



INVESTIGA I+D+i 2012/2013

GUÍA ESPECÍFICA DE TRABAJO SOBRE "PERCEPCIÓN ARTIFICIAL"

Texto de D. Francisco Serradilla García

Octubre de 2012

Parece fácil definir la percepción artificial, mejor llamada percepción computacional: es la percepción hecha por máquinas. Pero, ¿qué es la percepción?

Según el diccionario de la Real Academia la percepción es "sensación interior que resulta de una impresión material hecha en nuestros sentidos", es decir, lo que nuestra mente produce a partir de lo que llega a nuestros órganos sensoriales.

En este sentido tenemos que distinguir entre varias cosas:

1. El objeto en sí que produce la información que percibimos.
2. La información que recibe el órgano sensorial, es decir, el estímulo.
3. La traducción de la información recibida a estímulos eléctrico-químicos que son propagados por las neuronas sensoriales hasta el cerebro.
4. La interpretación que hace el cerebro de esos estímulos.

La primera de ellas pertenece al ámbito de la filosofía y, aunque hay asuntos interesantes que podemos plantearnos sobre ella, en general no afecta al procesamiento realizado para percibir, ni en los humanos ni en las máquinas.

En cuanto a las segunda, la información que recibe el órgano sensorial puede ser de varios tipos: luminosa, acústica, olfativa, táctil o de sabor, en el caso humano, pero en el caso de las máquinas (y de algunos animales) puede ser también de otras procedencia, como ultrasonidos, temperatura, magnética o de distintos tipos de radiaciones, por poner algunos ejemplos.

La tercera de ellas, la traducción de la información a algo manipulable por el sistema, y que en el humano es realizada por los órganos de los sentidos, en las máquinas se lleva a cabo mediante dispositivos transductores, que son componentes electrónicos que convierten la información de origen en corrientes eléctricas que posteriormente son digitalizadas para que puedan ser utilizadas en un programa de ordenador.

Por ejemplo, un micrófono es un dispositivo que transforma ondas acústicas en electricidad gracias a una membrana (diafragma) que al moverse por la vibración del aire genera una corriente eléctrica por algún mecanismo físico, por ejemplo, mediante un condensador. Esa corriente eléctrica es muestreada posteriormente (medida a intervalos regulares y cada medida representada por un número) produciendo una secuencia de números que serán tratados por un programa de ordenador.

La misión de este programa corresponde al punto cuarto, es decir, la interpretación de esos números para identificar el objeto o la situación que produjo el estímulo.

Los humanos somos capaces de hacer (de modo casi inconsciente) muchas tareas con la información sensorial que percibimos. Por poner algunos ejemplos continuando con el caso del sonido, podemos extraer el mensaje que transporta ese sonido cuando hablamos con alguien, podemos saber quién nos está hablando, percibir una melodía en una canción, sentir el ritmo o acudir en la ayuda de alguien que grita. Todo eso va cifrado en el sonido y, aunque para nosotros es muy sencillo extraerlo de la señal sonora, para una máquina es una tarea complejísima, seguramente más compleja que jugar al ajedrez o resolver un problema matemático.

Curiosamente en los orígenes de la Inteligencia Artificial se creyó que todo esto no representaba un problema, precisamente porque a los humanos nos cuesta mucho resolver problemas o jugar bien al ajedrez, y nos cuesta muy poco saber quién nos está hablando o reconocer a una persona en una fotografía, pero cuando los científicos intentaron escribir programas que lo hicieran se encontraron con enormes problemas teóricos y prácticos.

En la actualidad nos conformamos con hacer programas que cubren algún aspecto muy concreto y delimitado del problema, por ejemplo, hay sistemas que inspeccionan imágenes tomadas con una cámara de piezas para medir su tamaño y detectar una pieza defectuosa, o hay programas, como el asistente Siri, de Apple, que son capaces de entender lo que le decimos y actuar en consecuencia, aunque de un modo algo primitivo y con cierta probabilidad de error.

Usuario: "Despiértame a las ocho en punto".

Siri: "He programado una alarma para las 8:00".

Usuario: "Cámbiala a las 8:30".

Siri: "He cambiado la alarma para las 8:30".

Usuario: "Gracias".

Siri: "Mi mayor satisfacción es que estés contento".

Ejemplo de conversación con Siri

Estos programas requieren el desarrollo de algoritmos muy sofisticados y con un enorme coste computacional. El trabajo de Siri, por ejemplo, no lo hace el teléfono, sino que la señal de voz es enviada a servidores de la compañía en los que se traduce el sonido a texto y se determina cuál es la respuesta adecuada. Esta respuesta entonces es enviada de vuelta al teléfono para que este ejecute la acción.

En general, el problema de la percepción computacional se descompone en varias fases y se aborda cada una de ellas por separado. Estas fases son las siguientes:

- Captación.
- Preproceso.
- Segmentación.
- Descripción.
- Reconocimiento.
- Actuación.

Captación

Como se ha comentado, en esta fase el dispositivo transductor convierte la información en señal eléctrica y esta es digitalizada.

Preproceso

En esta fase se mejora la información ya digitalizada. Por ejemplo si es sonora se intenta eliminar el ruido o si el sonido es débil se intentar incrementarlo. En

esta fase también se pueden aplicar filtros para eliminar parte de la señal que no nos interese.

Segmentación

Es una de las fases más importantes. Aquí se trata de separar "la figura del fondo", es decir, el objeto que queremos reconocer de todo lo demás. En una foto, por ejemplo, podría consistir en separar la cara de una persona del resto de la foto: cuerpo, paisaje, etc., o en una canción separar la melodía del acompañamiento.

Descripción

Una vez segmentado el objeto de interés, en la fase de descripción se intenta representar este objeto utilizando menos información, y si es posible más relevante que la original. Por ejemplo, en el caso de la melodía podríamos intentar representarla por la secuencia de notas musicales que suenan en ella, o en el caso de la fotografía podríamos intentar representar la cara por números como la altura de los ojos, la distancia que los separa, la posición de la nariz, el tipo de nariz, etc.

Reconocimiento

Una vez que tenemos la descripción del objeto a reconocer, estamos en condiciones de intentar reconocerlo, es decir, saber qué objeto es: saber qué persona es la que está en la fotografía, saber a qué canción pertenece la melodía que está sonando, etc. Otras veces el objetivo no es tan ambicioso. Quizá nos baste con determinar si en la pieza que estoy explorando hay una grieta, por ejemplo, o si la medida del lado de una tuerca que se ha fabricado es correcta o no. Esta fase depende muchísimo del objetivo final que tenga el sistema que estemos desarrollando.

Actuación

Generalmente, una vez reconocido el objeto, en sistema debe tomar algún tipo de decisión, más o menos compleja. A modo de ejemplo, si un sistema de ayuda a la conducción ha determinado que hay un peatón en la calzada y que el automóvil va a una velocidad en la que podría atropellarlo, puede tomar la decisión de frenar automáticamente. Si una pieza es defectuosa tomará la decisión de enviarla a la zona de piezas defectuosas, etc.

Como hemos visto, la Percepción Computacional es un área de investigación, pero también de desarrollo tecnológico con infinidad de aplicaciones prácticas en el mercado, con productos tanto de uso industrial como de uso cotidiano, y que se encuentra a caballo de varias disciplinas, como son la Ingeniería Industrial, la Informática y la Física, entre otras.

Bibliografía

D. Maravall "Reconocimiento de formas y visión artificial". RAMA. 1993.

A. De la Escalera: "Visión por computador. Fundamentos y métodos". Prentice Hall, 2001.

José Francisco Vélez Serrano, Ana Belén Moreno Díaz, Ángel Sánchez Calle, José Luis Esteban Sánchez-Marín. "Visión por computador". Universidad Rey Juan Carlos. En <http://www.terra.es/personal/jfvelez/libro2/libro.html>