
Análisis de captura de movimiento para
inferencia de la comunicación no verbal
Motion capture analysis for inference of
non-verbal communication



Trabajo de Fin de Grado
Curso 2024–2025

Autor

Alejandro Barrachina Argudo, Pablo Sánchez Martín

Director

Alejandro Romero Hernández, Ismael Sagredo Olivenza

Colaborador

Grado en Ingeniería Informática y Grado en Desarrollo
de Videojuegos

Facultad de Informática

Universidad Complutense de Madrid

Análisis de captura de movimiento para
inferencia de la comunicación no verbal
Motion capture analysis for inference of
non-verbal communication

Trabajo de Fin de Grado en Ingeniería Informática y Grado
en Desarrollo de Videojuegos

Autor

Alejandro Barrachina Argudo, Pablo Sánchez Martín

Director

Alejandro Romero Hernández, Ismael Sagredo Olivenza

Colaborador

Convocatoria: *Junio 2025*

Grado en Ingeniería Informática y Grado en Desarrollo de
Videojuegos

Facultad de Informática

Universidad Complutense de Madrid

2 de mayo de 2025

Dedicatoria

*A Pedro Pablo y Marco Antonio, por crear
TeXiS e iluminar nuestro camino*

Agradecimientos

A Guillermo, por el tiempo empleado en hacer estas plantillas. A Adrián, Enrique y Nacho, por sus comentarios para mejorar lo que hicimos. Y a Narciso, a quien no le ha hecho falta el Anillo Único para coordinarnos a todos.

Resumen

Análisis de captura de movimiento para inferencia de la comunicación no verbal

Un resumen en castellano de media página, incluyendo el título en castellano. A continuación, se escribirá una lista de no más de 10 palabras clave.

Palabras clave

Máximo 10 palabras clave separadas por comas

Abstract

Motion capture analysis for inference of non-verbal communication

An abstract in English, half a page long, including the title in English. Below, a list with no more than 10 keywords.

Keywords

10 keywords max., separated by commas.

Índice

1. Introducción	1
1.1. Motivación	2
1.2. Objetivos	2
1.3. Plan de trabajo	2
2. Estado de la Cuestión	3
2.0.1. Perception Neuron 3	4
2.0.2. Meta Quest	4
2.0.3. Unity	4
2.0.4. TensorFlow	4
2.0.5. Random Forest	4
2.0.6. Recurrent Neural Network	4
2.0.7. Long Short-Term Memory	4
2.0.8. Convolutional Neural Network	4
3. Descripción del Trabajo	5
3.1. Traje	5
3.2. Búsqueda de un dataset	5
3.3. Dataset Artificial	5
3.4. Dataset real	6
3.5. Modelos de Inteligencia Artificial (IA)	7
3.6. Aplicación final	7
3.6.1. Funcionamiento de la aplicación	7
4. Conclusiones y Trabajo Futuro	9
Introduction	11
4.1. Motivation	12
4.2. Objectives	12
4.3. Work Plan	12
Conclusions and Future Work	13

Contribuciones Personales	15
Bibliografía	17
A. Título del Apéndice A	19
B. Título del Apéndice B	21
Glosario	23
Glosario	23

Índice de figuras

Índice de tablas

Capítulo 1

Introducción

*“La revolución industrial y sus consecuencias han sido un
desastre para la raza humana”*
— Theodore Kaczynski

En los últimos años se ha visto un gran interés y evolución de las tecnologías de realidad virtual a mano de empresas como Apple con su lanzamiento de las Apple Glasses o Meta con el lanzamiento de las Meta Quest 3 o su interés por el metaverso con su aplicación de Meta Horizon Worlds.

Es por ello que es interesante investigar la comunicación no verbal en entornos virtuales, ya no solo para aplicaciones con varios usuarios si no también para poder mejorar interacciones con Non Playable Characters (NPCs) en este tipo de entornos.

Nuestra propuesta en este Trabajo de Fin de Grado es una primera aproximación de cómo se puede usar la IA y la tecnología de captura de movimiento para lograr esa mejora en las interacciones en los mundos virtuales mediante la comunicación no verbal, siendo este caso la exploración de la capacidad de clasificar distintos gestos. Los gestos seleccionados han sido seis gestos que se han útiles a la hora de que un NPC pueda reconocerlo en un videojuego. Estos gestos han sido:

1. Bailar
2. Saludar
3. Señalar
4. Sentarse
5. Pelear
6. Correr

Para poder realizar este trabajo se ha requerido usar las gafas de realidad virtual Meta (antes conocidas como Oculus) Quest 2 y 3 y el traje de captura de movimiento Perception Neuron 3, de la empresa Noitom, como hardware y Python C# y Unity como los requisitos de software.

1.1. Motivación

Estudio del reconocimiento de gestos mediante IA para posibles implementaciones en el estudio y mejora de la comunicación no verbal en entornos virtuales.

1.2. Objetivos

Implementación de un modelo de IA que, con baja latencia, permita identificar el gesto que se esté realizando con un traje de captura de movimiento.

1.3. Plan de trabajo

Nuestro plan de trabajo consiste en varios pasos:

1. Búsqueda de un dataset: generar un dataset lo suficientemente grande con varios ejemplos de gestos como para poder entrenar de forma adecuada diferentes modelos.
2. Implementación de modelos de IA: implementación de varios modelos de IA para poder hacer una comparativa entre ellos y decidir cuál es el más adecuado teniendo en cuenta su velocidad de predicción y su precisión.
3. Desarrollo de una aplicación final: desarrollo de una aplicación para las Oculus Quest a forma de demo en la que se conecte al modelo elegido y se pueda ver en tiempo real su uso.

Capítulo 2

Estado de la Cuestión

En el estado de la cuestión es donde aparecen gran parte de las referencias bibliográficas del trabajo. Una de las formas más cómodas de gestionar la bibliografía en \LaTeX es utilizando **bibtex**. Las entradas bibliográficas deben estar en un fichero con extensión *.bib* (con esta plantilla se proporciona el fichero *biblio.bib*, donde están las entradas referenciadas más abajo). Cada entrada bibliográfica tiene una clave que permite referenciarla desde cualquier parte del texto con los siguiente comandos:

- Referencia bibliografica con cite: Bautista et al. (1998)
- Referencia bibliográfica con citep: (Oetiker et al., 1996)
- Referencia bibliográfica con citet: Krishnan (2003)

Es posible citar más de una fuente, como por ejemplo (Mittelbach et al., 2004; Lamport, 1994; Knuth, 1986)

Después, \LaTeX se ocupa de rellenar la sección de bibliografía con las entradas **que hayan sido citadas** (es decir, no con todas las entradas que hay en el *.bib*, sino sólo con aquellas que se hayan citado en alguna parte del texto).

Bibtex es un programa separado de latex, pdflatex o cualquier otra cosa que se use para compilar los *.tex*, de manera que para que se rellene correctamente la sección de bibliografía es necesario compilar primero el trabajo (a veces es necesario compilarlo dos veces), compilar después con bibtex, y volver a compilar otra vez el trabajo (de nuevo, puede ser necesario compilarlo dos veces).

- 2.0.1. Perception Neuron 3
- 2.0.2. Meta Quest
- 2.0.3. Unity
- 2.0.4. TensorFlow
- 2.0.5. Random Forest
- 2.0.6. Recurrent Neural Network
- 2.0.7. Long Short-Term Memory
- 2.0.8. Convolutional Neural Network

Capítulo 3

Descripción del Trabajo

3.1. Traje

3.2. Búsqueda de un dataset

Para poder entrenar modelos de IA se requiere una gran cantidad de datos, en este caso de animaciones de los gestos escogidos que se han mencionado en la introducción. Además, estos datos tienen que estar en un formato en el que se puedan extraer la información de los huesos (posición y rotación) para que pueda ser usado por los modelos, y tenían que tener los mismos huesos que el traje de captura de movimiento.

TO DO: HABLAR DE PÁGINA WEB DEL SEÑOR CON SU PROPIO TRAJE

TO DO: HABLAR DE LA ESCASEZ DE DATASET DE ANIMACIONES EN KAGGLE

TO DO: HABLAR DE POR QUÉ DESCARTAMOS IMÁGENES Y NOS CENTRAMOS MÁS EN ESQUELETOS

Ya que no se encontró un gran dataset que cumpliera con nuestros requerimientos se tomó la decisión de buscar en un banco de animaciones los gestos requeridos y transformar esas animaciones en un formato que pudiesen ser procesados.

3.3. Dataset Artificial

- MIXAMO COMO BANCO DE ANIMACIONES
- HERRAMIENTA PARA SACAR ANIMACIONES DE MIXAMO MEDIANTE UNA PALABRA CLAVE
- FALLO CON LOS ASSETS BUNDLES (AUNQUE SEA COMENTARLO)
- HERRAMIENTA DE UNITY PARA GENERAR EL ANIMATOR EN TIEMPO REAL (ALGUNA CAPTURA DEL TIME LAPSE CON REMY)
- CÓMO GUARDA LOS DATOS ESA MISMA HERRAMIENTA
- RESULTADOS FINALES OBTENIDOS

3.4. Dataset real

Debido a que las animaciones encontradas en el banco de animaciones no eran las suficientes y estaban descompensadas (TO DO: METER GRÁFICA DE LA CANTIDAD DEL DATASET Y QUE SE VEA LA DESCOMPENSACIÓN) se decidió crear una herramienta para poder recoger los datos de usuarios.

(TO DO: HABLAR EN PROFUNDIDAD DE LA HERRAMIENTA. UN UML ESTARÍA BIEN)

Esta recogida de datos consistió en ponerle el traje de captura de movimiento a los usuarios y pedirles que realizasen tres tomas de todos los gestos, a excepción del gesto de baile, ya que de ese gesto teníamos basantes más datos que el resto.

Para ello se creó un formulario en el que usuarios interesados en ello se apuntasen. En el formulario se explicaba el objetivo de la prueba y se numeraban los datos que iban a ser pedidos en la prueba. Los campos a rellenar eran:

1. Correo
2. Aceptar recogida de datos
3. Posible hueco en un horario, en huecos de una hora de lunes a viernes y de 9:00 a 20:00

Una vez creado el formulario se creó un cartel (TO DO: PONER FOTO DEL CARTEL) el cual se colgó en redes y se colgó en distintas facultades (NO SÉ SI PROCEDE COMENTARLO PERO BUENO), teniendo como resultado que se apuntasen 75 personas en el formulario.

Lo siguiente era tener un sitio en el que hacer las pruebas. Debido a que las pruebas requerían movimiento era preciso un lugar en el que se pudiesen mover sin dificultades y que no estuviese a la vista para preservar la intimidad de los usuarios.

TO DO: HABLAR DE LA RAMIFICACIÓN Y PODA HECHA PARA CONSEGUIR LOS HUECOS

Finalmente desde el día 14/04/2025 a las 9:00 hasta el día 17/04/2025 a las 19:30 se pudo grabar en el despacho (NO ME ACUERDO DEL NÚMERO) de la Facultad de Informática de la Universidad Complutense de Madrid. De los 75 usuarios apuntados finalmente se presentaron 69, consiguiendo así 975 animaciones en formato CSV, 195 de cada gesto. (TO DO: PONER FOTOS DE LAS PRUEBAS Y LOS USUARIOS)

TO DO: HABLAR DE LOS DATOS PEDIDOS DE LOS USUARIOS Y SUS RESULTADOS, SUS ESTADÍSTICAS Y TODO ESO

TO DO: HABLAR DE LAS ANIMACIONES YA PROCESADAS EN QUÉ SE QUEDA Y DAR PASO A LA PARTE DE LOS MODELOS

3.5. Modelos de IA

3.6. Aplicación final

Una vez los modelos estaban preparados era necesaria una aplicación a modo de demostración que se pueda ejecutar en las gafas de VR y con el traje de caputra de movimiento. La aplicación consta de un panel en el que introduces la IP de la máquina que tiene el servidor de Tensor ejecutándose con el modelo y la aplicación de Axis Studio para poder utilizar el traje. Además, tiene un panel de texto en el que aparecen los resultados de la predicción y un NPC que reacciona a tus gestos.

TO DO: CAPTURA DE LA APLICACIÓN

3.6.1. Funcionamiento de la aplicación

- CORRUTINA
- API REST Y HTTP PARA MANDAR Y RECIBIR LA INFO
- CON LO QUE DEVUELVE SE CAMBIA EL TEXTO Y EL NPC REACCIONA CAMBIANDO LA ANIMACIÓN
- EN CASO DE FALLO SE PUEDE REINTENTAR / SALE DE LA APLICACIÓN

Capítulo 4

Conclusiones y Trabajo Futuro

Conclusiones del trabajo y líneas de trabajo futuro.

Antes de la entrega de actas de cada convocatoria, en el plazo que se indica en el calendario de los trabajos de fin de grado, el estudiante entregará en el Campus Virtual la versión final de la memoria en PDF.

Introduction

*“La revolución industrial y sus consecuencias han sido un
desastre para la raza humana”*
— Theodore Kaczynski

In recent years, there has been a great interest and evolution of virtual reality technologies led by companies such as Apple with the launch of Apple Glasses or Meta with the launch of Meta Quest 3 or its interest in the metaverse with its application “Meta Horizon Worlds”.

It is for this reason that it is interesting to investigate non-verbal communication in virtual environments, not only for applications with multiple users but also to improve interactions with NPCs in these types of environments.

The proposal in this Bachelor’s Thesis is a first approach to how we can use Artificial Intelligence and motion capture technology to achieve this improvement in interactions in virtual worlds through non-verbal communication, being our case the exploration of the ability to classify different gestures. The gestures we have focused on are six gestures that we have considered useful for an NPC to recognize in a video game. These gestures are:

1. Dance
2. Greet
3. Point
4. Sit down
5. Fight
6. Run

To carry out this work, we have used the Meta (formerly Oculus) Quest 2 and 3 virtual reality glasses and the Perception Neuron 3 motion capture suit, from the company Noitom, as hardware and Python, C# and Unity as the software requirements.

4.1. Motivation

Research on gesture recognition using Artificial Intelligence (AI) for possible implementations in the study and improvement of non-verbal communication in virtual environments.

4.2. Objectives

Implementation of an AI model that, with low latency, allows identifying the gesture being performed with a motion capture suit.

4.3. Work Plan

The Work Plan consists of several steps:

1. Search for a dataset: generate a dataset large enough with several examples of gestures to adequately train different models.
2. Implementation of AI models: implementation of several AI models to make a comparison between them and decide which one is the most suitable considering its prediction speed and accuracy.
3. Development of a final application: development of an application for the Meta Quest as a demo that connects to the chosen model and allows real-time usage.

Conclusions and Future Work

Conclusions and future lines of work. This chapter contains the translation of Chapter 4.

Contribuciones Personales

En caso de trabajos no unipersonales, cada participante indicará en la memoria su contribución al proyecto con una extensión de al menos dos páginas por cada uno de los participantes.

En caso de trabajo unipersonal, elimina esta página en el fichero `TFGTeXiS.tex` (comenta o borra la línea `\include{Capitulos/ContribucionesPersonales}`).

Estudiante 1

Al menos dos páginas con las contribuciones del estudiante 1.

Estudiante 2

Al menos dos páginas con las contribuciones del estudiante 2. En caso de que haya más estudiantes, copia y pega una de estas secciones.

Bibliografía

*Y así, del mucho leer y del poco dormir, se
le secó el cerebro de manera que vino a
perder el juicio.*
(modificar en Cascaras\bibliografia.tex)

Miguel de Cervantes Saavedra

BAUTISTA, T., OETIKER, T., PARTL, H., HYNA, I. y SCHLEGL, E. *Una Descripción de \LaTeX 2 ϵ* . Versión electrónica, 1998.

KNUTH, D. E. *The \TeX book*. Addison-Wesley Professional., 1986.

KRISHNAN, E., editor. *\LaTeX Tutorials. A primer*. Indian \TeX Users Group, 2003.

LAMPORT, L. *\LaTeX : A Document Preparation System, 2nd Edition*. Addison-Wesley Professional, 1994.

MITTELBACH, F., GOOSSENS, M., BRAAMS, J., CARLISLE, D. y ROWLEY, C. *The \LaTeX Companion*. Addison-Wesley Professional, segunda edición, 2004.

OETIKER, T., PARTL, H., HYNA, I. y SCHLEGL, E. *The Not So Short Introduction to \LaTeX 2 ϵ* . Versión electrónica, 1996.

Apéndice **A**

Título del Apéndice A

Los apéndices son secciones al final del documento en las que se agrega texto con el objetivo de ampliar los contenidos del documento principal.

Apéndice	B
----------	----------

Título del Apéndice B

Se pueden añadir los apéndices que se consideren oportunos.

Glosario

AI	Artificial Intelligence
IA	Inteligencia Artificial
metaverse	The metaverse is a set of digital spaces to socialize, learn, play and do other activities. (Definition by the company Meta)
metaverso	El metaverso es un conjunto de espacios digitales para socializar, aprender, jugar y realizar otras actividades. (Definición de la empresa Meta)
NPC	Non Playable Character