

# Clasificación de información táctil para la detección de personas

Juan M. Gandarias, Jesús M. Gómez-de-Gabriel y Alfonso J. García-Cerezo



UNIVERSIDAD DE MÁLAGA

Dto. de Ingeniería de Sistemas y Automática  
Universidad de Málaga  
jmgandarias@uma.es

## Introducción

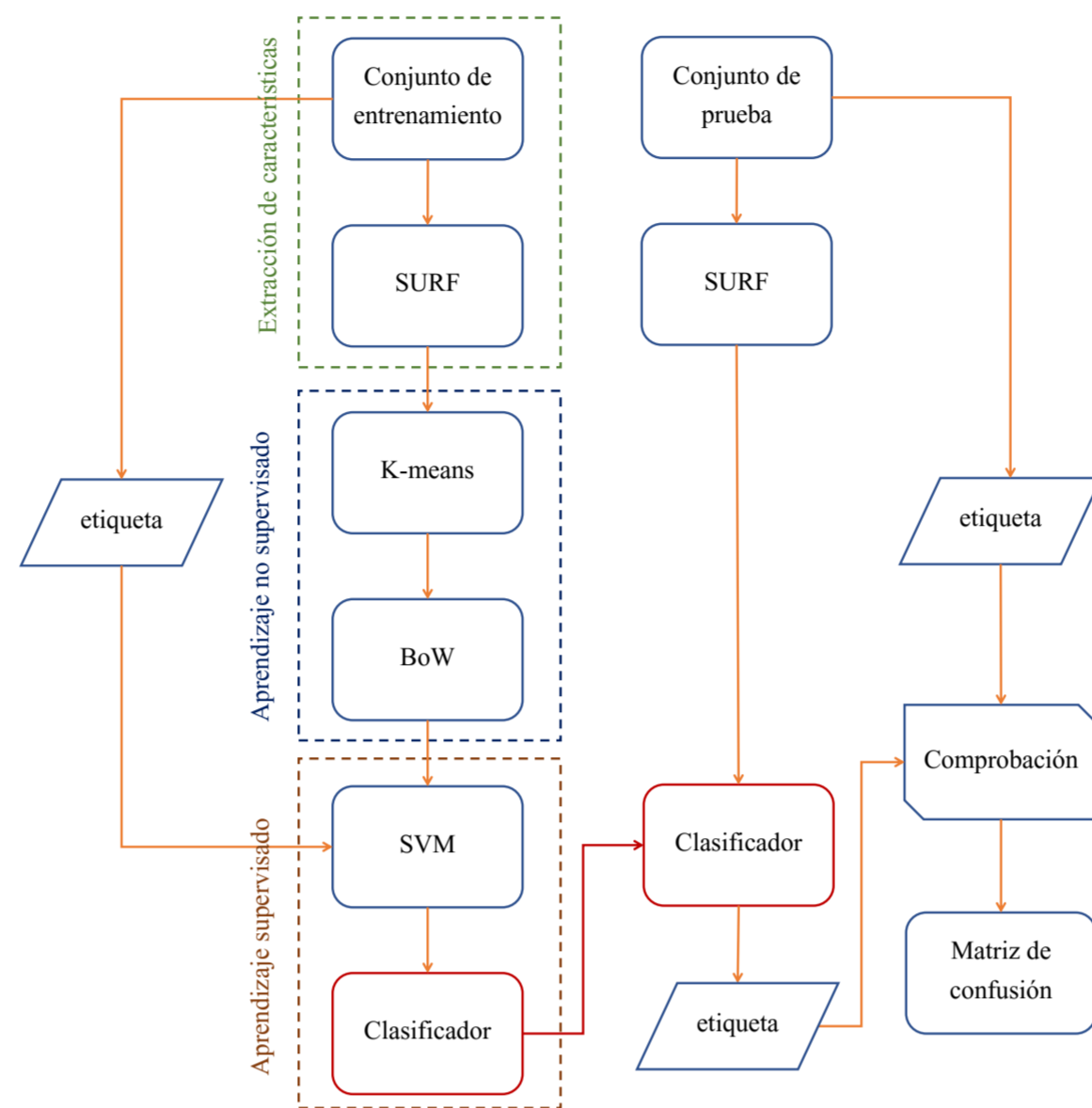
Este artículo presenta el diseño de un efector final táctil y la aplicación de técnicas de inteligencia artificial para la detección de personas mediante un brazo manipulador ligero de 6 grados de libertad. Este efector está compuesto por un sensor táctil de alta resolución que permite obtener imágenes de presión. El sistema extrae información háptica en situaciones de catástrofe en las que, generalmente, existe baja visibilidad, con el propósito de evaluar el estado de las víctimas en función de la urgencia de atención (triaje). Se han implementado dos métodos de inteligencia artificial para clasificar imágenes obtenidas por el sensor táctil, distinguiendo los contactos con personas de objetos inertes en escenarios de desastre. Cada método dispone de un extractor de características de imágenes de presión y un clasificador, obtenido por aprendizaje supervisado. Para validar los métodos se han realizado experimentos de clasificación en clases Humano y No humano. Finalmente, se ha realizado una comparación de ambos métodos en términos de porcentaje de acierto y tiempo empleado para la clasificación, en base a los resultados de los experimentos.



## Métodos

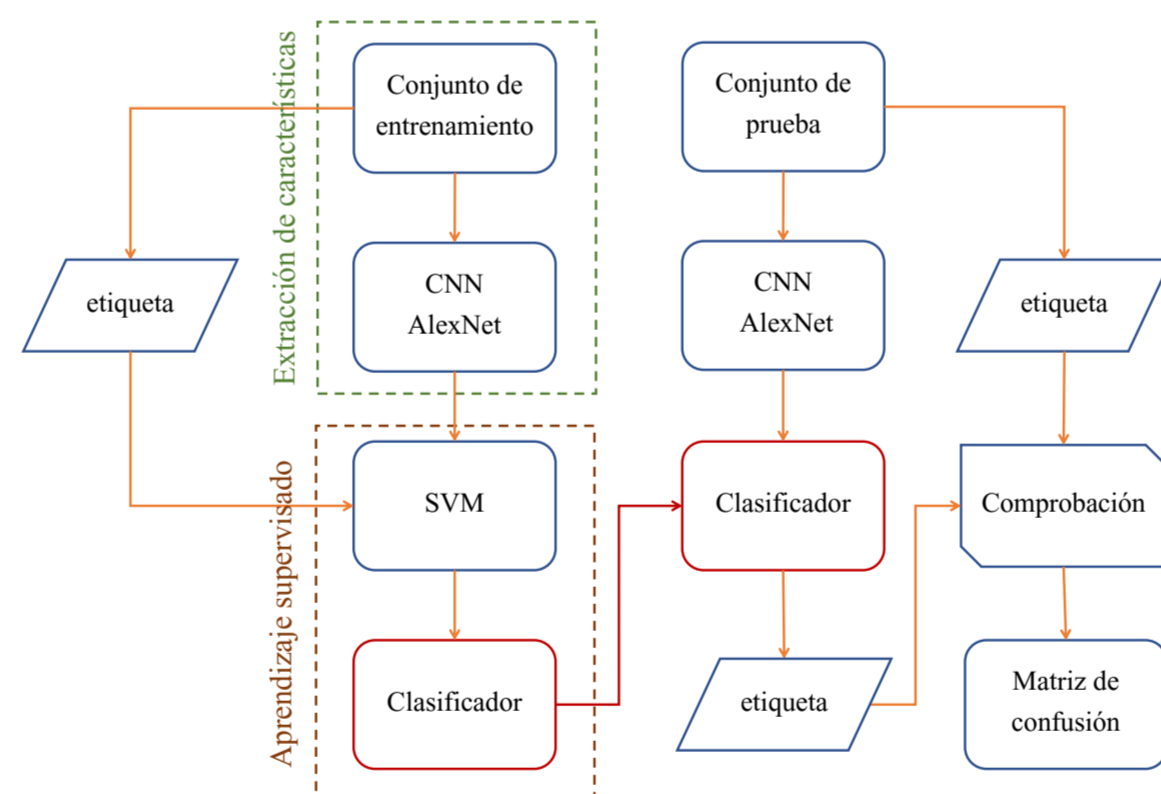
### SURF, k-means, BoW y SVM

Este método se basa en la extracción de características de las imágenes del conjunto de entrenamiento, diferente al conjunto de prueba, mediante el algoritmo de visión por computador SURF, que proporciona un detector y descriptor invariante a escala y rotación convirtiéndolo en un método rápido y robusto para describir imágenes. Posteriormente, basándonos en un modelo BoW, se genera un diccionario en el que las palabras se corresponden con las características previamente extraídas, de manera que los descriptores de las imágenes se agrupan en función de las similitudes encontradas por el método de agrupamiento k-means no supervisado. Finalmente, se obtiene un clasificador utilizando un método de aprendizaje supervisado, que se basa en el entrenamiento de una SVM.



### CNN AlexNet y SVM

El segundo método emplea una red neuronal convolucional profunda (CNN), conocida como AlexNet. La red se ha importado del repositorio de *caffe*, y posee 650000 neuronas distribuidas en 8 capas, de forma que las 5 primeras son convolucionales y las 3 últimas son completamente conectadas. La red importada ha sido entrenada previamente con el fin de clasificar imágenes dentro de 1000 clases diferentes. Esta idea consiste en utilizar una red neuronal como extractor de características en lugar de como método de clasificación. De forma que se toman las activaciones de la última capa previa a la clasificación, para entrenar una SVM.



En la fase de entrenamiento se obtienen las activaciones de la capa 7 de la red neuronal, llamada *fc7*. Estas activaciones se corresponden con las de la última capa previa a la clasificación, y se utilizan para realizar un entrenamiento supervisado con una SVM que genera un clasificador.

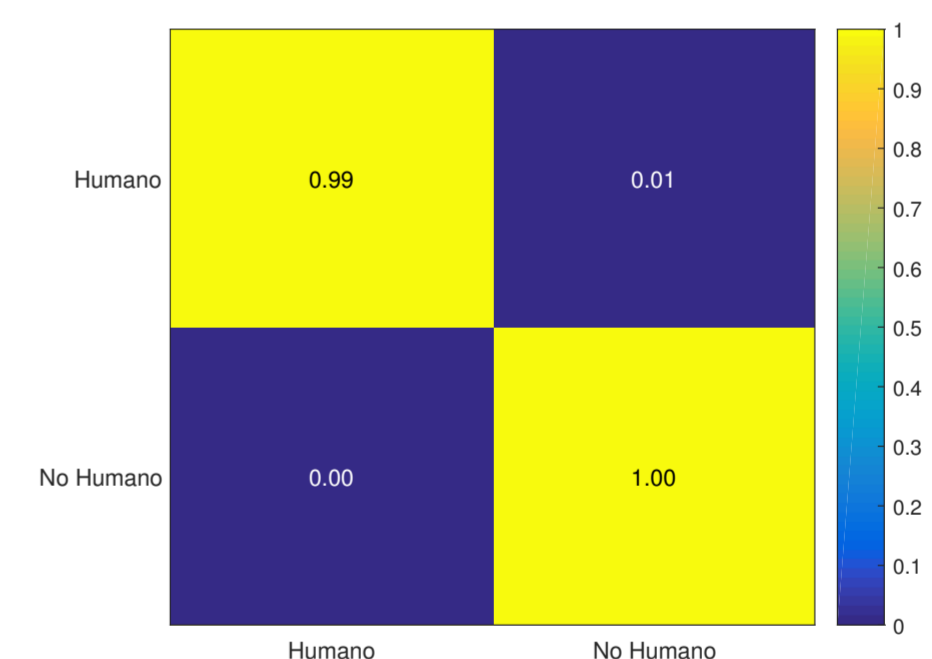
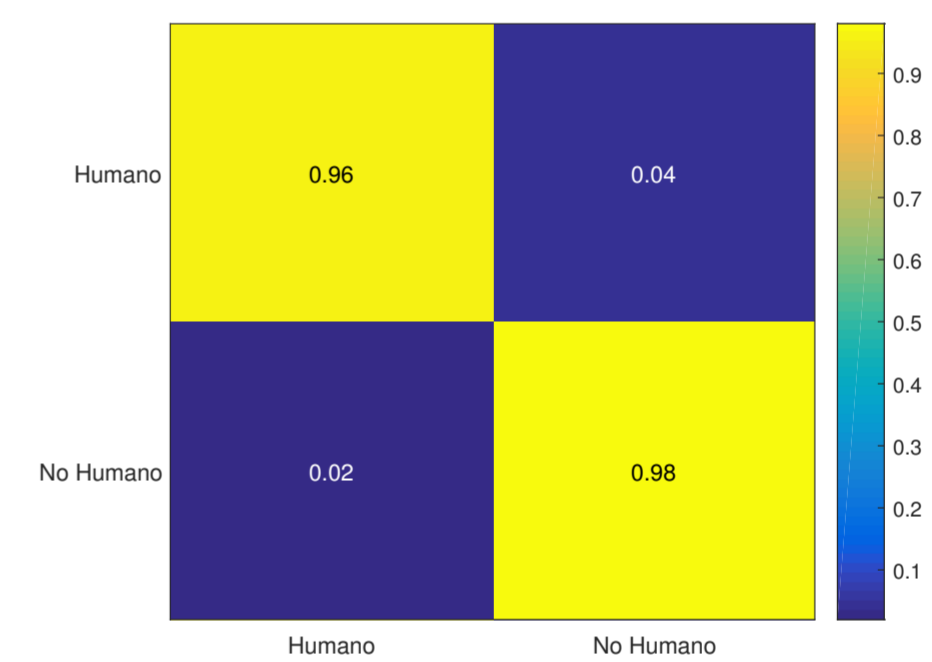
## Experimentos

Se ha desarrollado un efector final para un manipulador de 5 grados de libertad que incorpora el sensor 6077 de Tekscan. Este sensor posee un total de 1400 sensels resistivos de presión distribuidos en una matriz de 28 filas por 50 columnas con un tamaño de 53.3 mm x 95.3 mm y una densidad de 27,6 sensels/cm<sup>2</sup>. El sensor está cubierto por una capa de caucho de silicona que sirve como elemento protector y conductor de fuerzas externas.

Los objetos utilizados en la clase no humano han sido: bolígrafo, alicates y tijeras, mientras que las partes del cuerpo a las que hace referencia la clase humano han sido: brazo, mano y dedos. Los métodos se han entrenado utilizando un conjunto de entrenamiento de 120 imágenes, y ha sido evaluado utilizando un conjunto de prueba de 180 imágenes etiquetadas en las clases Humano y No humano.

## Resultados y Conclusiones

La matriz de confusión resultante del método 1 muestra un índice de acierto del 96.67%. Por otro lado, el tiempo que se tarda en clasificar una nueva imagen de presión es del orden de 10ms. Sin embargo, la matriz de confusión resultante del método 2 muestra un índice de acierto del 99.44%, siendo el tiempo de clasificación es del orden de 700ms.



En resumen, se ha obtenido una mejora del segundo método con respecto al primero del 2.77% en la clasificación Humano - No humano. Sin embargo, el tiempo empleado es, aproximadamente, 70 veces mayor en el segundo. En la tabla se recogen los resultados de tiempo y porcentaje de acierto de cada método, así como el porcentaje de mejora, tomando el primer método como referencia.

Método	Acierto (%)	Mejora (%)	Tiempo (s)
SURF	96.67	-	0.01
CNN	99.44	2.77	0.7

## Agradecimientos

Este trabajo ha sido financiado por el proyecto DPI2015-65186-R y por la ayuda BES-2016-078237 del fondo social europeo FSE.