



Herramienta para el Aprendizaje del Lenguaje Dactilológico Mediante Visión Artificial

Iris Aguilar, Antonio J. Reina, y Anthony Mandow

*Universidad de Málaga – Andalucía Tech
Departamento de Ingeniería de Sistemas y Automática, 29071 Málaga, Spain.
Email: ajreina@uma.es*

Abstract - El artículo aborda el uso de técnicas de visión artificial como herramienta didáctica para el aprendizaje de la lengua de signos para sordos. En particular, se propone una solución para el reconocimiento del alfabeto dactilológico con una cámara 2D. Con este fin, se presenta un conjunto de características visuales que permiten la clasificación de los signos obtenidos mediante un árbol de decisión. Este sistema de reconocimiento se ha integrado en una herramienta interactiva de aprendizaje que permite el dictado mediante gestos a la vez que puntúa, a modo de juego, la correcta realización de los signos. El artículo presenta resultados experimentales preliminares tanto del reconocimiento de signos como de su aplicación educativa.

Keywords: Computer vision, sign language recognition, visión por computador, reconocimiento de formas, descripción de imágenes, dactilología, educación

Citation Information:

Aguilar, I.; Reina, A. J.; Mandow, A.,
"Herramienta para el Aprendizaje del Lenguaje Dactilológico Mediante Visión Artificial,"
Actas de las XXXVI Jornadas de Automática, Bilbao, Spain, pp.127-133, Sep. 2015

```
@INPROCEEDINGS{Aguilar:JA2015,  
author={Iris Aguilar and Antonio J. Reina and A. Mandow},  
title={Herramienta para el Aprendizaje del Lenguaje Dactilológico Mediante Visión Artificial},  
booktitle={Actas de las XXXVI Jornadas de Automática},  
year={2015},  
pages={127-133},  
address={Bilbao, Spain}  
}
```

Herramienta para el Aprendizaje del Lenguaje Dactilológico Mediante Visión Artificial

Iris Aguilar, Antonio J. Reina, y Anthony Mandow
Universidad de Málaga, Andalucía Tech, Dpto. Ingeniería de Sistemas y Automática
ajreina@uma.es

Resumen

El artículo aborda el uso de técnicas de visión artificial como herramienta didáctica para el aprendizaje de la lengua de signos para sordos. En particular, se propone una solución para el reconocimiento del alfabeto dactilológico con una cámara 2D. Con este fin, se presenta un conjunto características visuales que permiten la clasificación de los signos obtenidos mediante un árbol de decisión. Este sistema de reconocimiento se ha integrado en una herramienta interactiva de aprendizaje que permite el dictado mediante gestos a la vez que puntúa, a modo de juego, la correcta realización de los signos. El artículo presenta resultados experimentales preliminares tanto del reconocimiento de signos como de su aplicación educativa.

Palabras clave: Visión por computador, reconocimiento de formas, descripción de imágenes, dactilología, educación.

1. Introducción

El reconocimiento de gestos constituye un problema de interés para aplicaciones como realidad virtual y aumentada [16], interacción con sistemas robóticos [7, 8, 10, 12] o interfaces humanos para aplicaciones informáticas [9, 11]. Asimismo, el reconocimiento del lenguaje de signos o dactilológico para sordos (véase la Figura 1) ha sido objeto de desarrollos orientados a la traducción y el aprendizaje [1, 6].

Para el reconocimiento de gestos de la mano, el uso de guantes sensorizados permite la captura realista de posturas [3], si bien la visión artificial puede ser adecuada para facilitar una interfase más natural al usuario [5]. El uso del Modelo Oculto de Markov ha servido para identificar el movimiento del lenguaje de signos a partir de imágenes sin necesidad de obtener información detallada de la posición de los dedos

de la mano [13, 18]. Otras soluciones se han basado en métodos como las redes neuronales o las redes dinámicas bayesianas [1]. También se ha propuesto la identificación de signos mediante la descomposición de formas, lo cual se puede conseguir con o sin modelos de referencia [2].

Respecto al tipo de cámaras, tanto las de profundidad (RGBD) como las de tiempo de vuelo (ToF) y estereoscópicas generan información tridimensional (3D) que resulta valiosa para el reconocimiento de gestos [14]. Especialmente relevante es el caso del sensor Kinect debido a su bajo coste y a la disponibilidad de *drivers* que facilitan su aplicación en tiempo real [15, 17]. No obstante, las cámaras monoculares presentan la ventaja de estar incluidas de serie en ordenadores personales, dispositivos móviles y televisores con funciones *smart*.

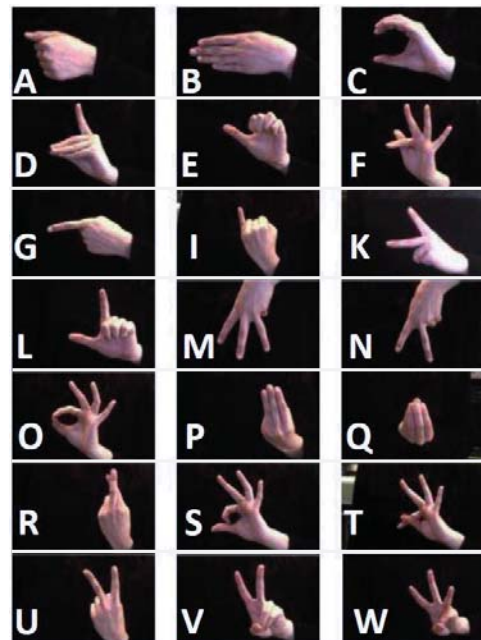


Figura 1: Signos estáticos del alfabeto dactilológico español.

Las herramientas para el aprendizaje del lenguaje de signos resultan útiles para sordos, su entorno familiar y otras personas interesadas en el estudio de esta forma de comunicación. En muchos casos, estas herramientas consisten en traductores entre el lenguaje escrito y una representación gráfica del lenguaje de signos [4]. Se han desarrollado herramientas educativas consistentes en juegos en tiempo real para la traducción de frases basadas en la posición de las manos [17] o multimodales empleando una combinación de movimientos de manos y cabeza [1]. También se ha contemplado el uso de robots humanoides como tutores [19], si bien esta solución es específica para niños con sordera y requiere una infraestructura más costosa.

En este artículo se propone una herramienta interactiva para el aprendizaje del alfabeto dactilológico. Esta aplicación se fundamenta en un sistema de reconocimiento basado en una cámara 2D para el que se ha definido un conjunto de características visuales que permiten la clasificación mediante un árbol de decisión. Así, la aplicación proporciona al usuario realimentación tanto en la realización de signos individuales como en la formación de frases.

El resto del artículo se organiza de la siguiente forma. La Sección 2 presenta los descriptores visuales y el método de reconocimiento propuesto. En la Sección 3 se describe la herramienta para el aprendizaje del alfabeto dactilológico. La Sección 4 ofrece resultados experimentales. Las conclusiones se presentan en la Sección 5.

2. Reconocimiento de Signos

Esta sección describe el procedimiento propuesto para el reconocimiento de signos dactilológicos mediante una cámara 2D. Este procedimiento consta de tres fases principales: i) adquisición y mejora de la imagen, donde se supone que aparece la mano suficientemente iluminada y distinguible del fondo; ii) obtención de características; y iii) reconocimiento mediante un árbol de decisión. Puesto que la primera fase se refiere a técnicas habituales en visión por computador, los epígrafes siguientes detallan las dos últimas fases.

El conjunto de signos que la aplicación es capaz de reconocer se muestran en la Figura 1, no habiéndose considerado aquellos signos que requieren de movimiento para su identificación.

2.1. Características para el Reconocimiento Dactilológico

La postura de la mano en el alfabeto dactilológico incluye signos con diferencias sutiles en su apariencia, como por ejemplo los signos F y T [6]. Así, la definición de un conjunto suficiente de características no resulta trivial. En este trabajo se propone un conjunto de siete características que se han obtenido empíricamente a partir de muestras de imágenes capturadas de distintos individuos. Estas características se describen a continuación:

Orientación (vertical u horizontal). Se trata de un descriptor binario en el que el número de píxeles predominante en las dos posibles direcciones del vector gradiente binarizado determina la orientación principal de la mano. La Figura 2 ilustra las discontinuidades en ambos ejes para un signo en el que domina la orientación horizontal. Existen posturas que son independientes de esta característica, puesto que la imagen correspondiente no posee una orientación dominante (por ejemplo, las letras S y O).

Grupos de dedos. El objetivo no es diferenciar los cinco dedos de la mano, sino el número de agrupaciones separadas de los mismos. Este dato se obtiene como el número de regiones conectadas separadas de la región central la cual se obtiene aplicando un operador de apertura morfológica. Dicho operador emplea un elemento estructura circular de diámetro aproximado al tamaño de los dedos en la imagen, con el objetivo de eliminarlos, siendo por tanto la diferencia entre la imagen binaria original y el resultado de la apertura la que contiene los grupos de dedos. La Figura 3 corresponde a un ejemplo con cuatro grupos de dedos.

Discontinuidades en la región central. Estas discontinuidades están presentes en mayor número en signos con dedos flexionados sobre la palma. Este descriptor, obtenido aplicando



Figura 2: Discontinuidades en los ejes horizontal (izquierda) y vertical (derecha) para la letra B.

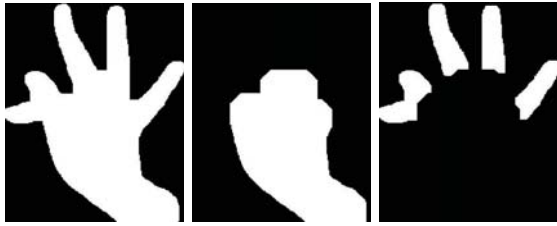


Figura 3: Ilustración de obtención de grupos de dedos para una imagen del signo F.



Figura 4: Discontinuidades de las regiones centrales de las letras U y V

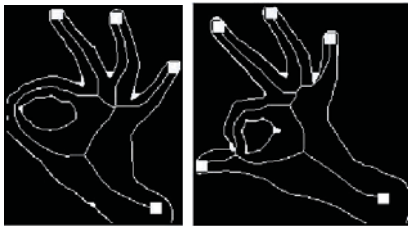


Figura 5: Contorno y esqueletos de las letras O y S, donde se han señalado los puntos finales.

un umbral, se aplica únicamente a la región central y permite distinguir entre una vista dorsal y una vista palmar de la mano. Así, por ejemplo, es posible diferenciar las letras U y V gracias a las discontinuidades presentes en esta última (véase la Figura 4).

Presencia de huecos. La identificación de huecos en la región conectada correspondiente a la mano permite identificar las letras O, S, y T.

Número de puntos finales del esqueleto. El esqueleto se obtiene mediante adelgazamiento morfológico. Los puntos finales del mismo representan las puntas de los dedos extendidos así como la muñeca (véase la Figura 5).

Diferencia convexa. El número de regiones conectadas de la diferencia convexa permite identificar concavidades en la postura de la mano. Previamente a la obtención de este descriptor



Figura 6: Ilustración de regiones cóncavas del signo S junto con la imagen de partida.

se han eliminado los huecos presentes en la región original. En la Figura 6 se muestra el resultado de esta operación para el signo S, junto con la imagen de partida antes de la eliminación del hueco.

Excentricidad. Es una medida que determina el grado de desviación de una sección cónica con respecto a una circunferencia, y permite determinar la agrupación de los dedos en torno a la palma de la mano. Su valor está comprendido entre 0 y 1, donde una excentricidad próxima a uno se corresponde con el puño cerrado, como ocurre con las letra A y E.

2.2. Árbol de decisión

Las siete características seleccionadas para la imagen de la mano permiten realizar el reconocimiento de la postura de la misma mediante un árbol de decisión, donde cada hoja representa uno de los signos del alfabeto. De esta manera, el reconocimiento se realiza sin comparar la imagen con un modelo de referencia.

El árbol completo posee un total de 22 hojas correspondientes a la totalidad de signos estáticos del alfabeto dactilológico español (ver Figura 1) mas un gesto para representar el espacio en blanco consistente en la mano abierta con los cinco dedos separados. Por motivos de espacio, la Figura 7 ofrece una versión reducida del árbol de decisión para un alfabeto compuesto únicamente por las vocales. En este árbol se han incluido las siete características lo cual permite identificar las posturas que no se corresponden con ninguna de las cinco vocales. También se puede apreciar cómo algunos signos pueden ser independientes de ciertas características, como la orientación en el caso de la letra O.

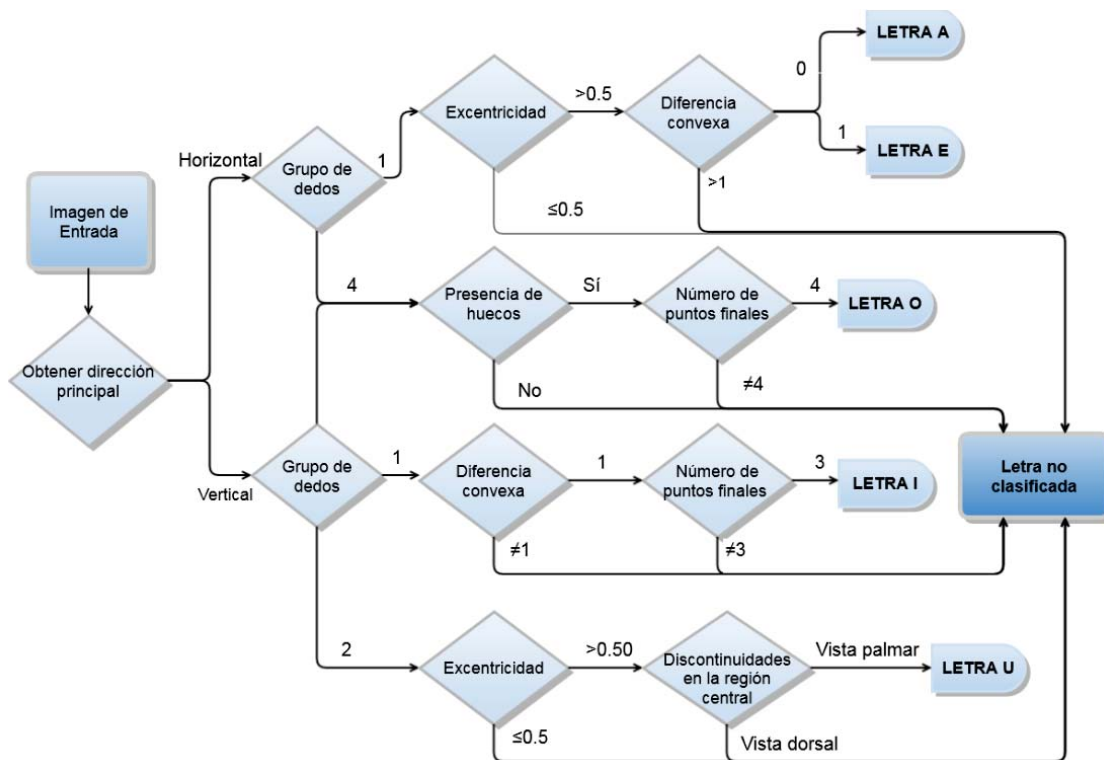


Figura 7: Ejemplo de árbol de decisión simplificado: Reconocimiento de las vocales.

3. Aplicación para la ayuda al aprendizaje del alfabeto de signos

Esta aplicación se ha desarrollado íntegramente en MatLab y hace uso de la toolbox para el procesamiento de imágenes y de su interfaz de usuario para el desarrollo de herramientas gráficas (GUIDE). La aplicación ofrece dos funcionalidades: un juego para el aprendizaje (Figura 8) y un traductor de signos dactilológicos a caracteres (Figura 9).



Figura 8: Interfaz gráfica de la herramienta de aprendizaje



Figura 9: Interfaz gráfica de la herramienta de traducción

3.1. Juego: Aprendizaje del Alfabeto

Se presentan dos formas de aprendizaje: una primera más básica en la que se muestra, a modo de ayuda, una imagen del signo que se debe formar, y un segundo modo en el que solo se especifica la letra. Estas dos fases facilitan que el usuario aprenda de forma gradual como se forman cada uno de los caracteres dactilológicos.

Ambas formas de aprendizaje incluyen 3 pasos: en el primero se solicita al usuario que forme una deter-

minada letra del alfabeto dactilológico (seleccionada al azar); en el segundo, se realiza la captura de la imagen en tiempo real; y finalmente, se premia o penaliza el gesto realizado.

En la Figura 8 se muestra el interfaz de usuario de esta aplicación, en la que se pueden apreciar dos ventanas que contienen imágenes. En el modo básico la segunda ventana se emplea para mostrar la imagen al signo a formar ya gesticulado, mientras que la primera ventana muestra la imagen captada por la cámara. En el segundo modo de funcionamiento solo se muestra la imagen que esta siendo capturada.

3.2. Traductor de signos

El traductor de signos es una herramienta adicional que sirve para avanzar en el aprendizaje, permitiendo al usuario escribir frases completas empleando el alfabeto dactilológico. Esta herramienta va mostrando los signos reconocidos en secuencia, si bien no se realiza ningún análisis sintáctico de la escritura realizada.

La Figura 9 muestra la interfaz de usuario de este traductor, donde se aprecia el reconocimiento de todos los signos representados en la Figura 1.

4. Resultados Experimentales

Esta sección presenta resultados preliminares del sistema propuesto. Con este fin, se analiza el rendimiento del método desarrollado para el reconocimiento de signos y se valora la aplicación de aprendizaje.

4.1. Condiciones experimentales

Para la captura de imágenes se ha empleado una cámara industrial USB 2 uEye con un sensor CMOS a la que se le ha dotado de una lente con distancia focal y apertura variables. Se trata de una cámara pequeña, compacta y robusta, con un interfaz de usb que hace que su conexión sea simple.

Como parte adicional del equipo de adquisición se ha incorporado un sistema de iluminación frontal, consistente en dos anillos de diodos led. Los objetivos de este sistema de iluminación son el de reducir la presencia de sombras y el hacer al sistema robusto frente a variaciones de luz. La Figura 10 muestra el conjunto cámara/lente/iluminación.



Figura 10: Disposición de la cámara y sistema de iluminación

4.2. Reconocimiento de signos

Los resultados preliminares del algoritmo de clasificación han alcanzado una tasa de aciertos (clasificaciones correctas) superior al 80% sobre un total de 620 imágenes. Esta tasa de aciertos incluye como tales la no clasificación de signos que no se corresponden con ninguno de los presentados en la Figura 1. Los signos que han tenido una tasa de acierto menor han sido la V y la C, que dependen de la detección de discontinuidades en la zona central de la mano, siendo esta muy sensible a las condiciones de iluminación.

De forma adicional se han realizado pruebas con diversas cámaras comerciales de inferiores prestaciones y costes, habiéndose obtenido resultado similares bajo las mismas condiciones de iluminación.

4.3. Aprendizaje del alfabeto

El uso del aprendizaje visual propuesto para alfabeto dactilológico ha permitido que personas sin ningún conocimiento previo del lenguaje de signos fuesen capaces de realizar de forma correcta la gesticulación de los diversos caracteres. El tiempo empleado en aprender cada signo en particular estuvo directamente relacionado con la complejidad del mismo, así pues, para la correcta realización de las letras A, por ejemplo, fue necesaria una única prueba, habiéndose revelado como las complicadas de aprender las letras F y T.

5. Conclusiones y trabajos futuros

Este trabajo se ha orientado a la aplicación de técnicas de visión artificial como herramienta didáctica para el aprendizaje de la lengua de signos para sordos. En particular, se ha propuesto una solución para el reconocimiento del alfabeto dactilológico con una cámara 2D. Para ello, se ha definido un conjunto de siete características visuales que permiten la clasificación de los signos obtenidos mediante un árbol de decisión sin necesidad de utilizar imágenes de referencia. Este sistema de reconocimiento se ha integrado en una herramienta interactiva de aprendizaje que permite el dictado mediante gestos a la vez que puntúa, a modo de juego, la correcta realización de los signos.

La orientación de la mano y el número de grupo de dedos han sido las características primordiales para una primera clasificación. Dentro de esta diferenciación ha sido necesaria la incorporación de cinco descriptores más para poder llevar a cabo con éxito la clasificación de signos de similar apariencia. Con esta solución se logra reconocer la totalidad de los signos estáticos del alfabeto dactilológico español, incluyendo aquellos que presentan diferencias sutiles entre ellos, como es el caso de la F y la T.

Como trabajo futuro se considera ampliar el sistema para incluir la clasificación de las letras que se gesticulan con movimiento. Además, se pretende mejorar la aplicación para hacerla más robusta frente a condiciones no controladas de iluminación y posibles variaciones del fondo de la imagen, y el desarrollo de una versión para dispositivos móviles.

Referencias

- [1] O. Aran, I. Ari, L. Akarun, B. Sankur, A. Benoit, A. Caplier, P. Campr, A. H. Carrillo, and F. X. Fanard. Signtutor: An interactive system for sign language tutoring. *IEEE Multimedia*, 16(1):81–92, 2009.
- [2] J. Choi and J. I. Park. Hand posture recognition using shape decomposition. In *IEEE International Symposium on Virtual Reality Innovations*, pages 347–348, 2011.
- [3] S. Cobos, M. Ferre, M.A. Sánchez-Urán, J. Ortego, and R. Aracil. Human hand descriptions and gesture recognition for object manipulation. *Computer Methods in Biomechanics and Biomedical Engineering*, 13(3):305–317, 2010.
- [4] N. Ditcharoen, K. Naruedomkul, and N. Cercone. SignMT: An alternative language learning tool. *Computers and Education*, 55(1):118–130, 2010.
- [5] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73, 2007.
- [6] I. García Incertis, J. Gómez García-Bermejo, and E. Zalama Casanova. Hand gesture recognition for deaf people interfacing. In *International Conference on Pattern Recognition*, volume 2, pages 100–103, 2006.
- [7] V. Gonzalez-Pacheco, M. Malfaz, F. Fernandez, and M. A. Salichs. Teaching human poses interactively to a social robot. *Sensors (Basel, Switzerland)*, 13(9):12406–12430, 2013.
- [8] Y. Lei, W. Hongpeng, T. Dianxiong, and W. Jue. A real-time hand gesture recognition algorithm for an embedded system. In *IEEE International Conference on Mechatronics and Automation*, pages 901–905, 2014.
- [9] A. Melero Gil, I. García Incertis, J. Gómez García-Bermejo, and E. Zalama Casanova. Seguimiento y reconocimiento de los gestos de una mano mediante visión por computador. In *XXVIII Jornadas de Automática*, 2007.
- [10] S.E. Nope, H. Loaiza, and E. Caicedo. Modelo bio-inspirado para el reconocimiento de gestos usando primitivas de movimiento en visión. *Revista Iberoamericana de Automatica e Informatica Industrial*, 5(4):69–76+95, 2008.
- [11] S. S. Rautaray and A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2012.
- [12] A. Sanfeliu, J. Andrade-Cetto, M. Barbosa, R. Bowden, J. Capitán, A. Corominas, A. Gilbert, J. Illingworth, L. Merino, J.M. Mirats, P. Moreno, A. Ollero, J. Sequeira, and M. T. J. Spaan. Decentralized sensor fusion for ubiquitous networking robotics in urban areas. *Sensors*, 10(3):2274–2314, 2010.
- [13] T. Starner and A. Pentland. Real-time american sign language recognition from video using hidden markov models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 265–270, 1995.

- [14] J. Suarez and R. R. Murphy. Hand gesture recognition with depth images: A review. In *IEEE International Workshop on Robot and Human Interactive Communication*, pages 411–417, 2012.
- [15] H. D. Yang. Sign language recognition with the kinect sensor based on conditional random fields. *Sensors*, 15(1):135–147, 2015.
- [16] M. T. Yang and W. C. Liao. Computer-assisted culture learning in an online augmented reality environment based on free-hand gesture interaction. *IEEE Transactions on Learning Technologies*, 7(2):107–117, 2014.
- [17] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti. American sign language recognition with the kinect. In *International Conference on Multimodal Interfaces*, 2011.
- [18] Z. Zafrulla, H. Brashear, P. Yin, P. Presti, T. Starner, and H. Hamilton. American sign language phrase verification in an educational game for deaf children. In *International Conference on Pattern Recognition*, pages 3846–3849, 2010.
- [19] A. Özkul, H. Köse, R. Yorganci, and G. Ince. Robostar: An interaction game with humanoid robots for learning sign language. In *IEEE International Conference on Robotics and Biomimetics*, pages 522–527, 2014.