
UNIVERSIDAD
DE LOS ANDES
VENEZUELA

Proyecto de Grado 

Presentado ante la ilustre Universidad de Los Andes como requisito parcial para obtener el Título de Ingeniero de Sistemas

# Reconocimiento multimodal de estados afectivos para el robot social LRS2

Por

Br. Nerio Moran

Tutor: Prof. Jesús Pérez

Marzo 2018

# Reconocimiento multimodal de estados afectivos para el robot social LRS2

Br. Nerio Moran

Proyecto de Grado — Sistemas Computacionales, 3 páginas
Escuela de Ingeniería de Sistemas, Universidad de Los Andes, 2018

**Resumen:** En este trabajo se presenta una propuesta para el reconocimiento de emociones a partir de señales multimodales, específicamente expresiones faciales y el habla, en el contexto de interacción humano-robot. El problema planteado hace parte de los estudios actuales que buscan mejorar la comunicación humano-robot mediante la inclusión de diferentes señales que permitan determinar el estado emocional de los humanos de una manera continua y precisa. El sistema propuesto aplica técnicas de aprendizaje de máquina para lograr la detección de emociones en una persona mientras interactúa con el robot LRS2.

**Palabras clave:** Interaccion Humano Robot, Aprendizaje de Maquina, emociones, Redes neuronales, Multimodal

acento en la a

*No hay dedicatoria todavia*

acento en la i

# Índice general

# Introducción

Dentro del área de la interacción humano-robot, se ha generado una gran iniciativa de investigación en la detección de emociones, esto bajo la justificación de acortar la brecha de comunicación entre humanos y robots. Mejorar la capacidad de interacción mediante el reconocimiento del estado emocional del humano es de gran importancia para permitir que los robots puedan desempeñar tareas cooperativas y de una manera natural con los humanos.

Actualmente las áreas de aplicación para la interacción humano-robot son diversas y se mantienen en expansión. Alguna de ellas: robots sociales, entornos inteligentes, enseñanza, terapias y tratamientos, asistencia entre otros.

La importancia que ha adquirido durante los últimos años la interacción entre humanos y robots es debido a la necesidad de desarrollar tareas de cooperación entre ellos, de tal forma que cualquier persona pueda interactuar de forma natural sin la necesidad de entender el robot.

La creación de sistemas lo suficiente robustos para permitir este tipo de interacción requieren de distintas señales de entrada independientes que se transformaran en información multimodal. El procesamiento de información proveniente de múltiples fuentes es un problema que se encuentra en diferentes áreas de investigación como la inteligencia artificial y la robótica.

# Capítulo 1

# Planteamiento del problema

## 1.1. Antecedentes

Los antecedentes de esta investigación se dividen principalmente en dos categorías: primero, trabajos relacionados con el desarrollo de modelos de aprendizaje para distintos tipos de señales de entrada; segundo, investigaciones que utilizan mas de 1 señal de entrada en sistemas de reconocimiento multimodal. acento en la a

Faria (2017) [2] realizo acento en la o titulado "Affective Facial Expressions Recognition for Human- quita la fecha, deja solo el numero de referencia un marco de trabajo para el reconocimiento de emociones en la interacción humano-robot. Para el marco de trabajo utilizo una base de datos "Karolinska Directed Emotional Faces (KDEF)", la cual utilizo para entrenar mediante aprendizaje supervisado el clasificador. El modelo se enfoco en detectar las 7 emociones universales, para esto utilizaron un modelo dinámico bayesiano mezclado, el cual es una especialización de las redes dinámicas bayesianas.

Razuri [5] estudio las características cruciales del habla para el reconocimiento de emociones es su investigación titulada "Speech emotion recognition in emotional feedback for Human-Robot Interaction". En su investigación utilizo diferentes tipos clasificadores para probar cuales de ellos se acercaba a un sistema de tiempo real, la base de datos de entrenamiento fue eNTERFACE05. La investigación tuvo como conclusión que los clasificadores: Maquina de vectores de soporte (SVM) y las redes bayesianas son buenos candidatos para los sistemas de tiempo real.

Por otro lado, Ménard (2015) [3] utilizo como señales de entrada el ritmo cardiaco y la conductancia de la piel para determinar las emociones en su investigación titulada "Emotion Recognition Based on Heart Rate and Skin Conductance ". Ambas entradas fueron usadas en clasificadores distintos, crearon su propia base de datos utilizando sensores de respuesta biológica. Utilizaron como clasificador para ambas entradas una maquina de vectores de soporte (SVM) y el cual fue entrenado con los coeficientes de fourier de ambas entradas. Se determinó que ambas entradas fisiológicas sirven como fuente de información para determinar una emoción y a diferencia de otras entradas como la voz o la imagen digital estas señales tienen disponibilidad continua.

En los sistemas de reconocimiento multimodal [4] en su tesis de maestria titulada "Identificación de señales multimodales para reconocimiento de emociones en el contexto de interacción humano-robot ", realiza un sistema de reconocimiento bimodal cuyas entradas son imágenes y el habla. Para las imágenes utiliza una red neuronal convolucional y para el habla utiliza una maquina de vectores de soporte (SVM). Para la de fusión utiliza un sistema basado en decisiones basándose en el resultado de ambas entradas.

Castillo (2018) [1] en su investigacion titulada "Emotion Detection and Regulation from Personal Assistant Robot in Smart Environmentrealiza un diseño para un robot social asistente el cual cuenta con sistema de reconocimiento bimodal basado en imágenes y el habla. Para detectar la emoción del usuario con el que interactuá utiliza la representación continua de las emociones basada en valencia y excitación, solo detecta 4 estados emocionales: Sorpresa, Felicidad, tristeza y neutralidad. Las características del audio fueron: Pitch, Flux, Rolloff-95, Centroid, Zero-crossing rat, SNR y el ritmo comunicativo. Para el audio se utilizaron dos modelos de clasificación uno basado en un árbol de decisión y el otro basado en reglas de decisión. Para el reconocimiento de las imágenes utilizaron dos softwares CERT y SHORE. En el componete de fusión utilizaron un método basado en reglas de decisión.

# Bibliografía


[1] J. Castillo, A. Castro Gonzalez, F. Alonso Martin, A. Fernandez Caballero, and M. Salichs, "Emotion detection and regulation from personal assistant robot in smart environment," pp. 179–195, 01 2018.

[2] F. C. F. Diego R. Faria, Mario Vieira and C. Premebida, "Affective facial expressions recognition for human-robot interaction," *IEEE*, pp. 805–810, 2017.

[3] M. Ménard, P. Richard, H. Hamdi, B. Daucé, and T. Yamaguchi, "Emotion recognition based on heart rate and skin conductance," pp. 26–322, 01 2015.

[4] A. K. Perez Hernandez, "Identificación de señales multimodales para reconocimiento de emociones en el contexto de interacción humano-robot," Universidad de los Andes Colombia, Facultad de Ingenieria, Departamento de Ingeniería de Sistemas y Computación, Proyecto de Maestria , 2016.

[5] J. G. Rázuri, D. Sundgren, R. Rahmani, A. Larsson, and A. M. Cardenas, "Affective facial expressions recognition for human-robot interaction," *(IJARAI) International Journal of Advanced Research in Artificial Intelligence,*, 2015.