla y tomar una decisión basándose en lo aprendido.

## **Deep Learning (DL)**

El aprendizaje profundo o Deep Learning es una subrama de Machine Learning que se basa en el uso de Neural Network, empleando numerosas capas y nodos para el aprendizaje, que en conjunto a toda la base de datos previa da lugar las decisiones en las que se basa.

El Deep Learning utiliza múltiples capas de decisión, siendo el dato de ingreso interpretado por el logaritmo en diferentes niveles, cada uno repercuten en el anterior y produce un aprendizaje basado en intento y error para aprender, corrigiendo los datos de salida de cada capa para llegar a una respuesta cada vez más aproximada a la deseada. Una diferencia importante entre Machine Learning y Deep Learning, es que, Machine Learning en general mejora con el tiempo, pero requiere de corrección de vez en cuando, si se desvía de los resultados que se buscan. En cambio, Deep Learning determina si su propia predicción es correcta haciendo uso de su red neuronal, no requiere de intervención.

En el desarrollo del proyecto se hace mención al modelo CAFFE y COCO, los cuales son parte fundamental en el proyecto, ya que estas son los modelos de Deep Learning empleados para el Body Tracking.

## **Modelo DL CAFFE**

El modelo CAFFE es un Framework opensource trabajado con las librerías C++ y CUDA para Deep Learning, tiene interfaces en Command Line, Python y MathLab, incluye soporte en problemas con el modelo, tiene referencias, demos y herramientas para poder usarlo, además de la característica de usarse con CPU y GPU[26].

Además de poseer una comunidad abierta a ampliar el desarrollo y trabajar con la herramienta, ventaja característica del código open source. El modelo CAFFE ofrece el poder definir modelos, optimizar opciones para el desarrollo y pre-entrenar para poder realizar el aprendizaje.

En la sección del proyecto, CAFFE realizó sus estudios para poder guardar Dataset de miles de cuerpos humanos, otorgándole puntos clave para formar el esqueleto. La herramienta OpenPose selecta hace uso de la CMU Panoptic Dataset, que junto a sus miles de miles de esqueletos 3D permite estimar la pose humana, produciendo el Modelo BODY\_25 visto previamente en la figura 3.3a.

## Modelo DL COCO

EL modelo COCO es un Dataset reciente del año 2014, desarrollado para el reconocimiento de objetos en el contexto requerido, a través de la obtención de imágenes complejas que contengan la información requerida en su estado natural. En su mayoría, posee objetos reconocibles por niños pequeños, tiene 2.5 millones de etiquetas en mas de 300000 imágenes.

En el contexto, el modelo COCO posee miles de imágenes relacionadas a la identificación del cuerpo humano y la construcción de su esqueleto, marcando con etiquetas sus puntos claves del cuerpo y la confiabilidad de esos datos[30].

## Neural Network

Neural Network inicialmente es inspirado por la compleja red de neuronas del cuerpo humano, el cual transmite información, ordenes a través de impulsos eléctricos, donde miles de millones de conexiones existen, con ese fin[41].

Formalmente, el modelo de Neural Network consiste en un set de unidades Computacionales y un set de conexiones de una vía unidas. A través del ingreso de datos, las unidades Computacionales lo examinan y computan, generando una activación como dato de salida. La activación atraviesa la conexión que tiene la unidad Computacional con otra unidad y procede a examinarlo y computarlo para generar otra activación. Este proceso se repite, con el fin de aumentar el llamado peso que determina la influencia que se obtiene al examinar y computar el valor recibido, dependiendo el objetivo, mientras mayor o menor termine siendo el peso en las conexiones a medida que se computan las activaciones en las unidades Computacionales, mayor efecto tendrá en el dato de salida. Un ejemplo de ello se observa en la figura 3.5 [22], donde los círculos representan unidades Computacionales, las flechas las conexiones, los valores los pesos, ingreso de datos es el 1 y el dato de salida es 0. El 0.73, 0.79 y 0.69 es el resultado de una Activación después del primer computo.

Esta herramienta es empleada por OpenPose y PoseNet para el entrenamiento de las herramientas de estimación de pose, que lo utilizan para que con los Dataset de miles de imágenes que sirven como imágenes de ingreso y sus etiquetas, puedan ser analizadas por extensas capas de unidades Computacionales que evaluaran cada imagen y proporcionaran pesos, que a mayor peso, mayor precisión en la estimación de pose tendrá.

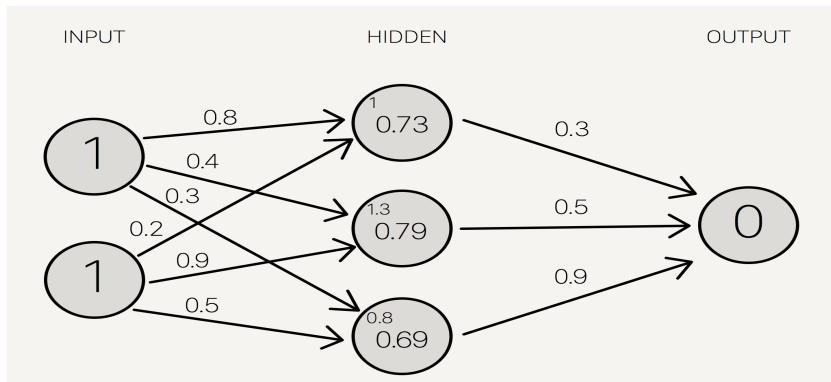


Figure 3.5: Ejemplo de Neural Network  
Fuente:[14]

# **Estudio de alternativas**

## **4.1 Análisis de Alternativas actualmente disponibles en el mercado**

En la actualidad, en la extensión de la investigación, existen productos comerciales y proyectos privados que se asemejan a las características ofrecidas por el proyecto, son en su mayoría de índole privatizada especialmente en el campo de la medicina, donde solo se hace mención a sus resultados, sin embargo, ninguna resalta por su accesibilidad gratuita, distinguiendo el proyecto del factor común. Siendo las aplicaciones y programas desarrollados con un enfoque principal en entretenimiento, como es el caso de videojuegos y medicina, en analizadores médicos y programas de rehabilitación.

### **Analizadores Médicos**

En la rama de la medicina, la información es más privatizada o proporciona un menor alcance al público, solo reporta resultados y posibles cambios y amplitudes afirmando la falta de pruebas para dar mayor información, por lo que profundizar en ello no otorgara profundidad al proyecto, no obstante, se ha encontrado que se utiliza la tecnología planeada para controlar la pose del usuario, ya sea en pruebas médicas para hallar anomalías físicas en la posición del cuerpo al realizar diversas actividades y en entrenamiento físico para mantener una posición estable y evitar dañarse a uno mismo.

#### **4.1.1 Entrenamiento**

Cuando se referencia a un sistema de estimación de pose del cuerpo, se nombran varios títulos de videojuegos de baile, siendo las primeras y más importantes sagas, las desarrolladas por Ubisoft como Just Dance o Dance Experience, enfocadas en realizar coreografías pre-diseñadas para canciones populares en la época del lanzamiento de sus entregas; a pesar de que en Just Dance han existido niveles con temática de artes marciales, estas no contaban con la intención de ser un entrenamiento educativo, sino un baile con el propósito de entretener utilizando movimientos basados del arte marcial.

Además de las mencionadas, existe otro género en los videojuegos enfocado al entrenamiento físico, como Shape Up de Ubisoft y Wii Fit de Nintendo, estas enfocadas

más a un entrenamiento físico casual como es el caso de Shape Up, donde se incentiva al jugador a realizar ejercicios anaeróbicos de alta intensidad como son flexiones, abdominales o sentadillas y Wii Fit enfocado a rutinas de Yoga, equilibrio y aeróbicos; demostrando las diferentes aplicaciones y posibilidades de desarrollo y resultados.

### **Just Dance**

Just dance es una serie de sistemas interactivos con Body Tracking, es un producto pionero en el tema desarrollado por Ubisoft París que debutó en octubre del 2010. La serie de Just Dance cuenta con un total de 30 títulos con un amplio rango de categorías de edades para llegar a la audiencia, es ya una serie que libera un nuevo título al menos una vez al año, demostrando ser un éxito comercial. Gran parte de la inspiración del proyecto actual es este sistema interactivo, que emplea al máximo el controlador Kinect y el Body Tracking con comandos de voz para ofrecer un uso casual y entretenido, sin embargo, la carencia del equipo y la frustración de no ser capaces de aprovecharlo, se desvían en cambio a la posibilidad de crear una alternativa, el cual se espera que con el equipamiento provisto para el proyecto, carente de una cámara de profundidad (elemento sobresaliente en el controlador Kinect), permita el desarrollo adecuado de las características de Body Tracking.

## **4.2 Análisis de Alternativas En Herramientas Disponibles Para el Desarrollo**

### **4.2.1 Trajes con sensores de movimiento**

La industria actual de cine, sistemas interactivos, deportes e incluso la medicina cuentan con una inversión que llega a niveles millonarios en recrear animaciones del cuerpo humano, empleando las herramientas externas que faciliten la lectura de movimiento, su uso es bastante extendido y llega a crear obras de gran calidad, sin embargo, en su mayoría esta tecnología no está disponible para el público en general debido a su elevado costo, y ya que no es un campo prioritario de investigación en la Universidad Privada Boliviana, por tanto, esta alternativa queda descartada.

Los trajes cuentan con 19 sensores distribuidos alrededor del cuerpo y recogen movimientos de un actor para transmitirlos en tiempo real. Una de sus ventajas es la reducción de costos en las producciones para motion capture, además de no requerir un set de estudio demasiado restrictivo para su implementación. Existen además, varios modelos del traje proporcionados por empresas como Xsens, Holosuit, Teslasuit y otros.

### **Creación de Software propio para Body Tracking**

Uno de los objetivos implícitos del desarrollo de este proyecto es el siguiente "El objetivo no es volver a crear la rueda, sino, crear algo nuevo utilizando esa rueda", palabras del

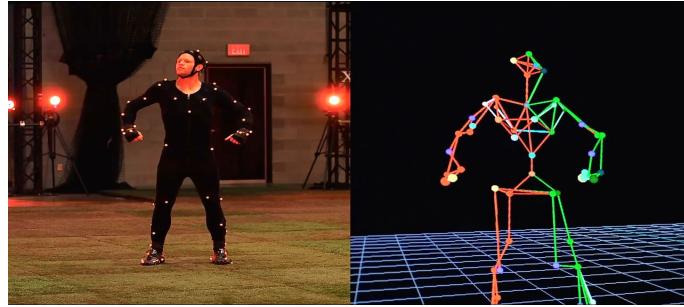


Figure 4.1: Traje con Sensores de Movimiento  
Fuente: [5]

supervisor del proyecto, por tanto, a pesar de ser un posible acercamiento a la solución, queda completamente descartada.

#### 4.2.2 TensorFlow

Tensorflow es una herramienta open source para Machine Learning provista de soporte por Microsoft, la comunidad la ha empleado extensamente en proyectos y estudios de múltiples campos de la inteligencia artificial, de Body Tracking con el uso de PoseNet, derivado de TensorFlow, cuenta con soporte y una documentación clara, siendo además sus requisitos recomendados para su uso relativamente bajos.

TensorFlow se basa en redes neuronales para Deep Learning, es una herramienta interfaz que implementa algoritmos para Machine Learning, puede ser ejecutada con una variedad de sistemas, desde celulares hasta Tablets, e incluso sistemas a gran escala. Tiene un amplio rango de algoritmos de entrenamiento, incluyendo entrenamiento e modelos de inferencia para Neural Network Models e investiga ramas de múltiples ramas, incluyendo recolección de información, proceso de entendimiento de un lenguaje natural, geografía, descubrimiento de drogas, ciencias Computacionales y otros. [13]

#### 4.2.3 PoseNet

PoseNet es una herramienta desarrollado por Google Creative Lab basada en TensorFlow que permiten demostrar una estimación a tiempo real de estimación de poses (Body Tracking) en tiempo real. Esta herramienta puede ser empleada tanto para una persona a la vez como para varias personas a la vez dependiendo del algoritmo que se emplee, la diferencia es en la velocidad y simpleza en su función [27].

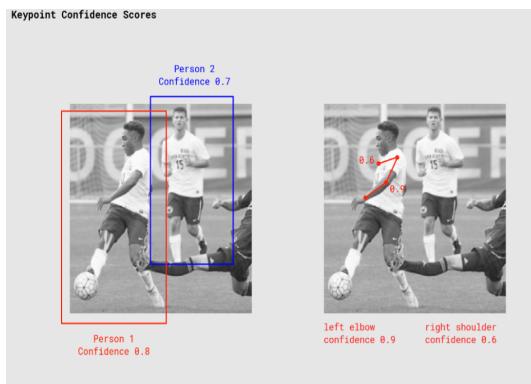
Debido a que el proyecto será para un uso singular, se mencionara su posibilidad a través del algoritmo de Body Tracking para una persona.

Emplea una imagen RGB que alimenta a la red neuronal y emplean un decodificador de poses, designando valores de confianza, posiciona puntos clave y valores de confianza para los puntos clave para el aprendizaje de la red neuronal en la lectura de imágenes con las cuales aprendió a estimar las posiciones en tiempo real[34].

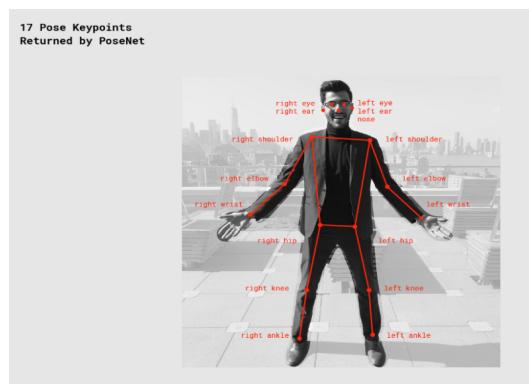
En cuanto los términos previos, son propios de la estimación de pose, por tanto se los debe explicar y mostrar visualmente para su entendimiento, los cuales se denotan en la

figura 4.2:

- Pose: Retorna un objeto que contiene una lista de puntos clave y valores de confidencia para cada persona detectada.
- Valor de confianza de pose: Determina la confianza que tiene la estimación de la pose, en un rango de 0 a 1 y puede ser usado en caso que las partes del cuerpo bloqueadas por el cuerpo mismo o elementos externos represente un obstáculo para su lectura exacta.
- Punto Clave: Son literalmente un grupo de puntos, es la parte que se estima del cuerpo humano para formar el esqueleto, como se observa claramente en la figura 3.3b. En la base de datos de modelo COCO, existen 17 puntos de lectura.
- Valor de Confianza de un punto clave: Determina en un rango de 0 al 1 la confianza que se tiene de la precisión del punto determinado, bajo el mismo concepto de la pose, en caso que existan obstáculos.
- La posición del punto clave: Las coordenadas (x,y) de la imagen en las que se encuentran los puntos claves.



(a) Ejemplo a



(b) Ejemplo b

Figure 4.2: Ejemplos de Datos en la Base de Datos COCO para la estimación de pose  
Fuente: [27] [34]

Si bien esta es una herramienta con soporte constante por parte de Google y una documentación clara y amplia, la herramienta requiere de características altas por parte de los equipos empleados para usarlo, su instalación y uso no es nada sencillo y es imposible cubrir todos sus requisitos en la medida de lo posible con el corto tiempo dado por la materia, concluyendo que se requerirá de buscar otras opciones para el proyecto.

#### 4.2.4 Wrnch

<https://wrnch.ai/>) Wrnch es una herramienta de calidad para el desarrollo de aplicaciones con Body Tracking, posee un gran potencial para el desarrollo del proyecto,

contando con el esqueleto que se forma al seguir los movimientos de la persona, además cuenta con una opción de multiples cámaras en distintos angulos, capaz de seguir el movimiento de los dedos al mismo tiempo que el cuerpo completo casi en tiempo real. Como debilidad, la dependencia del Hardware y sus elevados requisitos para emplear al máximo esta herramienta con el equipamiento disponible, limitó sus posibilidades y uso en el proyecto.

#### **4.2.5 OpenPose**

OpenPose es una herramienta open source desarrollada y mantenido principalmente por Gines Hidalgo, junto a su equipo de 6 personas y el apoyo de CMW Panoptic Studio Dataset. Es una herramienta principalmente programada en C++, empleando Cuda, Cmake y Shell, parte de ello para su instalación por parte de terceros que deseen desarrollar Software a partir de esta herramienta. Incluye APIs para desarrollo en Python y C++, posee un plugin para Unity desarrollado en el 2018, que a pesar de estar desactualizado, posee los mínimos requerimientos necesarios para el desarrollo del proyecto[42][15][40][14]. Es una herramienta similar a PoseNet en cuanto a los requerimientos del proyecto, es constantemente actualizado y permite incluso el reconocimiento de puntos clave del rostro y las manos, que si bien, merece ser mencionado, no se empleara a lo largo del proyecto.

Esta es la herramienta seleccionada para el desarrollo del proyecto, por tanto, fue previamente explicado y mencionado en el Marco Teórico, en la sección de Capture Motion y Skeletal Tracking, para más información, favor de revisarlo nuevamente.

#### **4.2.6 DeepMotion**

DeepMotion es una herramienta de body tracking proporcionada por la compañía DeepMotion, la cual, empleando machine learning, proveen de una solución para crear animaciones de modelos 3D con recursos mínimos en requerimientos de Hardware y experiencia requerida. Si bien, esta herramienta es rápidamente descartada debido a su alto coste y la lentitud de respuesta por parte del equipo en proporcionar el conocimiento para su adquisición, el soporte que se espera es amplio.

#### **4.2.7 Unity y Kinect**

El Kinect SDK provee a los desarrolladores de herramientas para el reconocimiento de voz empleando el uso de los sensores del dispositivo Kinect con el que funciona, además de una lectura de profundidad e infrarrojo para facilitar la motion capture. Al emplear este recurso y utilizar Unity para poder desarrollar tanto los modelos necesarios para el proyecto como la interfaz de la aplicación, se podría llegar a desarrollar el proyecto sin contratiempos e incluso con más funciones.

Sin embargo, hoy en día, el dispositivo Kinect fue descontinuado y las consolas a las que se conectan ya no incluyen el adaptador para el dispositivo. Además, se recuerda que la falta del Kinect y la búsqueda de una solución alternativa a su uso obligan a rechazar la consideración de emplear esta posible solución.

# Plan de Actividades

El desarrollo del proyecto inicia oficialmente el día 17 de Septiembre, con la definición del perfil del proyecto por parte del grupo de estudiantes.

El desarrollo fue interrumpido debido a la concentración del equipo en otras actividades, que exigió el 100% del tiempo libre y de estudio y progreso en el proyecto, siendo más específicos la materia de Aplicación con Redes, estimando entre 4 a 8 horas diarias (incluyendo fines de semana) de exigencia entre estudio y prácticas para mantener el ritmo solo a esa materia. Debido a ello, desde la fecha 18 de septiembre del 2020, a la fecha 4 de noviembre del 2020, el progreso fue mínimo, reduciéndose a la búsqueda de API's y herramientas, así como la redacción base del proyecto y la oficialización interna del producto deseado.

La metodología SCRUM fue inicializada oficialmente y de manera constante el día 10 de Noviembre del 2020, planteando un rango de horas de trabajo de 2 a 5 horas diarias (incluyendo fines de semana), con la finalización del Sprint los sábados por la mañana, a menos que sea una entrega de avance, la cual requerirá de una finalización del Sprint hasta donde se encuentra el proyecto en ese momento.

## 5.0.1 Cronograma

	September		October						November					December					
Week	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14					
3	4	5	6	0	1	2	3	4	5	6	0	1	2	3	4	5			
0	1	2	3	4	5	6	0	1	2	3	4	5	6	0	1	2	3	4	5
T1																			
T2																			
T3																			
T4																			
T5																			
T6																			
T7																			

Cada TX representa un título

- T1 *Redacción del Perfil.*
- T2 *Investigación.*
- T3 *Redacción del documento.*
- T4 *Implementación de la aplicación.*
- T5 *Revisión final del documento.*
- T6 *Revisión final de la aplicación.*
- T7 *Redacción de la presentación.*

# Bibliography

- [1] Esrb game ratings. <https://www.esrb.org/>, 1994.
- [2] Las etiquetas de edad pegi. <https://pegi.info/es>, 2003.
- [3] The eye of judgment (ps3). <https://forum.gamesaktuell.de/blogs/imperatcore/361571-augen-zu-und-durch-prev.html>, 2007.
- [4] Holiday gaming gift guide: Xbox 360 kinect. <https://owtk.com/2011/11/holiday-gaming-gife-guide-xbox-360-kinect/>, 2011.
- [5] The last of us - motion capture. <https://www.youtube.com/watch?v=9TxJWY19xxk>, 2013.
- [6] Crearon una tÉcnica de rehabilitaciÓn virtual para pacientes con discapacidad. <https://www.elfederal.com.ar/crearon-una-tecnica-de-rehabilitacion-virtual-para-pacientes-con-discapacidad/>, 2016.
- [7] Playstation camera. <https://www.playstation.com/en-us/accessories/playstation-camera/>, 2016.
- [8] Sony eye-camera, ps3 cámara web 640 x 480 pixeles usb 2.0. <https://icecat.biz/es/p/sony/9473459/webcams-eye-camera-+ps3-1269549.html>, 2016.
- [9] Intel lanza dos cámaras realsense con sensor de profundidad: visión en 3d para cualquier dispositivo. <https://www.xataka.com/realidad-virtual-aumentada/intel-lanza-dos-camaras-realsense-con-sensor-de-profundidad-vision-en-3d-para-cualquier-dispositivo>, 2018.
- [10] [visual body tracking]. <https://www.youtube.com/watch?v=o1WRrjs0m8o&app=desktop>, 2018.
- [11] Azure kinect dk. <https://developer.microsoft.com/en-us/mixed-reality/>, 2020.
- [12] Working with the azure kinect developer kit. <https://www.infoworld.com/article/3562738/working-with-the-azure-kinect-developer-kit.html>, 2020.

- [13] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [14] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Real-time multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [15] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.
- [16] X. Chang, Z. Ma, M. Lin, Y. Yang, and A. G. Hauptmann. Feature interaction augmented sparse learning for fast kinect motion detection. *IEEE Transactions on Image Processing*, 26(8):3911–3920, 2017.
- [17] N. George H. Peter A. Barbara L. Fuzhong Chen P., L. Mao. Coronavirus disease (covid-19): The need to maintain regular physical activity while taking precautions. *Journal of sport and health science*, 9(2):103–104, February 2020.
- [18] Christensson. Hardware definition. <https://techterms.com/definition/hardware>, 2020.
- [19] Andy J Daly-Smith, Stephen Zwolinsky, Jim McKenna, Phillip D Tomporowski, Margaret Anne Defeyter, and Andrew Manley. Systematic review of acute physically active learning and classroom movement breaks on children’s physical activity, cognition, academic performance and classroom behaviour: understanding critical design features. *BMJ open sport & exercise medicine*, 4(1), 2018.
- [20] Dimiter Dobrev. A definition of artificial intelligence. *arXiv preprint arXiv:1210.1568*, 2012.
- [21] Dominic. Kinect como profesor de educación física: Vicent gadea, profesores innovadores. <https://tecnousuariostv.wordpress.com/2016/01/13/kinect-como-profesor-de-educacion-fisica-vicent-gadea-profesores-innovadores/>, 2016.
- [22] Stephen I Gallant and Stephen I Gallant. *Neural network learning and expert systems*. MIT press, 1993.
- [23] Julen Zaballa García. Euskadi será el escenario de la presentación en primicia de kinect. <https://www.videojuegosvascos.com/eventos-de-videojuegos/euskadi-primicia-kinect/>, 2010.
- [24] Daniel Grest, Jan Woetzel, and Reinhard Koch. Nonlinear body pose estimation from depth images. In *Joint Pattern Recognition Symposium*, pages 285–292. Springer, 2005.

- [25] Nathan B Herz, Shyamal H Mehta, Kapil D Sethi, Paula Jackson, Patricia Hall, and John C Morgan. Nintendo wii rehabilitation (“wii-hab”) provides benefits in parkinson’s disease. *Parkinsonism & related disorders*, 19(11):1039–1042, 2013.
- [26] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.
- [27] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015.
- [28] Belinda Lange, A Rizzo, Chien-Yen Chang, Evan A Suma, and Mark Bolas. Markerless full body tracking: Depth-sensing technology within virtual environments. In *Interservice/industry training, simulation, and education conference (I/ITSEC)*, 2011.
- [29] Belinda Lange, Evan A Suma, Brad Newman, Thai Phan, Chien-Yen Chang, Albert Rizzo, and Mark Bolas. Leveraging unencumbered full body control of animated virtual characters for game-based rehabilitation. In *International Conference on Virtual and Mixed Reality*, pages 243–252. Springer, 2011.
- [30] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [31] Krystof Litomisky. Consumer rgb-d cameras and their applications. *Rapport technique, University of California*, 20, 2012.
- [32] Ricky J Madsen, Jon A Holmes, and Trygve D Peterson. Depth-of-field indicator for a camera, June 27 2000. US Patent 6,081,670.
- [33] W James Murdoch, Chandan Singh, Karl Kumbier, Reza Abbasi-Asl, and Bin Yu. Interpretable machine learning: definitions, methods, and applications. *arXiv preprint arXiv:1901.04592*, 2019.
- [34] D Oved, I Alvarado, and A Gallo. Real-time human pose estimation in the browser with tensorflow. js. *TensorFlow Medium, May*, 2018.
- [35] Bruce Pirie. Meaning through motion: Kinesthetic english. *The English Journal*, 84(8):46–51, 1995.
- [36] Gaetano Raiola, Pio Alfredo Di Tore, et al. Motor learning in sports science: Different theoretical frameworks for different teaching methods. *Sport Science*, 10(S1):50–56, 2017.

- [37] Debbie Rand, Rachel Kizony, and Patrice Tamar L Weiss. The sony playstation ii eyetoy: low-cost virtual reality for use in rehabilitation. *Journal of neurologic physical therapy*, 32(4):155–163, 2008.
- [38] Pulkit Sharma. Image classification vs. object detection vs. image segmentation. <https://medium.com/analytics-vidhya/image-classification-vs-object-detection-vs-image-segmentation-f36db85fe81>, 2019.
- [39] Matheen Siddiqui and Gérard Medioni. Human pose estimation from a single view point, real-time range sensor. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 1–8. IEEE, 2010.
- [40] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.
- [41] Sun-Chong Wang. Artificial neural network. In *Interdisciplinary computing in java programming*, pages 81–100. Springer, 2003.
- [42] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016.