

Interfaz Natural de Usuario para el Control de Robot Móvil con Gestos Faciales y Movimientos del Rostro Usando Cámara RGB

César Osimani¹, Martín Salamero², Carlos Bartó³, Marcos Lopez⁴

Centro de Investigación Aplicada y Desarrollo en Informática y Telecomunicaciones (CIADE-IT)
Universidad Blas Pascal (UBP) - Córdoba, Argentina

¹cosimani@ubp.edu.ar, ²msalamero@ubp.edu.ar, ³carlosbarto@gmail.com, ⁴marcoslopez@ubp.edu.ar

Resumen

Actualmente existen muchos esfuerzos por parte de grupos de investigación en acelerar el desarrollo de algoritmos y técnicas que faciliten la cotidianidad de las personas a través de una interacción más natural con los dispositivos. La disciplina que busca mejorar la experiencia del usuario en el manejo de computadoras se la conoce como Interacción Humano Computadora (HCI - Human Computer Interaction), y en algunos de sus campos se pueden encontrar las Interfaces Cerebro Computadora (BCI - Brain Computer Interface) que permiten una interacción mediante nuestro pensamiento, realizando la adquisición de las ondas cerebrales, que luego son procesadas para generar las acciones en el entorno. También se encuentran las Interfaces Naturales de Usuario (NUI - Natural User Interface) con las cuales se ofrece la posibilidad de control a través de gestos, posturas y movimientos con el cuerpo.

La utilización de recursos de visión artificial para reconocer los movimientos y gestos del rostro de

también hace hincapié en la Interacción Humano Robot (HRI - Human RobotInteraction) para implementar las técnicas de visión artificial en el control del desplazamiento de un robot móvil.

Palabras clave: Visión Artificial; Interfaz Natural de Usuario; Reconocimiento de expresiones faciales; Control de robots

Contexto

La línea de investigación que se trata en este proyecto es el desarrollo de Interfaces Naturales de Usuario con el uso de técnicas de Visión Artificial. Se realiza en el marco de las actividades desarrolladas en el Centro de Investigación Aplicada y Desarrollo en Informática y Telecomunicaciones (CIADE-IT), dependiente de la Universidad Blas Pascal, específicamente en lo relativo al proyecto "Visión Artificial para la Interacción Natural Humano Computadora a través de los gestos del rostro" aprobado por el Ministerio de Ciencia y Tecnología del Gobierno de la Provincia de Córdoba en el llamado a presentación de proyectos 2015 para Grupos de Reciente creación de investigación científica y tecnológica (GRUPT) en la Resolución Nro.0045 del 12 de agosto de 2016.

Cabe destacar que se mantiene un proyecto en esta línea con investigadores de la Universidad de Almería (España) con quienes, a la fecha, se tiene la aceptación del artículo "Hand Posture Recognition with standard webcam for Natural Interaction" en el congreso WorldCist'17 (5th World Conference on Information Systems and Technologies).

El objetivo de este proyecto, además de contribuir con el avance de técnicas para el reconocimiento de gestos del rostro, es diseñar una interfaz de usuario compuesta por una pantalla de visualización y una cámara RGB estándar que permita controlar dispositivos a través de los movimientos de la cabeza y los gestos faciales, emulando gráficamente en la pantalla un conjunto de botones de control para facilitar al usuario reconocer con exactitud el comando que se está realizando. Este trabajo

Introducción

Para la creación de interfaces naturales contécnicas de visión artificial se requiere la elección de bibliotecas de programación para el procesamiento de las imágenes y el diseño de interfaces gráficas. Este grupo de trabajo tiene antecedentes en esta línea de investigación con el uso de las bibliotecas OpenCV (ver [1] y [2]) y Qt bajo lenguaje C++ (documentación en [3]). Se comienza adquiriendo las imágenes de video de una cámara RGB estándar. Cada una de las imágenes es analizada con uno de los métodos más utilizados, desarrollado por Viola & Jones [4], para la detección de rostros. Es un método estadístico que utiliza muestras de entrenamiento (imágenes con rostros e imágenes sin rostros) para extraer información que permite distinguir un rostro de aquello que lo no es. Este método también puede ser utilizado para la detección de ojos, bocas sonrientes o no, orejas, como también puede ser extendido para detectar otro tipo de objetos (aviones, autos, etc.). Para utilizar este método se realiza un entrenamiento con las imágenes de muestra, para así obtener un clasificador del objeto. Este tipo de clasificador entrega una región de interés (ROI - Region Of Interest) que es un rectángulo dentro de la imagen que contiene el objeto en cuestión.

Es importante indicar que un prototipo ya se encuentra desarrollado por este grupo de investigación, el cual permite controlar un menú de opciones a través de la orientación del rostro. El código fuente y un video demostrativo se encuentran en [5] y [6].

El control del menú se realiza con un clasificador de rostros (disponible en OpenCV) que identifica su posición dentro de la imagen y además reconoce su orientación a partir de un leve desplazamiento que sufre la región detectada cuando el rostro cambia su orientación. Esta característica se trata en [7].

OpenCV dispone de un clasificador para la sonrisa que genera un alto porcentaje de falsos positivos cuando se analiza una imagen que contiene diversos objetos. Sin embargo, da muy buenos resultados cuando la región de la imagen que se analiza es reducida y forma parte del rostro. Por este motivo, es altamente recomendable que en un primer paso se detecte el rostro, para luego dividir esta región de interés por la mitad (obteniendo una región

donde se encuentran los ojos y otra región con la boca). Por último, analizar la región inferior con el clasificador de la sonrisa para detectar si la persona está sonriendo o no. De esta forma, la sonrisa puede ser utilizada como acción de selección dentro de la interfaz natural de usuario. Para completar la interfaz natural de usuario con otra alternativa para realizar acción de selección, se puede optar por detectar la apertura de la boca bajando la mandíbula. Para ello, el uso de un clasificador que identifique la boca abierta es una opción. De esta forma, la interfaz quedaría preparada para interpretar lo siguiente:

- Cambios de orientación del rostro.
- Sonrisa y apertura de boca.

Entre los distintos usos que puede tener una interfaz natural de usuario podemos mencionar aquellos recursos tecnológicos para personas con problemas en su movilidad o que tengan limitaciones para la operación de aquellos dispositivos con los tradicionales mecanismos de control. Otro uso para este tipo de interfaz es el control de robots del tipo plataformas móviles, no sólo para el control de su desplazamiento sino también la posibilidad de controlar algún accesorio que pudiera estar montado sobre el robot, como por ejemplo, una cámara de video.

Líneas de Investigación y Desarrollo

Existen diversos trabajos en el estado del arte que estudian los mecanismos de control de robots a través de una interacción con el usuario por medio de la detección del rostro. En [8] se realiza la segmentación del rostro con la detección del color de la piel y un análisis geométrico para identificar la región del rostro. Luego se utiliza el algoritmo SURF (Speeded- Up Robust Features) para extraer descriptores visuales para por último comprobar la consistencia geométrica a través del método iterativo RANSAC (Random sample consensus) contra las imágenes de rostros de una base de datos.

El trabajo [9] utiliza el método de Viola &

Jones para detectar el rostro y embebe este algoritmo en un FPGA (Field Programmable Gate Array), logrando realizar el procesamiento de 30 cuadros por segundo. Esto lo hace apto para interfaces de comunicación con el usuario en sistemas embebidos montados en un robot.

Otros trabajos, tal como [10], también utilizan el método de Viola & Jones para detectar el rostro y es apto para ser ejecutado en un computador Raspberry Pi (creado por Raspberry Pi Foundation [11]).

Partiendo del prototipo [6] ya desarrollado por este grupo de investigación, el alcance del proyecto incluye lo siguiente:

- Incorporar detección de sonrisa y apertura de la boca para ejecutar acciones.
- Construcción de un robot móvil comandado de manera remota con los movimientos y gestos del rostro.
- Diseñar una interfaz gráfica de usuario en el que se visualicen botones de control.

El sistema completo consta de dos programas informáticos: uno para ser ejecutado en una laptop para comandar de manera remota al robot móvil, recibir las imágenes desde una cámara montada en el mismo y reconocer los movimientos y gestos del rostro; por otro lado, un programa para ser ejecutado en un computador Raspberry Pi para ocuparse de los motores y controlar la cámara que redirige las imágenes a la computadora remota.

Para el diseño de la interfaz gráfica se utilizará lenguaje C++ con la biblioteca Qt para visualizar, en la pantalla de la computadora, las imágenes que son recibidas desde el robot. La interfaz despliega botones de control, como se observa en la Figura 1. Estos botones pueden ser accionados con el mouse, con el teclado o a través de los gestos del rostro, formando con esta característica la Interfaz Natural de Usuario.



Figura 1. Botones de control del robot móvil y la cámara

Mediante los movimientos del rostro, se podrá controlar el desplazamiento del robot y accionar los botones u optar por el control del movimiento de la cámara, que dispondrá de dos grados de libertad (horizontal y vertical). Cabe destacar en este punto que, según la extensión de la ley de Fitts [12], el tiempo requerido para alcanzar un objetivo (en este caso, un botón) decrece cuando se incrementa el tamaño del mismo. Por ello se proponen botones de gran tamaño y en transparencia con el video a fin de reducir el tiempo de selección de los botones. En la interfaz se resaltaré un botón cuando el usuario oriente su cabeza y mirada hacia dicho botón. Para ejecutar la acción asociada al botón, se deberá abrir la boca. De esta forma, si el usuario se posiciona sobre el botón de avance y abre la boca, el robot avanzará mientras se mantenga la boca abierta. Si el usuario sonríe, se podrá alternar entre las siguientes tres posibilidades: controlar el desplazamiento del robot, controlar la orientación de la cámara o dejar de controlar el robot. Como se observa en la Figura 1, los botones de control cambiarán, como también lo hará el ícono mostrado en la esquina superior derecha, para que se pueda identificar sobre qué dispositivo se está realizando el control.

El robot móvil se confeccionará a partir del chasis provisto por la empresa Makerfire (Figura 2). Es una plataforma plana con cuatro motores de corriente continua. Al chasis se le incorporará la electrónica de control y se le suministrará energía mediante una batería de gel de 6 volts. Para permitir lograr que los motores giren en ambos sentidos, se utilizarán circuitos electrónicos Puente H, controlados con una placa Intel Galileo [13] que genera una señal tipo PWM proporcional a la velocidad de desplazamiento del móvil. El chasis también tendrá montado una cámara de video en un

soporte con dos servomotores (ver Figura 3). La placa Intel Galileo controla el movimiento de los dos servomotores para cambiar la elevación y el ángulo horizontal de la cámara. Adicionalmente, un sensor ultrasónico para medir proximidad colocado en el frente del robot actuará como un interruptor de emergencia cuando el robot se dirija hacia algún obstáculo, evitando la colisión.

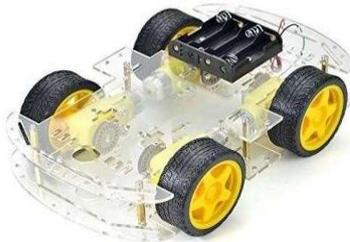


Figura 2. Chasis del robot Makerfire

La computadora Raspberry Pi montada en el robot dispone de pines GPIO [14] (General Purpose Input/Output) y a través de ellos se realizará la comunicación con la placa Intel Galileo. Para la comunicación inalámbrica desde el programa ejecutado en la laptop y el programa en la Raspberry se utilizará un punto de acceso de red inalámbrica basado en el estándar IEEE 802.11. La biblioteca Qt dispone de recursos para utilizar el protocolo TCP para el envío de datos, que posibilita tanto la comunicación para el control del robot como para la transmisión de las imágenes.

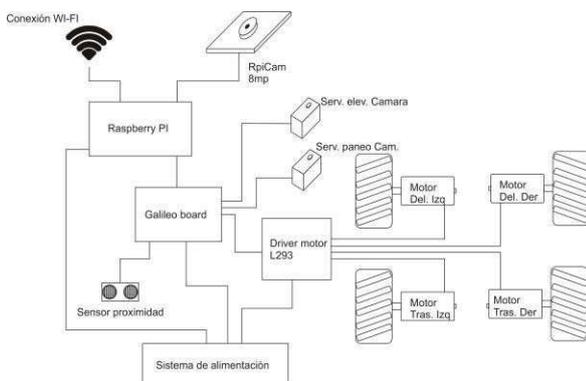


Figura 3. Diagrama en bloques del robot móvil

Resultados y Objetivos

Los objetivos de desarrollo se centran en el diseño de las herramientas para acercarnos a una interacción natural con las máquinas que nos rodean. El desarrollo de un robot móvil

controlado con el movimiento del rostro servirá como prueba de concepto acerca de la viabilidad y usabilidad de este tipo de tecnología para un usuario con movilidad reducida (o usuarios en general). Como se puede apreciar en [15], es posible su uso para el control de una silla de ruedas eléctrica.

Con el desarrollo de una interfaz natural, como la propuesta en este trabajo, facilitaría una mayor integración para personas que tienen dificultades en la comunicación oral. El desarrollo de un producto de apoyo que utilice técnicas de Comunicación Aumentativa y Alternativa, como por ejemplo ARASAAC [16], en donde el usuario, mediante pictogramas, genera una oración con la selección de un número reducido de imágenes, permitiendo expresarse mejorando su calidad de vida y autonomía personal.

Para lograr estos objetivos de desarrollo se realizarán las siguientes tareas, en tres momentos del proceso de diseño de la interfaz:

- **Antes de comenzar el diseño final:** Observación de distintos usuarios interactuando con el prototipo ya desarrollado [6] para conocer sus necesidades en el contexto real.
- **Durante la incorporación de detección de sonrisa y boca abierta:** Estas pruebas entregarán datos importantes para identificar una elaboración poco clara de la interfaz, que no se entiende y confunde, permitiendo realizar correcciones.
- **Con el producto final:** Las pruebas en esta etapa apuntarán a estudios pragmáticos aplicados al control de la interfaz para comandar el robot. Se recabarán datos para conocer su usabilidad, por ejemplo: conocer las opciones fáciles de comprender, en la que se desempeñan con más soltura, la que más agrada. Estas pruebas pueden conllevar a un rediseño completo del producto.

Se espera obtener resultados sobre la usabilidad de la interfaz tanto para usuarios en general como también para usuarios con movilidad reducida, sobre todo en sus miembros superiores. La Universidad Blas Pascal cuenta con la Comisión de Inclusión de Personas con Discapacidad que tiene por objeto impulsar actividades para el progresivo desarrollo de personas con dificultades que impiden su participación plena en la comunidad. Es importante destacar que se cuenta con el apoyo de esta comisión para realizar las pruebas con personas con movilidad reducida y con dificultades para su comunicación oral.

Formación de Recursos Humanos

El grupo se compone por cuatro investigadores con experiencias y roles distintos: el Ing. Carlos Bartó como asesor en metodología de la investigación, el Ing. César Osimani como director del proyecto y el Ing. Martín Salamero y el Lic. Marcos Lopez como investigadores en formación. Se cuenta con la colaboración de dos alumnos pasantes de Ing. Informática. Además, se asocia a la tesis "Reconocimiento de posturas de las manos para la Interacción Natural Humano Computadora en Ambientes Inteligentes con cámaras RGB" de César Osimani por el título Doctor en Ciencias de la Ingeniería de la FCEfyN de la UNC.

Financiamiento

La investigación en esta comunicación cuenta con el financiamiento para Grupos de Reciente Formación y Tutoría del Gobierno de Córdoba, y dos becas de investigación y una dedicación exclusiva a investigación por parte de la Universidad Blas Pascal.

Referencias

- [1]. "OpenCV - Open source Computer Vision", Online: <http://docs.opencv.org/trunk/index.html>, [Consulta: marzo de 2017].
- [2]. G Bradski, A. Kaehler, "Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library", O'Reilly Media, 2016.
- [3]. "Qt Library - Documentación", Disponible: <http://doc.qt.io>, [Consulta: marzo de 2017].
- [4]. P. Ithaya Rani, K. Muneeswaran, "Robust Real Time Face detection automatically from video sequence based on Haar features", "Communication and Network Technologies (ICCNT 2014)", Dec. 2014.
- [5]. "Menú controlado con orientación del rostro", <http://github.com/cosimani/menupencv>, 2015.
- [6]. "Video del menú controlado con el rostro", <https://www.youtube.com/watch?v=BUmQNs7GxPY>, 2015.
- [7]. C. Osimani y E. Kohmann, "Simple método para la Interacción Natural Humano Computadora con movimientos del rostro", CoNaIISI 2015, Bs. As. Arg., Nov. 2015.
- [8]. Shan An, Xin Ma, Rui Song, Yibin Li, "Face detection and recognition with SURF for human-robot interaction", 10th Asian Control Conference (ASCC 2015), June 2015.
- [9]. S.-S. Lee, S.-J. Jang, J. Kim, B. Choi, "A hardware architecture of face detection for human-robot interaction and its implementation", (ICCE-Asia), Oct. 2016.
- [10]. S. Fernandes, J. Bala, "Low Power Affordable and Efficient Face Detection in the Presence of Various Noises and Blurring Effects on a Single-Board Computer", 49th Annual Convention of the Computer Society of India (CSI), Volume 1. vol. 337, pp. 119-127, 2015.
- [11]. Raspberry Pi Foundation, About us, [online] Disponible:<https://www.raspberrypi.org/about>, [Consulta: marzo de 2017].
- [12]. R. A. Burnoa, B. Wub, R. Dohertya, H. Coletta, R. Elnaggar, "Applying Fitts' law to gesture based computer interactions", 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015), 2015.
- [13]. Placa Intel Galileo, [online] Disponible:<https://software.intel.com/es-es/iot/hardware/galileo>, [Consulta: marzo de 2017].
- [14]. S. Ahsan, "Peripheral control with a Raspberry Pi - An ultrasonic sensor", [online]. Disponible:<https://wiki.aalto.fi/download/attachments/84751400/burglaralarm.pdf>, [Consulta: marzo de 2017].
- [15]. M. A. Eid, N. Giakoumidis, A. El Saddik, "A Novel Eye-Gaze-Controlled Wheelchair System for Navigating Unknown Environments: Case Study With a Person With ALS", IEEE Access, Jan. 2016.
- [16]. F. C. Luque, E. B. López, "Símbolos pictográficos de ARASAAC: ¿son adecuados?". J. Navarro, M. T. Fernández, F. J. Soto, F. Tortosa (Coords.), (2012), Respuestas flexibles en contextos educativos diversos. ISBN: 978-84-616-0718-1.