

Clasificación de información táctil para la detección de personas

Juan M. Gandarias, Jesús M. Gómez-de-Gabriel y Alfonso García-Cerezo
Dto. de Ingeniería de Sistemas y Automática
Universidad de Málaga
jmgandarias@uma.es

Resumen

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.

Palabras clave: Sensores táctiles, robótica de rescate, reconocimiento de objetos, aprendizaje automático.

1. INTRODUCCIÓN

La teleoperación supone un elemento fundamental en el campo de la robótica de rescate, debido a la complejidad de las operaciones a realizar en un entorno no estructurado [7]. Experiencias previas en situaciones reales han evidenciado la problemática de los sistemas basados en percepción visual. En entornos con escasa iluminación, polvo o humo, los sistemas con percepción háptica aportan información adicional que puede compensar las limitaciones visuales [13].

Una de las primeras tareas de la robótica de rescate consiste en clasificar el estado de las víctimas, una vez localizadas, en función de la urgen-

cia de atención (*triaje*). Esta tarea presenta retos tecnológicos como la interacción robot-humano o Human-Robot Interaction (*HRI*), considerado uno de los mayores desafíos de la robótica de rescate [12]. Un primer enfoque a la resolución del problema supondría el reconocimiento de víctimas y de las distintas partes del cuerpo, de cara a poder realizar mediciones posteriormente. Utilizándose para ello percepción táctil únicamente.

En este sentido, existen diversos trabajos cuyo objeto consiste en reconocer objetos utilizando sensores táctiles. La mayoría de estos trabajos están basados en el uso de algoritmos de inteligencia artificial. Algunos sugieren el uso de herramientas de aprendizaje profundo o *Deep Learning*. Así, en [14] se presenta el uso de aprendizaje profundo con técnicas de *dropout* para la reducción del sobreajuste (*overfitting*), y se presentan las mejoras obtenidas al mezclar información táctil y la posición del robot para el reconocimiento. En [1], por el contrario, se aplican herramientas de aprendizaje profundo para la clasificación de los materiales en contacto.

Por otro lado, en [10, 9] se presenta el uso de herramientas de aprendizaje automático o *machine learning* para el reconocimiento de objetos utilizando el descriptor *SIFT* (*Scale-Invariant Feature Transform*) [8] y métodos de clasificación basados en bolsas de palabras o *Bag of Words* (*BoW*). Un trabajo posterior incorpora información de la posición del objeto para crear un algoritmo de clasificación más robusto y eficiente [11].

El Departamento de Ingeniería de Sistemas de la Universidad de Málaga ha contribuido a la aplicación de sensores táctiles a la robótica de rescate [15], mediante un sensor de presión que proporciona una imagen de las fuerzas de contacto con el entorno, o las víctimas, instalado en la pinza del brazo hidráulico de un robot de rescate [4].

Este artículo presenta una aplicación de los sensores táctiles al campo de la robótica de rescate, consistente en el desarrollo de un efector final de un manipulador ligero de 6 grados de libertad, en el que se ha dispuesto un sensor táctil con el objetivo de obtener imágenes de presión del en-



(a)



(b)

Figura 1: Desarrollo del sensor. (a) Recubrimiento del sensor con caucho de silicona. (b) Montaje del sensor en el brazo manipulador.

torno. Asimismo, se presentan y comparan dos métodos de inteligencia artificial para la clasificación de imágenes de presión. El primer método está compuesto por el descriptor *Speeded-Up Robust Features (SURF)* [2], un modelo *Bag of Words (BoW)* y una *Supported Vector Machine (SVM)* [3]. Mientras que el segundo método se basa en el uso de una *Deep Convolutional Neural Network (CNN AlexNet)* [6] y una *SVM*. De esta forma, un clasificador mediante aprendizaje supervisado, identifica si el contacto se realiza con un humano o con objetos. Finalmente, se evalúan ambos métodos comparando sus porcentajes de aciertos y tiempos de cómputo empleados.

El resto del artículo se organiza de la siguiente manera. En la sección 2 se resume la aplicación junto con el efector táctil desarrollado. En la sección 3 se describen las implementaciones realizadas de los métodos de clasificación utilizados. Posteriormente, en la sección 4 se describen los experimentos y se discuten los resultados obtenidos. Finalmente, se incluyen conclusiones y se plantean trabajos futuros.

2. SENSOR TÁCTIL PARA ROBÓTICA DE RESCATE

Los escenarios de desastre suelen caracterizarse por la limitada percepción visual debido a la presencia de humo o polvo, tanto en suspensión como en superficie. La información táctil puede ser crucial para la identificación de víctimas potenciales. En este sentido, un primer enfoque consistiría en saber con qué clase de objeto se va a interactuar.

Esta primera aproximación permite conocer, de forma automática, si se está tratando con una víctima o no, utilizando únicamente información táctil en forma de mapas de presión. De manera

que se pretende obtener una distinción de estos mapas en dos clases, *Humano* o *No humano*. Para lo que se han desarrollado dos métodos de aprendizaje automático, cuyas características se detallan en el apartado 3.

El sistema que se presenta pretende dotar de capacidad táctil a un brazo manipulador de 6 grados de libertad, el modelo *OUR-i5* de la familia *AU-BO*. En las imágenes de la figura 1 se presenta el desarrollo del sensor táctil que se ha acoplado al robot. El sensor está dispuesto sobre una base de PVC y recubierto de una capa protectora de caucho de silicona que permite la percepción de fuerzas externas. Todo el conjunto se ha diseñado en *SolidWorks* y se han utilizado técnicas de fabricación aditivas de prototipado rápido para la construcción de los soportes y el acoplamiento al brazo.

3. IMPLEMENTACIÓN DE LOS CLASIFICADORES

Se han implementado dos algoritmos de inteligencia artificial. Ambos métodos se basan en un extractor de características de las imágenes de presión y en la obtención de un clasificador aplicando métodos de aprendizaje automático.

3.1. Extracción de características con *SURF*, agrupamiento con *k-means* y *BoW* y clasificación con *SVM*

El esquema de la figura 2 representa las fases de implementación de este método, que consta de dos fases: entrenamiento y prueba. En la fase de entrenamiento se extraen las características de las imágenes del conjunto de entrenamiento mediante el algoritmo de visión por computador *SURF*, que proporciona un detector y descriptor invarian-

te 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*. Cabe aclarar que las imágenes pertenecientes al conjunto de entrenamiento son distintas a las del conjunto de prueba.

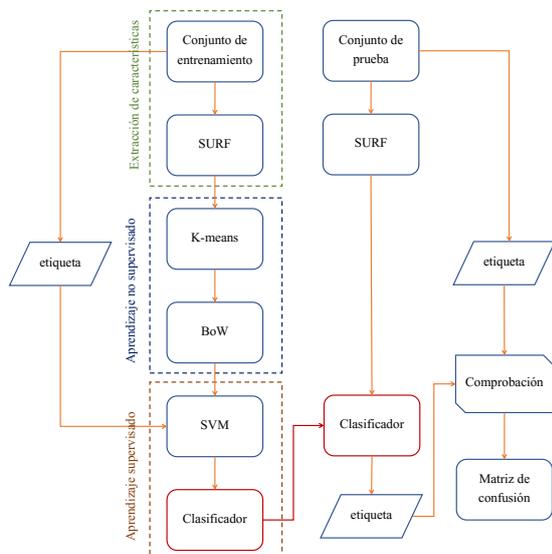


Figura 2: Esquema del método 1. Extracción de características con *SURF*, agrupamiento con *k-means* y *BoW* y clasificación con *SVM*.

Este clasificador permite identificar imágenes de presión dentro de unas clases pre-determinadas durante la fase de entrenamiento. Para evaluar el clasificador, en la fase de prueba se obtienen los descriptores de las imágenes del conjunto de pruebas mediante el descriptor *SURF* y se evalúa el método comparando la clase obtenida de cada imagen con la clase real que le corresponde, de forma que se puede obtener la matriz de confusión y el porcentaje de acierto.

Una vez evaluado el funcionamiento del método, se pueden identificar nuevas imágenes de presión, cuya clase es desconocida *a priori*, obteniendo los descriptores mediante *SURF* y utilizando el clasificador obtenido en la fase de entrenamiento y evaluado en la fase de pruebas. Así, se puede calcular el tiempo que se emplea en clasificar una nueva imagen, lo que sirve también como herramienta para evaluar el rendimiento.

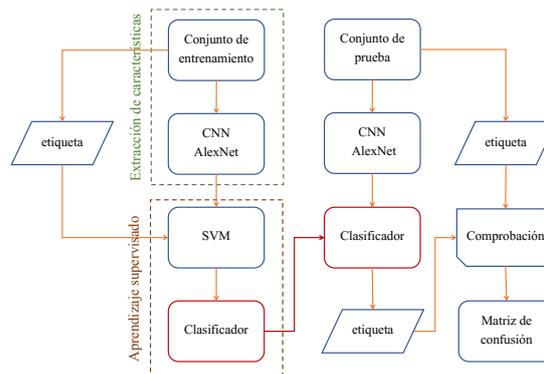


Figura 3: Esquema del método 2. Extracción de características con la *CNN AlexNet* y clasificación con *SVM*

3.2. Extracción de características con la *CNN AlexNet* y clasificación con *SVM*

Por otro lado, el segundo método emplea una red neuronal convolucional profunda (CNN), conocida como *AlexNet* [6]. La red se ha importado del repositorio de *caffe* [5], 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.

En concreto, el método desarrollado implementa una variante de esta red, aprovechando el concepto utilizado en [16]. 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 este sentido, se puede elaborar un esquema similar al del método 1 que describa el flujo de trabajo desarrollado. Este esquema se encuentra en la figura 3 y sigue las dos mismas fases del método 1: entrenamiento y prueba.

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 activaciones de la última capa previa a la clasificación, y se utilizan para realizar un entrenamiento supervisado con una *SVM* que genera un clasificador.

Este clasificador se evalúa extrayendo las características de las imágenes del conjunto de prueba y comparando las clases generadas con las clases conocidas. De manera que se puede obtener la matriz de confusión y el porcentaje de acierto del clasificador. Finalmente, el clasificador se puede utilizar para identificar nuevas imágenes de presión cuya

clase es desconocida a priori.

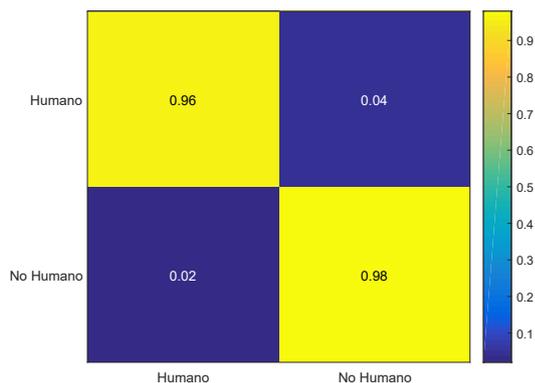


Figura 4: Matriz de confusión resultante de la evaluación del método 1 para clasificar imágenes de presión en las 2 clases *Humano* y *No humano*

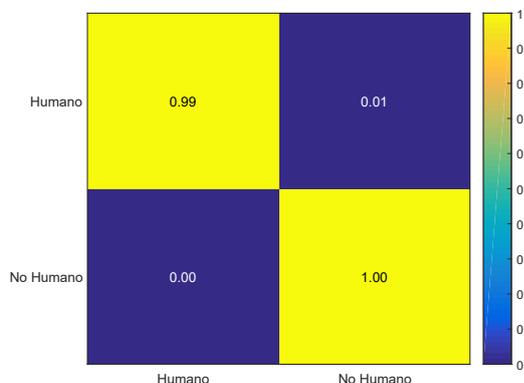


Figura 5: Matriz de confusión resultante de la evaluación del método 2 para clasificar imágenes de presión en las 2 clases *Humano* y *No humano*

4. EXPERIMENTOS Y RESULTADOS

4.1. Descripción

Para llevar a cabo los experimentos se ha utilizado el sensor táctil modelo 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/cm². Además, la presión máxima admitida es de 34kPa y el rango de temperatura de trabajo es -40°C a 60°C. El sensor está cubierto por una capa de caucho de silicona que sirve como elemento protector y conductor de fuerzas externas.

Para evaluar los métodos desarrollados para la aplicación de detección de personas en situacio-

nes de catástrofe, se han utilizado 300 imágenes etiquetadas en las clase *Humano*(150) y *No humano*(150). De las cuales 120 se han utilizado para la fase de entrenamiento, y 180 para la fase de pruebas, siendo la mitad de cada conjunto de una clase diferente. Hay que aclarar que las imágenes incluidas en la clase *Humano* se han escogido de las partes: mano, brazo y dedos, debido a que en una situación de catástrofe serían las más propensas a encontrarse desnudas. Por tanto, serían accesibles al robot para realizar la identificación. Además, todas las imágenes de presión etiquetadas en la clase *Humano* se han tomado de una misma persona. Por otro lado, los objetos utilizados en la clase *No humano* han sido un bolígrafo, unos alicates y unas tijeras.

En las imágenes de la figura 6 se pueden ver algunas de las imágenes de presión utilizadas, y etiquetadas en sus correspondientes clases.

4.2. Resultados

En la figura 4 se muestra la matriz de confusión resultante de la fase de evaluación del método 1. A partir de la cual se obtiene 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.

En cuanto al segundo método, se puede observar la matriz de confusión resultante en la figura 5. A partir de la cual se determina un índice de acierto del 99.44%. 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 1 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.

5. CONCLUSIONES Y TRABAJO FUTURO

Se ha presentado el desarrollo de un sensor táctil y su aplicación al campo de la robótica de rescate. Para ello, el sensor se ha instalado en un brazo manipulador ligero, con objeto de identificar el estado de víctimas potenciales en función de la urgencia de atención (triaje). Además, se han implementado dos métodos de inteligencia artificial para clasificar imágenes de presión. Cada método dispone de un extractor de características y un clasificador. El clasificador, en ambos métodos, se ha obtenido mediante aprendizaje supervisado utilizando una

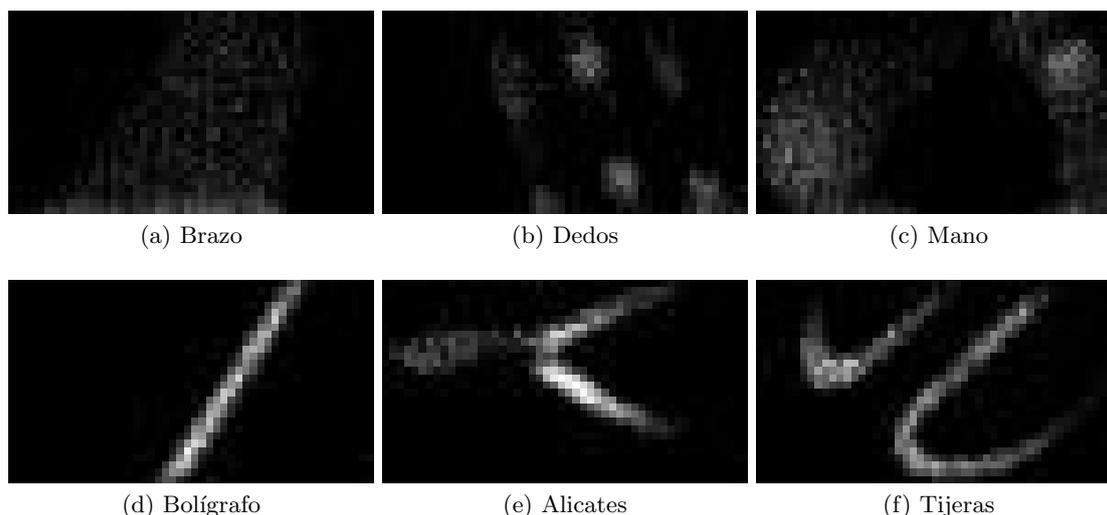


Figura 6: Ejemplos de imágenes de presión. Las imágenes superiores (a), (b) y (c) pertenecen a la clase *Humano* y las inferiores (d), (e) y (f) pertenecen a clase *No humano*

Método	Acierto (%)	Mejora (%)	Tiempo (s)
<i>SURF + SVM</i>	96.67	-	0.01
<i>CNN + SVM</i>	99.44	2.77	0.7

Tabla 1: Resumen de los resultados de los experimentos

SVM (Supported Vector Machine). En cuanto al extractor de características, el primer método incorpora el descriptor *SURF (Speeded-Up Robust Features)*, mientras que el segundo método utiliza una variante de la red neuronal convolucional conocida comúnmente como *AlexNet*. Para validar los métodos se han realizado experimentos de clasificación en las clases *Humano* y *No-humano*, y en subclases de estas. Finalmente, se han realizado los experimentos que sirven como comparativa de ambos métodos.

En este sentido, se ha comprobado que el método 2 obtiene una mejora del 2.77% con respecto al método 1. Sin embargo, el tiempo empleado en la clasificación de una nueva imagen de presión es mayor de un orden de magnitud en el método 2 (del orden de 0.7s) que en el primero (del orden de 0.01s). Es decir, pese a que el método 2 obtiene un índice de aciertos mayor que el primero, esta diferencia no resulta significativa en comparación la diferencia de tiempo de ejecución de ambos clasificadores. Por lo tanto, se puede concluir que, para llevar a cabo la clasificación entre humanos y objetos, resulta más apropiado el uso del método 1.

Actualmente se están llevando a cabo nuevos experimentos que permitan, no sólo distinguir si el contacto se realiza con un humano o no, si no ser capaces de detectar, en el caso de que el contacto se realice con un humano, la parte del cuer-

po concreta que se está tocando. Futuros trabajos consistirán en utilizar algoritmos basados en gradientes de presión y palpación activa. Asimismo, se pretende observar el comportamiento de ambos métodos en un entorno de desastre simulado en el que las subclases de objetos sean las típicas de estas situaciones, como piedras o ramas por ejemplo.

Agradecimientos

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

Referencias

- [1] Baishya, S. S. and Bauml, B. (2016). Robust material classification with a tactile skin using deep learning. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8–15. IEEE.
- [2] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359.
- [3] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.
- [4] García-Cerezo, A., Mandow, A., Martínez, J. L., Gómez-de Gabriel, J., Morales, J., Cruz,

- A., Reina, A., and Serón, J. (2007). Development of ALACRANE: A mobile robotic assistance for exploration and rescue missions. In *SSRR2007 - IEEE International Workshop on Safety, Security and Rescue Robotics Proceedings*, pages 1–6. IEEE.
- [5] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe. *Proceedings of the ACM International Conference on Multimedia - MM '14*, 1436:675–678.
- [6] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks.
- [7] Liu, Y. and Nejat, G. (2013). Robotic Urban Search and Rescue: A Survey from the Control Perspective. *Journal of Intelligent & Robotic Systems*, 72(2):147–165.
- [8] Lowe, D. (1999). Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pages 1150–1157. IEEE.
- [9] Luo, S., Liu, X., Althoefer, K., and Liu, H. (2015a). Tactile object recognition with semi-supervised learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9245, pages 15–26. Springer, Cham.
- [10] Luo, S., Mou, W., Althoefer, K., and Liu, H. (2015b). Novel Tactile-SIFT Descriptor for Object Shape Recognition. *IEEE Sensors Journal*, 15(9):5001–5009.
- [11] Luo, S., Mou, W., Althoefer, K., and Liu, H. (2016). Iterative Closest Labeled Point for Tactile Object Shape Recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.
- [12] Murphy, R. R., Tadokoro, S., Nardi, D., Jaffcoff, A., Fiorini, P., Choset, H., and Erkmén, A. M. (2008). Search and Rescue Robotics. In *Springer Handbook of Robotics*, pages 1151–1173.
- [13] Ranasinghe, A., Sornkarn, N., Dasgupta, P., Althoefer, K., Penders, J., and Nanayakkara, T. (2016). Salient Feature of Haptic-Based Guidance of People in Low Visibility Environments Using Hard Reins. *IEEE Transactions on Cybernetics*, 46(2):568–579.
- [14] Schmitz, A., Bansho, Y., Noda, K., Iwata, H., Ogata, T., and Sugano, S. (2014). Tactile object recognition using deep learning and dropout. *2014 IEEE-RAS International Conference on Humanoid Robots*, pages 1044–1050.
- [15] Vidal-Verdú, F., Barquero, M. J., Castellanos-Ramos, J., Navas-González, R., Sánchez, J. A., Serón, J., and García-Cerezo, A. (2011). A Large Area Tactile Sensor Patch Based on Commercial Force Sensors. *Sensors*, 11(12):5489–5507.
- [16] Wang, Y. and Cottrell, G. W. (2015). Bikers are like tobacco shops, formal dressers are like suits: Recognizing urban tribes with caffe. In *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015*, pages 876–883. IEEE.