Reconocimiento de Imágenes y Asistente de Voz Utilizando Robótica Social

Ing. Alicia Alexis Llusco Cuba
JUNIOR Carrera de Ingeniería de Sistemas, Escuela Militar de Ingenieria
La Paz, Bolivia
alluscoc@est.emi.edu.bo

Image Recognition and Voice Assistant Using Social Robotics

Resumen – El objetivo general es desarrollar un software de reconocimiento de siluetas, que este basado en redes neuronales y asistente de voz con órdenes predeterminadas que cuenta con una interfaz visual que brinde información.

Palabras Claves— Alrededor de cuatro palabras o frases clave en orden alfabético, separadas por comas. Para obtener una lista de palabras claves sugeridas, envíe un correo en blanco a keywords@ieee.org o visite el sitio web de IEEE

Abstract - The general objective is to develop a silhouette recognition software. It is based on neural networks and voice assistant with predetermined orders that has a visual interface that provides more information.

Keywords— About four key words or phrases in alphabetical order, separated by commas. For a list of suggested keywords, send a blank email to keywords@ieee.org or visit the IEEE website at

I. INTRODUCCIÓN

En la actualidad, los robots comerciales e industriales son ampliamente utilizados, y realizan tareas de manera más exacta o barata que los humanos. Los robots son muy utilizados en plantas de manufactura, montaje y embalaje, en transporte, en exploración en la Tierra y en el espacio, cirugía, armamento, investigación en laboratorio, turismo y en la producción en masa de bienes industriales y de consumo, llamados también Robot Sociales.

Por otro lado, el enorme desarrollo que está viviendo la tecnología asociada a la Inteligencia Artificial (IA) está dando lugar en los últimos tiempos a nuevas herramientas y aplicaciones espectaculares en lo que es la robótica. Una de las áreas donde los avances han sido más notables es el reconocimiento de imágenes, en parte gracias al desarrollo de nuevas técnicas de Deep Learning o aprendizaje profundo. Hoy en día tenemos ya al alcance de nuestra mano sistemas más precisos que los propios humanos, en las tareas de clasificación y detección en imágenes.

Otro de los avances de la inteligencia artificial es el reconocimiento de voz que trata de establecer una comunicación entre el hombre y los ordenadores, a través del lenguaje humano, asociado a un asistente personal inteligente de voz que pueda realizar tareas u ofrecer servicios a un individuo. Estas tareas o servicios están basados en datos de entrada de usuario, reconocimiento de voz y la habilidad de acceder a información de una variedad de recursos programados o establecidos como ser las bases de datos.

Tomando en cuenta los avances tecnológicos mencionados anteriormente, el proyecto se enfoca en desarrollar un software como asistente virtual aplicable a la robótica social, diseñado para relacionarse con humanos que fomente la empatía y la confianza, dotando de capacidades que conviertan a este asistente inteligente en un avanzado punto de interacción con clientes. Capaz de interactuar con personas a través de comando de voz predeterminados, texto e imágenes.

II. INFORMACIÓN BÁSICA DEL PROYECTO

A. Planteamiento del problema

Hoy en día los visitantes de los museos desean conocer y comprender la información brindada en ellos, ya sea de forma interactiva o de manera personal, en definitiva, disfrutar y tener una experiencia que pueda recordar en un futuro.

Es necesario si se desea lograr una comunicación efectiva, implicarlos, hacerlos sentir participes de manera personal. La mayoría de los museos no cuentan con asistentes que brinden mayores detalles acerca de sus exposiciones, toda la información que el visitante obtiene es procedente de folletos, libros, paneles informativos, guías. Esto limita a aquellos visitantes que desean profundizar la información en algún aspecto en concreto.

Los museos se encuentran frente al reto de aumentar el interés por las obras dentro de las exposiciones tradicionales, con un contexto que lo respalde y lo explique. En La Paz Bolivia una estrategia para incrementar las visitas, están actividades como: la Larga Noche de Museos organizada por la Oficialía Mayor de Culturas, donde se presenta al público en general los distintos establecimientos, de esta manera se conoce más de la cultura e historia reflejada dentro de los espacios culturales e históricos, durante la actividad los museos cuentan con guías gratuitos que ayudan a brindar más información, el Museo Histórico Militar participa de esta actividad, pero esto solo sucede una vez al año.

B. Justificación

El presente documento es una propuesta para la ejecución de aplicar el reconocimiento de imágenes para las siluetas y así lograr el control y direccionamiento de una estructura robótica, utilizando un sistema con asistente de voz que permita brindar información almacenada en la base de datos a través de ordenes verbales predeterminadas de entrenamiento previo.

C. Pre – procesamiento de Imágenes

En computación una escala de grises es una escala empleada en la imagen digital, en donde el valor de cada píxel posee un valor equivalente a una graduación de gris, el equivalente a la luminancia de la imagen. Como se sabe el ojo percibe distintas intensidades de luz en función del color que se observe, esto es debido a la respuesta del ojo al espectro visible la cual se puede observar en la Figura 1, por esta razón el cálculo del equivalente blanco y negro (escala de grises o luminancia) de la imagen debe realizarse como una media ponderada de las distintas componentes de color de cada píxel.

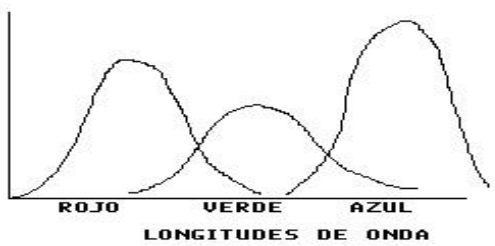

Fig. 1 Componentes RGB DE UNA IMAGEN

La umbralización es un proceso que permite convertir una imagen de niveles de gris o de color en una imagen binaria, de tal forma que los objetos de interés se etiqueten con un valor distinto al de los píxeles del fondo. En adelante sólo se hablará de imágenes en niveles de gris, aunque la extensión a color es inmediata si sólo se usa una de las componentes RGB o alguna mezcla de las tres.

La umbralización es una técnica de segmentación rápida, que tiene un coste computacional bajo y que puede ser realizada en tiempo real durante la captura de la imagen usando un computador personal de propósito general, como se observa en la Figura 2.

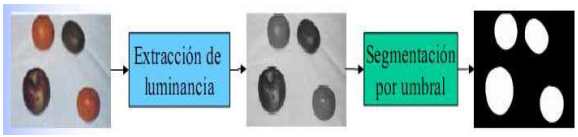

Fig. 2 Segmentación basada en el umbral

D. Redes neuronales convolucionales

La principal diferencia de la red neuronal convolucional con el perceptrón multicapa viene en que cada neurona no se une con todas y cada una de las capas siguientes, sino que solo con un subgrupo de ellas (se especializa), con esto se consigue reducir el número de neuronas necesarias y la complejidad computacional necesaria para su ejecución véase Figura 3.

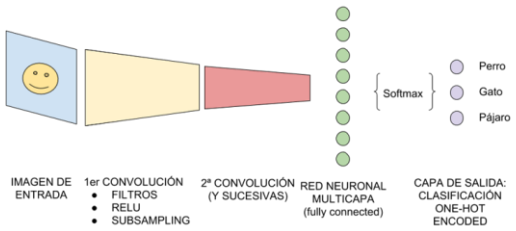

Fig. 3 Red neuronal Convolucional

La clasificación de imágenes a partir de una red neuronal como las vistas en la sección anterior puede ser un problema debido a la carga computacional que

esto supone. Como se comenta, en las redes neuronales todas sus neuronas están totalmente conectadas entre capas.

Si por ejemplo se dispone de una imagen con unas dimensiones de 640 x 640 píxeles, esta hace un total de 409.600 datos de entrada que estarán conectados a cada una de las neuronas de la primera capa. Esto produce un gran número de cálculos que ralentizará el proceso de entrenamiento.

Para mitigar este problema se presentan las redes neuronales convolucionales. Las

redes neuronales convolucionales están formadas por diferentes, donde las primeras capas se encargan de extraer características como los bordes o esquinas de una imagen. Una vez detectados estos bordes son utilizados para detectar formas en las capas posteriores.

Cuando se detectan las formas se procede a detectar las características de alto nivel como puede ser una rueda en una imagen donde aparece un coche. Las últimas capas están totalmente conectadas y son las encargadas de dar una predicción a partir de estas últimas características extraídas ejemplo véase Figura 3.

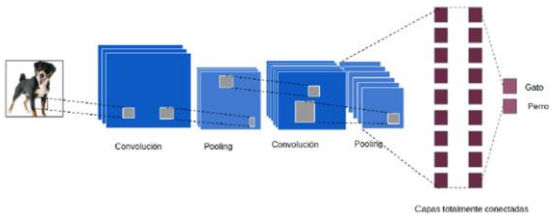

Fig. 3 Red convolucional

Reducir el número de conexiones entre la capa oculta y la capa de entrada es su principal funcionalidad. A partir de unos filtros o también llamados kernel se extraen las principales características de una imagen.

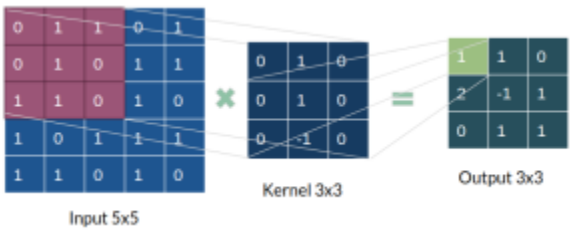

Fig. 4 se observa cómo se realiza un producto escalar entre el kernel y la matriz de la imagen de entrada. En esta imagen se ha establecido un único kernel pero a la hora de calcular las dimensiones de la salida se tiene en cuenta:

III. LA MATEMÁTICA

Las ecuaciones necesarias para el procesamiento de imágenes son las siguientes:

CONVERSIÓN ESCALA DE GRISES

$$Gris = 0{,}2989 * rojo + 0{,}5870 * verde + 0{,}1140 * azul$$

APLICACIÓN DE UMBRAL

$$L(x,y) = \begin{cases} 1, I\ (x,y) \leq u \\ 0, I\ (x,y) > u \end{cases}$$

$$L_u = \{(x,y) \in \Omega\ /\ I\ (x,y) < u$$

Método automático para separar objeto del fondo:

- Cálculo del histograma de gris
- Cálculo iterativo de media y varianza
- Hipótesis de umbral U=> divide el histograma en dos partes y se calcula la media varianza para cada parte iterativamente cambiando U.

SEPARAR OBJETO DEL FONDO

$$W_1(t+1) = \sum_{X=0}^{U-1} P_W(x)$$

$$W_1(t+1) = \frac{1}{W_1(t+1)} \sum_{X=0}^{U-1} x P_u(x)$$

$$\theta_1{}^2\ \ (t+1) = \frac{1}{W_1(t+1)} \sum_{X=0}^{U-1} (x - m_1(t+1))^2 P_u(x)$$

$$\begin{cases} D = 0\ \ si\ U\ no\ esta\ entre\ ambos \\ D = \theta^2\ \ si\ U\ está\ entre\ ambas \end{cases}$$

Este conjunto produce una división del espacio. La cantidad de componentes conexas de L_u determinan el número de regiones.

IV. RESULTADOS

La red neuronal utiliza el modelo de Aprendizaje Supervisado ya que entrenaremos a la red con datos conocidos de imágenes de diferentes siluetas. Se utilizará dichos datos para poder entrenar la red neuronal y obtener los resultados correctos. En ciertos puntos la red neuronal podría llegar a cometer ciertos errores, pero con el modelo que elegimos podremos minimizar los errores que se puedan generar al momento de emplear la red neuronal. La red neuronal convolucional nos permite tener datos digitales como entrada como también de salida, por

lo que se segmentará a las imágenes de entrenamiento en matrices y se lo convertirá a binario con una conversión previa a escala grises de la imagen. Así se tendrá las posiciones que serán en 1 y en 0, se podrá comparar la base de conocimiento con las imágenes que se captura por la cámara integrada a la arquitectura robótica.

Fig. Conversión escala de grises

Fig. Conversión a binario

Una vez determinada la adquisición de datos de ingreso y el pre - procesamiento de las imágenes de entrada se tiene la siguiente arquitectura que se puede observar en la Figura con las dos convoluciones determinadas y la conexión a una red neuronal monocapa.

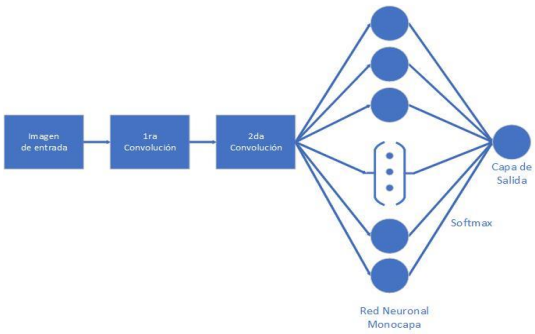

Fig. Arquitectura de la red neuronal convolucional

Los módulos trabajados y desarrollado en la fase de las iteraciones deben ser integradas bajo una aplicación y la estructura Robótica

Para la integración de cada uno de los subsistemas del software se tiene el siguiente diagrama:

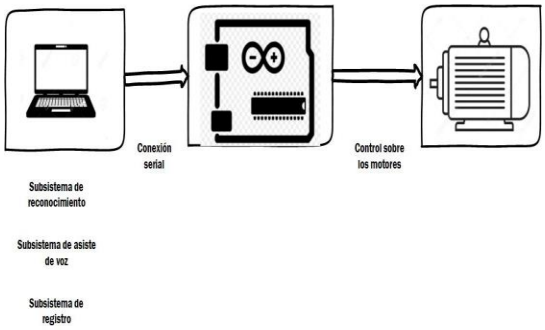

Fig. Integración del sistema

V. CONCLUSIÓN

Las redes neuronales convolucionales es un modelo que mejor funciona para el procesamiento de imágenes debido a que reduce la carga en cuanto a la cara computacional.

REFERENCES

[1] Barca, Rafael (2013) "Introducción a la Robótica", Universidad Alcalá Madrid España
[2] Vélez Serrano, J. F., Moreno Díaz, A. B., Sánchez Calle, A. Sánchez Marín, J. L. E. (2003). Visión por Computador 2da Edic. [en línea].
[3] Sucar, L., E., Gómez, G. (2011). Visión Computacional. [en línea]. Instituto Nacional de Astrofísica, Óptica y Electrónica. Puebla, México.
[4] Sobrado Malpartida, E. A. (2003). Sistema de visión artificial para el reconocimiento y manipulación de objetos utilizando un brazo robot. (Tesis de grado de magister en ingeniería de control y automatización). Pontificia Universidad Católica del Perú. Lima.
[5] Inteligencia Artificial 1a ed. - Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn), 2014. 225 Primera Edición: Marzo 2014 Iniciativa Latinoamericana de Libros de Texto Abiertos (LATIn)
[6] koro irusta gonzalo aprendizaje en robots sociables 2017
[7] Kendall, J; Kendall, K.2011 2009. Análisis y diseño de sistemas. 8 ed. PEARSON EDUCACIÓN, MX. Pag. 8-12, 14-19.
[8] García, J. (2017). Dinámica de Sistemas, La Teoría General de Sistemas. Segunda Edición.
[9] (IEEE, 2015). IEEE. (1977). Standard Dictionary of Electrical and Electronic Terms.

Fecha de Envió del Articulo: 14/05/2020
Fecha de Aceptación de artículo: 28/05/2020