



UNIVERSIDAD DE GUADALAJARA
CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍAS

DIVISIÓN DE ELECTRÓNICA Y COMPUTACIÓN

DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

INGENIERÍA EN COMPUTACIÓN

DISPOSITIVO PARA PROYECTOR INTERACTIVO

PARA ACREDITAR LOS MÓDULOS

MÓDULO I. ARQUITECTURA Y PROGRAMACIÓN DE SISTEMAS

MODULO II. SISTEMAS INTELIGENTES

MODULO III. SISTEMAS DISTRIBUIDOS

P R E S E N T A

**ARREOLA CÁRDENAS
IGNACIO**

A S E S O R

MARTÍNEZ SOLTERO ERASMO GABRIEL

Guadalajara, Jalisco, a 15 de julio del 2022

Índice

3.....	Introducción
3.....	Planteamiento del problema
4.....	Objetivos
5.....	Justificación
5.....	Estado del arte
6.....	Hipótesis
6.....	Implementación
8.....	Conclusión
9.....	Referencias

Introducción

Nuestra forma de comunicarnos e interactuar con dispositivos de proyección a distancia ha sido la misma por mucho tiempo. Los comúnmente llamados proyectores suelen tener una interfaz y una forma de conexión muy rudimentaria, nuestras opciones para interactuar con ellos se ven muy limitados si no se cuenta con los aditamentos necesarios o el modelo más nuevo con periféricos extra, que además de tener un alto costo son difíciles de configurar de manera correcta. El objetivo de este proyecto es resolver esto mediante un pequeño dispositivo que se encargará de ser el puente de comunicación e interacción entre nosotros y el proyector; de esta manera no se necesita cambiar de proyector para obtener una mejora en sus capacidades.

A través del uso de algoritmos de visión por computadora y de aprendizaje automático (machine learning) empleados con el lenguaje de programación Python el software es desarrollado completamente para ser utilizado en una computadora de una sola placa Raspberry Pi.

Planteamiento del problema

Actualmente si se desea obtener capacidades mayores o mejoradas de un simple proyector a distancia es casi imposible de realizar sin tener que recurrir a la compra de otro proyector completamente. Esto puede causar un dilema si encontramos que el proyector antiguo sigue funcionando y realizando sus actividades de forma correcta y eficiente, no se debería de tener que conseguir un aparato totalmente nuevo solo para conseguir unas cuantas mejoras en la forma de interacción con el proyector.

Objetivos

Objetivos generales

Generar un dispositivo que de manera externa agregue capacidades de interacción de forma inalámbrica a un proyector; estas capacidades de interacción serán controles por señas de las manos captadas a través de una pequeña cámara web y procesadas dentro de un Raspberry Pi 4 con algoritmos de visión por computadora al igual que algoritmos de aprendizaje automático.

Objetivos específicos

Con este dispositivo se mejora la conexión e interacción que se tiene con un proyector común sin periféricos extra. Anteriormente solo se podía interactuar conectando un ordenador y dando las señales a la computadora a través de periféricos como teclado, ratón, apuntadores, etc; ahora será posible establecer comunicación con esos mismos periféricos agregando gestos realizados con las manos estando frente a la cámara web del dispositivo, al igual que por un smartphone utilizando la tecnología de transmisión de datos Bluetooth.

El dispositivo hace uso de algoritmos de aprendizaje automático para encontrar y descifrar los gestos realizados con las manos.

La computadora de una sola placa Raspberry Pi más novedosa al momento es la Raspberry Pi 4, la cual se puede encontrar en variaciones de 2, 4 y 8 GB de memoria RAM. Gracias al poder que se puede conseguir de dicha computadora es posible utilizar los nuevos algoritmos de aprendizaje automático, los cuales han sido optimizados para su uso en máquinas de recursos más limitados como lo es el Raspberry Pi. El software dentro del Raspberry Pi fue desarrollado en el lenguaje de programación Python, el cual cuenta con una biblioteca de módulos muy extensa comprendiendo casi cualquier utilidad que sea necesaria.

Metas

Desarrollar un dispositivo capaz, responsivo y eficiente. El dispositivo tiene hardware plug-and-play, ya que solo se conectaría un Raspberry Pi al proyector y una cámara web al Raspberry Pi. En el software se encuentra la mayor parte del trabajo; aunque se utilicen

bibliotecas que aportan grandes funcionalidades es necesario desarrollar software escalable, robusto y eficiente.

Justificación

Año con año se desarrollan tecnologías nuevas y se mejoran las ya existentes. Lamentablemente, algunos equipos no se benefician de estas grandes mejoras debido a su simplicidad y a veces falta de visión de las mismas empresas que los maquilan venden para dar seguimiento a su producto y ofrecer mejoras del mismo al consumidor; el cual puede volver a ser fuente de ingresos vendiéndole una mejora de capacidades a algún dispositivo con el que cuente, de igual manera, si se crea un gran producto innovador que de nueva vida a su antiguo dispositivo el cliente quedará satisfecho y se generará afecto hacia la empresa.

Es el deber de los desarrolladores e innovadores el explotar las nuevas tecnologías, encontrando sus límites, puntos débiles y puntos fuertes. De esa manera es posible estar en constante evolución y mejora para destacar en un mercado altamente cambiante y poder brindar más comodidades a la humanidad y conocimientos a la comunidad científica.

Estado del arte

La gran disminución de complejidad y requerimiento de poder computacional para el uso de técnicas de Machine Learning en conjunto a técnicas de Visión por Computadora da un gran campo en el que anteriormente solo las grandes empresas podrían competir. Al día de ahora es posible para cualquier persona enfocada a crear un producto innovador generar una Start Up y llegar a competir con las grandes empresas en esta 4^{ta} revolución industrial. Aunque las grandes empresas son las desarrolladoras y creadoras de avances tecnológicos en estas áreas, es común que liberen dichos avances (aunque no los más grandes, son su secreto a la victoria total) para ser utilizados por desarrolladores fuera de su empresa.

Antecedentes

Qué es una Red Neuronal, Qué es una Red Neuronal Convolutacional, Qué es Visión por Computadora, qué es Raspberry Pi,
Salones de Alfa y Beta donde hay pantalla inteligente, pero no se utiliza

Hipótesis

Desarrollar un dispositivo externo para aumentar las capacidades de un proyector ordinario mediante el manejo de algoritmos de visión por computadora en conjunto a herramientas de machine learning. Se podrá interactuar de manera inalámbrica y sin periféricos con el proyector. Se busca prototipar el dispositivo, que será capaz de realizar las tareas, utilizando un ordenador de placa única asequible como lo es la Raspberry Pi.

Implementación

El dispositivo está compuesto en su totalidad por un Raspberry Pi modelo 4 de 4 GB de memoria RAM y una cámara Logitech C920S. Se utiliza la interpretación de gestos realizados con las manos para poder interactuar con el dispositivo. luego poner todos los controles realizables con las manos.

Esto es posible gracias al uso de MediaPipe, un marco para crear soluciones de aprendizaje automático (ML) multiplataforma. Dentro de MediaPipe encontramos la solución Hands, que es una solución de seguimiento de manos en el dispositivo en tiempo real que predice el esqueleto de la mano de un ser humano a partir de una sola cámara RGB. La canalización consiste de dos modelos: 1) un detector de palma, que provee un cuadro delimitador de una mano, para luego seguir a 2) un modelo de puntos de referencia de la mano, que predice el esqueleto de la mano.

Primera fase:

Las pruebas para el desarrollo del dispositivo comenzaron en la computadora de placa única Raspberry Pi 3B+, la cual cuenta con 1 GB de RAM y GHz.

Para los primeros acercamientos se utilizó el módulo de cámara de Raspberry Pi conocido como PiCamera V1, pero se encontró que no tenía una resolución adecuada para la captación de poses y fallaba cuando se trabajaba con poca iluminación. A continuación se cambió a una cámara web Logitech C920 con la que se rindieron muchos mejores resultados para la obtención de imágenes y video.

Algoritmos de aproximación de pose

Primeramente se intentó generar la interacción del usuario realizando poses utilizando todo su cuerpo, principalmente los brazos. El primer modelo probado fue PoseNet de TensorFlow, que utiliza un modelo de red neuronal de convolución (CNN) para hacer una regresión de la pose a partir de una imagen RGB. Se encontró que PoseNet no contaba con la precisión necesaria para aproximar la posición del cuerpo y sus extremidades de manera adecuada, al mismo tiempo, era un algoritmo muy pesado para la pequeña computadora Raspberry Pi 3B+ que solo pudo procesar un cuadro por segundo, o en ocasiones cuando el dispositivo se sobre calentaba procesaba menos de un cuadro por segundo.

La siguiente implementación que se puso a prueba fue MoveNet de TensorFlow. Esta contaba con una versión de TensorFlow Lite y modelos pre-entrenados y cuantizados para su mejor desempeño en equipos de pocos recursos. A comparación de su predecesor PoseNet, encontramos una precisión muy buena para reconocer y aproximar la posición de los cuerpos humanos y sus extremidades encontrados en la toma. Aunque se desempeñó mejor en el uso de los recursos de la Raspberry Pi, lamentablemente, solo agregó otro cuadro procesado por segundo; lo cual dejaba el procesamiento en únicamente dos cuadros por segundo.

Segunda fase:

Al encontrar que los algoritmos de aproximación de cuerpo completo eran muy demandantes de recursos computacionales se optó por utilizar algoritmos para aproximación de la posición

de manos, los cuales obviamente al reconocer una menor cantidad de puntos son más veloces y menos intensivos computacionalmente.

El primer método probado fue umbralizar la imagen para así conseguir encontrar la mano en el fondo, pero no fue posible, ya que el dispositivo no tendrá un fondo estable en la mayoría de los casos y se tiene que buscar la mano en toda la imagen. Rápidamente se encontró y probó la sustracción de fondo de manera adaptativa para arreglar este problema, lamentablemente debido al tiempo que se tenía para completar el desarrollo del dispositivo esta técnica no sería costeable.

Tercera fase:

Finalmente, se decidió utilizar la solución Hands de MediaPipe, el cual es desarrollado por Google y es de código abierto. Hands aproxima la posición de 21 puntos clave en la mano. Al igual que con los algoritmos de aproximación de pose, es algo intensivo en el uso de recursos, pero en un Raspberry Pi 4 se obtienen alrededor de 6 cuadros procesados por segundo reconociendo hasta dos manos en la imagen.

Al mismo tiempo fue posible obtener un Raspberry Pi 4 de 4 GB de RAM, con el cual se obtuvieron grandes resultados utilizando los algoritmos de aprendizaje automático de MediaPipe.

Conclusiones

Debido al poco tiempo obtenido para la realización del proyecto modular, no fue posible concretar la hipótesis. Las bases para continuar y terminar de desarrollar el proyecto son firmes y con poco más de tiempo sería posible tener un prototipo funcional.

El sistema puede ser aumentado considerablemente para convertir a todo un salón en un aula inteligente; tomando asistencia por reconocimiento facial, capturando y enviando las notas de la clase a un repositorio remoto al que los estudiantes del curso tengan acceso, al igual que transmitir en vivo las clases para estudiantes en modalidad virtual o híbrida.

Otras mejoras para el sistema son los modos recreativos; uno de ellos dando posibilidad de realizar gráficos cambiantes respondiendo a movimientos de personas en tiempo real, otro permitiendo realizar mapeo de proyección con facilidad a todas las personas.

Referencias

<https://blog.tensorflow.org/2018/05/real-time-human-pose-estimation-in.html>

<https://www.geeksforgeeks.org/posenet-pose-estimation/>

<https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html>

<https://arxiv.org/pdf/2006.10214.pdf>

https://drive.google.com/file/d/1-rmIgTfuCbBPW_IFHkh3f0-U_lnGrWpg/preview