



UNIVERSIDAD POPULAR AUTÓNOMA DEL ESTADO DE PUEBLA

CENTRO INTERDISCIPLINARIO DE POSGRADOS

INVESTIGACIÓN Y CONSULTORÍA

DEPARTAMENTO DE INGENIERÍA

DOCTORADO EN INGENIERÍA MECATRÓNICA

INTERACCIÓN HUMANO-ROBOT CON ADEMANES USANDO KINECT 2.0

TESIS QUE PARA OBTENER EL GRADO DE

DOCTOR EN INGENIERÍA MECATRÓNICA

PRESENTA

HAROLD ANDRES VASQUEZ CHAVARRIA

PUEBLA, MÉXICO.

30, OCTUBRE, 2018



UPAEP – Secretaría General

Dirección General de Apoyos Académicos

Dirección del Centro de Recursos para el Aprendizaje y la Investigación.

Biblioteca Central - **Karol Wojtyła**

Tesis Digitales Restricciones de uso:

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de textos, imágenes, gráficas, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente de donde la obtuvo mencionando el autor o autores involucrados en el documento.

Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD POPULAR AUTÓNOMA DEL ESTADO DE PUEBLA

CENTRO INTERDISCIPLINARIO DE POSGRADOS

INVESTIGACIÓN Y CONSULTORÍA

DEPARTAMENTO DE INGENIERÍA

DOCTORADO EN INGENIERÍA MECATRÓNICA

SE APRUEBA LA TESIS:

INTERACCIÓN HUMANO-ROBOT CON ADEMANES USANDO KINECT 2.0

HAROLD ANDRES VASQUEZ CHAVARRIA

COMITÉ DE TESIS DOCTORAL

DR. HÉCTOR SIMÓN VARGAS MARTÍNEZ

COORDINADOR

DR. VITTORIO ZANELLA PALACIOS

SECRETARÍO

DRA. GEORGINA GUADALUPE ROSAS GUEVARA

VOCAL

Puebla, México.

09, 11, 2018

Agradecimiento

A Dios, mi esposa, mi madre y padre, mis hijas, mis hermanas y amigos; que aportaron paciencia y ánimos para alcanzar este fin

Dedicatorias

A todos los compatriotas venezolanos que, al igual que mi persona, dejaron atrás una patria llena de familia, vivencias y recuerdos que nos genera constantemente nostalgia y esperanzas. Un coctel de sentimientos que nos impulsa a seguir adelante y luchar por alcanzar más logros como este.

Índice General

Agradecimiento	3
Dedicatorias	4
Índice General.....	5
Índice de Tablas.....	7
Índice de Gráficas y Figuras.....	8
Abstract.....	10
Resumen en español	10
Introducción	11
Planteamiento de la Investigación	13
Antecedentes	16
Marco Teórico	18
Robots	18
Robótica de servicio	18
Interacción humano-robot.....	23
Kinect	24
Visual Gesture Builder	24
Metodología	26
Plan de pruebas	27
Resultados	29
Robot de aplicación	29
Construcción del DataSet.....	32
DataSet	33

Consideraciones de captura de datos.....	36
Captura de datos	39
Estadísticas	40
Aplicación	41
Conclusiones	44
Referencias bibliográficas	45

Índice de Tablas

Tabla 1 Numero totales de gestos en el DataSet	40
Tabla 2 Algunos datos del DataSet. Distancias (Mínima, Media y Máxima), ocasiones o sesiones, de frente o de espalda al Kinect, y los totales por ademán.	41

Índice de Gráficas y Figuras

Figura 1 Evolución en la aparición de nuevas Sociedades Técnicas de la IEEE relacionadas con la Robótica de servicio (Moradi et al., 2013)	19
Figura 2 Evolución de la investigación en robótica (Liu et al., 2009)	19
Figura 3 Mercado actual y estimación de los robots de servicio según sector de uso (International Federation of Robotics, 2018a)	20
Figura 4 Robot de asistencia para la cirugía (International Federation of Robotics, 2018a).....	21
Figura 5. Robot de servicio en el campo, para la agricultura (International Federation of Robotics, 2018a)	21
Figura 6 Distribución del volumen de ventas para robots de servicio de asistencia en el hogar y de entretenimiento (International Federation of Robotics, 2018a) ...	22
Figura 7 Distribución de compañías constructoras de robots de servicio por continente (International Federation of Robotics, 2018a)	23
Figura 8 Kinect versión 2.0 de la compañía Microsoft	24
Figura 9 Modelo propuesto para el reconocimiento de ademanes	27
Figura 10 DONAXI, Robot de Servicio de la UPAEP (Foto propia).....	30
Figura 11 Robots de servicio que compitieron en el TMR2015 (Foto propia)	31
Figura 12 Pasos a seguidos para la construcción del DataSet de gestos entrenados (Fuente propia)	32

Figura 13 se muestra algunas secuencias (por columna) de los ademanes considerados. Desde la primera a la última columna los ademanes son: attention, come, left, right, pointing, stop y turn.....	35
Figura 14 Kinect Estudio de Microsoft (Captura de pantalla)	35
Figura 15 Visual Gesture Builder VGB (Captura de pantalla)	36
Figura 16 Kinect y el trípode usado para la captura de los datos (Imagen propia).	37
Figura 17 Tres ademanes entrenados para crear la base del sistema prototipo (Foto propia).....	42
Figura 18 Capturas de pantalla de la herramienta Microsoft VGB Live Preview donde la persona ejecuta cada uno de los tres gestos.....	42

Abstract

In this research is presented a proposal for interaction between the robot and the person, using gestures, in order to make friendlier human-robot relationship. This functionality will be integrated into a service robot called Donaxi, from UPAEP's Robotics Lab. Some data from a gestures dataset being built for service robots is shown. It showed an application where the service robot Donaxi can recognize some of the gestures performed to it.

Resumen en español

En esta investigación se presenta una propuesta de interacción entre el robot y la persona, mediante gestos, para establecer una relación más amigable entre humanos y robots. Esta funcionalidad se integrará en un robot de servicio llamado Donaxi, del Laboratorio de Robótica de la UPAEP. Se muestran algunos datos de un conjunto de datos de gestos que se está construyendo para robots de servicio. Se muestra una aplicación donde el robot de servicio Donaxi puede reconocer algunos de los gestos que se le realizan.

Introducción

La robótica llegó para quedarse. Hoy en día no se trata de ciencia ficción y es una disciplina multidisciplinaria que incluso está generando carreras universitarias propias. En México es estudiada en distintos Institutos de Educación Superior (IES) y el campo laboral está requiriendo cada vez más especialistas en esta rama, en especial en el campo manufacturero. Según (Transparency Market Research, 2018) se calcula que en el año 2017 el mercado de la robótica manejó unos US\$ 1,108.4 Mn y se estima una tasa de crecimiento anual compuesto de 28.1% al año 2026.

En términos generales, la robótica se podría clasificar en dos grandes grupos, acorde a Federación Internacional de la Robótica (IFR, por sus siglas en inglés): robótica industrial y robótica de servicio (International Federation of Robotics, n.d.).

Acorde con la IFR, el valor de las ventas en robótica de servicio para uso profesional aumentó en un 39% este año, llegando a un valor de 6.6 mil millones de dólares. Esto indica un mercado potencial en crecimiento que amerita fuerza laboral e intelectual, y abre una oportunidad de negocios potenciales (International Federation of Robotics, 2018b).

Consciente de esta realidad, en la Universidad Popular Autónoma del Estado de Puebla (UPAEP), en su laboratorio de robótica, se viene construyendo y programando un robot de servicio para asistencia en el hogar, denominado Donaxi.

En la robótica de servicio la interacción entre el robot y el ser humano es más constante y en la mayoría de los casos, lo más parecido a lo que ocurre entre las personas. Por ello existe la necesidad que la interacción entre el usuario y el robot sea lo más natural posible. Para este fin, se ha estudiado distintos medios de comunicación que permitan a una persona, sin muchos conocimientos de tecnologías, poder comandar el robot que, por ejemplo, tendrá en su hogar. Puede ser mediante voz, un control remoto, celular o Tablet, etc. En el caso del control remoto, existen muchas funcionalidades posibles del robot y con alto nivel de complejidad, por lo que sería poco funcional contar con un solo equipo que cuente con todos estos comandos.

En el caso de la voz, aún existen muchos retos que solventar que permitan una comunicación clara y fluida entre un humano y el robot. Situaciones como distancia y ruido del ambiente, pueden entorpecer la interpretación correcta de los comandos por parte del robot.

Es por ello que, la interacción por medio de visión artificial, en este caso, mediante ademanes, puede ser una alternativa más adecuada a las tecnologías disponibles actualmente. Estos ademanes pueden fungir como algunos de los comandos que se quieren instruir al robot.

Por lo antes descrito, se propone el desarrollo de una aplicación que permita al robot de servicio de la UPAEP: Donaxi, poder interactuar con una persona por medio de ademanes.

En el presente documento se presenta los resultados de la aplicación de un reconocedor de ademanes para instruir a un robot de servicio. El documento primero presenta el planteamiento de este proyecto, los antecedentes, un marco teórico, la metodología de la investigación, los resultados obtenidos y las conclusiones.

Planteamiento de la Investigación

Muchos países, en especial los países europeos, están preocupados por el cuidado de los ancianos. Esto afecta los aspectos sociales y económicos de cualquier país. Dado el envejecimiento de la población, en el futuro habrá una falta de cuidadores para las personas mayores y los robots de servicio ofrecen una alternativa para ayudar a las personas mayores. Debido a esto, hay muchos programas de apoyo para ayudar a desarrollar este tipo de robots, llamados robots de servicios (Aracil, Balaguer, & Armada, 2008).

Debido a que estos robots van a interactuar con una persona, es necesario un modo cómodo de comunicación para que la persona sienta que el robot es capaz de entender muy bien las instrucciones dadas. La forma más natural de lograr esto es con voz, ya que es la forma de comunicación más común entre nosotros. Sin embargo, usar sólo la voz para comunicarse no es suficiente, debido a algunos problemas y también porque no todas las personas pueden usar la voz. Por lo tanto, es necesario utilizar otra forma de comunicación.

Está claro que los gestos son una alternativa muy útil, aun así, no es muy común entre las personas. Los gestos se han utilizado en muchos proyectos con robots de servicio y han demostrado ser muy eficientes en ambientes ruidosos, donde la voz no es suficiente.

Los gestos han demostrado ser muy útiles para interactuar con un robot (Droeschel, Stückler, Holz, & Behnke, 2011; Goodrich & Schultz, 2007; Liu, Yang, Wang, & Meng, 2009). Se han utilizado muchos tipos de dispositivos para lograr esto, pero los más comunes son las cámaras de video y el Kinect. Entre estos, el Kinect es el más utilizado en los robots de servicios, debido a que los programadores no tienen que lidiar con la segmentación ni con el seguimiento de las personas. Estas funcionalidades son proporcionadas por el Kinect SDK de Microsoft, utilizando una combinación de hardware y software. Con esto, el problema se redujo para identificar cuándo se está realizando el gesto y qué gesto realizó el usuario. El problema "cuando" se llama segmentación y el problema "qué" se llama

reconocimiento. Cada problema se puede resolver por separado, pero el objetivo es que ambos funcionen simultáneamente, por lo que se logra una interacción más natural (Vasquez Chavarria, Escalante, & Sucar, 2013).

Este modo de interacción está inspirado en los videos juegos que usan distintos dispositivos para fines parecidos. Uno de estos dispositivos es el Kinect® de la consola Xbox de Microsoft. Este dispositivo permite la interacción con el video juego con solo movimientos de los brazos y en general del cuerpo.

Por otro lado, la voz es la forma más natural de interactuar con las máquinas. El problema más común con esto es que la máquina no puede entender a la persona. Esto porque cada persona tiene diferente voz y pronunciación. Otro problema que afecta a esto son los ambientes ruidosos. Esto se puede resolver con: solo software, solo hardware o ambos. En esos casos, es necesario complementar la voz con otras vías de comunicación. Obviamente, los gestos son la mejor opción para hacer esto. Por lo tanto, podemos obtener una interacción multimodal con el robot de servicio mediante voz y gestos.

Las diferentes situaciones en las que se pueden utilizar la voz y los gestos juntos son:

1. El gesto es un refuerzo de voz. Es decir, cuando el robot no es capaz de entender el comando de voz, la persona realiza un gesto con el mismo significado que el comando de voz. Por ejemplo, si el usuario llama con voz al robot para que le preste atención pero, debido a que están en un Centro Comercial, el robot no puede escuchar a su dueño, entonces el usuario puede mover su mano para llamar la atención del robot.
2. El gesto es un complemento de voz. Es decir, en el entorno doméstico, el usuario desea que el robot le dé un nuevo medicamento sin que él lo sepa y por lo tanto, el robot no sabe dónde está. Luego, con voz, el usuario solicita el medicamento al robot y simultáneamente, con la mano, muestra el lugar donde se encuentra el medicamento.

Dado que la complejidad de este problema propuesto es alta, en el presente trabajo se enfocará en el tema de ademanes. Por lo tanto, se propone un sistema de

software que permita reconocer ademanes a través del Kinect versión 2, para comandar el robot de servicio Donaxi, de la Universidad Popular Autónoma del Estado de Puebla.

Antecedentes

Los robots de servicio han aparecido como una necesidad para satisfacer las labores cotidianas del hogar y vislumbran un futuro promisorio (Aracil et al., 2008). Debido a la necesidad de que puedan interactuar con personas, en ambientes no controlados y desconocidos al principio; existen un gran número de grupos de investigación que se dedican a solucionar el gran número de dificultades que existen y están por surgir para lograr este fin. En lo relacionado con la interacción humano robot, algunos trabajos han mostrado algunas propuestas para lograr esto.

Varios trabajos han tratado lo relacionado con el reconocimiento de ademanes (Mitra & Acharya, 2007; Watson, 1993; Wu & Huang, 1999). En la mayoría de los casos se plantea problemas de tratamiento de imágenes, en lo que se refiere a segmentar el área de interés (persona y manos) y seguimiento, es decir, poder determinar dónde se encuentra la persona o la mano en las siguientes imágenes de la secuencia de video. Algunas propuestas se enfocan solo en el reconocimiento, por lo que sus fuentes de datos son videos editados, donde sólo se ve a la persona realizando el ademan en estudio. Otros se enfocan en automatizar esta segmentación del video para lograr una interacción en tiempo real. En (Vasquez Chavarria et al., 2013) se busca combinar ambos aspectos: segmentación y reconocimiento de manera conjunta. Para ello, se usa Modelos Ocultos de Markov (HMM, por sus siglas en inglés), junto con un novedoso modelo de múltiples ventanas de tamaños dinámicos.

Aunque la mayoría de los trabajos no usan el Kinect como dispositivo de captura, por lo que deben lidiar con el procesamiento de imágenes, solo recientemente han empezado a aparecer trabajos que aprovechan este sensor (Bernstein, Lotocky, & Gallagher, 2012; Gowing et al., 2014; Pavlakos, Theodorakis, Pitsikalis, Katsamanis, & Maragos, 2014; Zaman, 2011).

Finalmente, en lo que se refiere a interacción humano robot multimodal, aun no existen muchos trabajos que traten con ademanes y voz como fuentes de señal, sino que se basan en otras (Goodrich & Schultz, 2007; Jaimes & Sebe, 2005;

Marcel, 2002; Pavlakos et al., 2014). En (Nieuwenhuisen, Stuckler, & Behnke, 2010) se menciona una interacción multimodal, utilizando reconocimiento de voz, gestos de cara por parte de uno de los robots y mímica. Por lo tanto, no se contempla el reconocimiento de ademanes. En (Pavlakos et al., 2014) se propone un modelo de interacción multimodal usando modelos ocultos de Markov (HMM por sus siglas en inglés), usando gestos que acompañan a las palabras pronunciadas en italiano. Además, en este trabajo aún se utilizó el Kinect versión 1 para sus fines.

Algunos mecanismos de interacción multimodal combinan tecnologías más avanzadas, como electroencefalograma (EEG), electromiografía (EMG) y voz (Yan Bian, Xuanfang Wang, Li Zhao, & Shigang Cui, 2016).

Marco Teórico

En la presente sección se presentará algunos temas concernientes a la presente investigación. La información presentada esta basada en fuentes fidedignas y confiables a modo de garantizar exactitud en la información aportada.

Robots

Bien declaró Joseph Engelberger, el padre de la robótica: “No puedo definir un robot, pero lo sé cuando lo veo” (Zielinska, 2016). Sin embargo, se hace necesario hacer un intento por definir y delimitar el alcance de este campo de investigación y desarrollo.

Según la ISO 8373:2012, un robot es un “Mecanismo accionado programable en dos o más ejes con un grado de autonomía, moviéndose dentro de su entorno, para realizar las tareas previstas” (ISO/TC 184, 2012). De igual forma, en este apartado se señala “La clasificación del robot en robot industrial o robot de servicio se realiza de acuerdo con su aplicación prevista”.

Robótica de servicio

Acorde con la ISO 8373:2012 un robot de servicio se define como “Robot que realiza tareas útiles para personas o equipos, excluyendo aplicaciones de automatización industrial”.

Es tal la importancia de esta rama de la robótica, que en 1995 se crea el Comité Técnico de Robótica de Servicio en la Sociedad de Robótica y Automatización de la IEEE. Incluso se han creado más comités técnicos relacionados con esta rama desde dicho año (Moradi et al., 2013). En la Figura 1 se muestra la evolución de estas.

Joseph Engelberger predijo que la robótica de servicio se convertiría un día en la aplicación de robots mas ampliamente usada, superando en numero a las aplicaciones industriales (Zielinska, 2016).

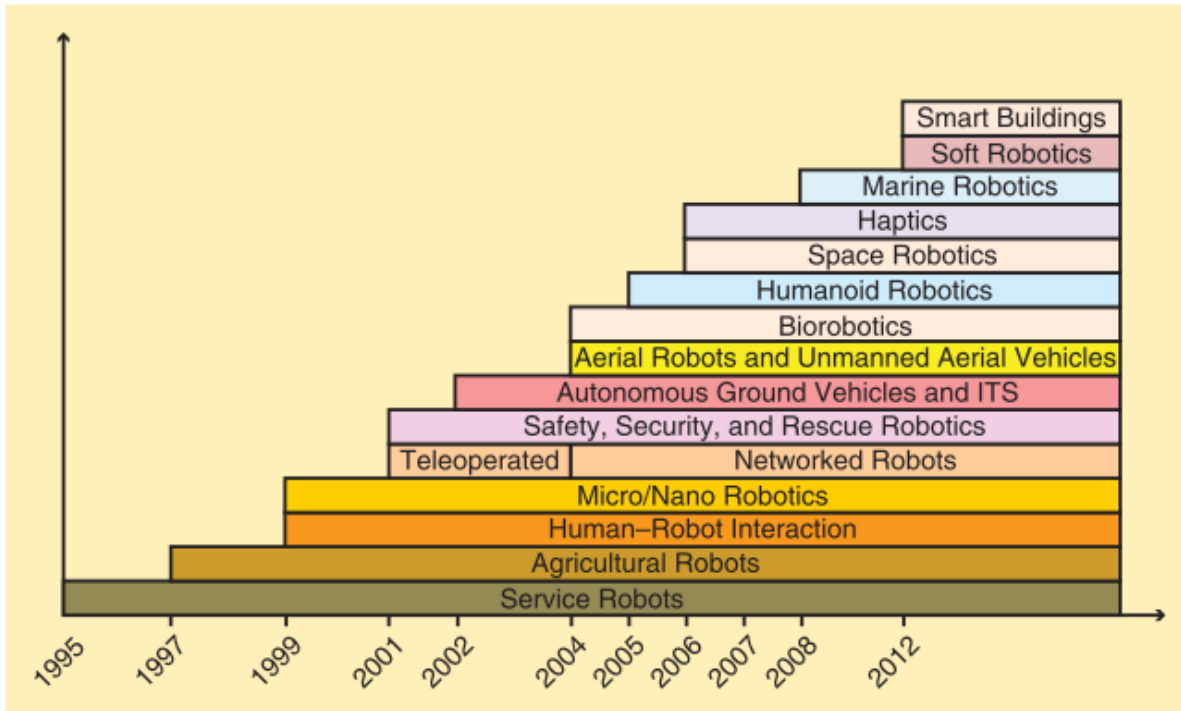


Figura 1 Evolución en la aparición de nuevas Sociedades Técnicas de la IEEE relacionadas con la Robótica de servicio (Moradi et al., 2013)

Este campo de la robótica ha ido evolucionando desde hace mucho tiempo. En la Figura 2 se muestra la evolución de este tipo de robots durante los últimos 50 años.

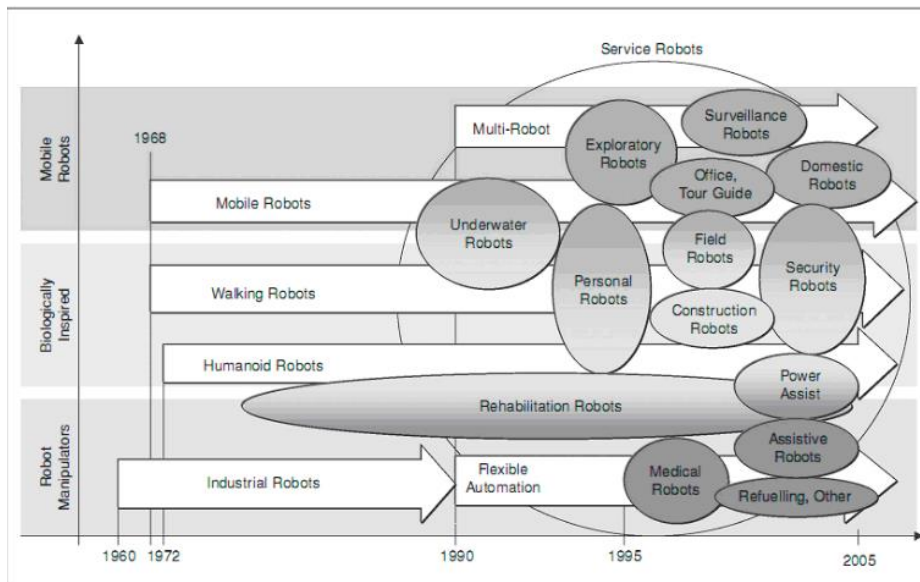


Figura 2 Evolución de la investigación en robótica (Liu et al., 2009)

Según la IFR, este sector posee un mercado potencial alto, representado este año un mercado de 6.6 billones de dólares, es decir, un incremento de 39% respecto al año anterior (a octubre de 2018). Los sistemas logísticos muestran la mayor demanda, representando el 63% del total de unidades y el 36% de las ventas totales (en valor) de robots de servicio profesional. Además, se pronostica para el periodo 2019-2021 un volumen acumulado de alrededor de 46 billones. De esto, los robots médicos, logísticos y de campo, serán los que mayor aporten a esta cifra (International Federation of Robotics, 2018b)

Estas cifras y las que siguen pueden apreciarse en la Figura 3.

Los robots en logística han crecido en unidades un 162% a nivel mundial en el periodo 2016 a 2017, dado que pasó de 26,300 unidades a 69,000. De esto, 62,200 vehículos guiados automatizados se encuentran en ambientes no manufactureros, tales como Amazon, eBay, etc.

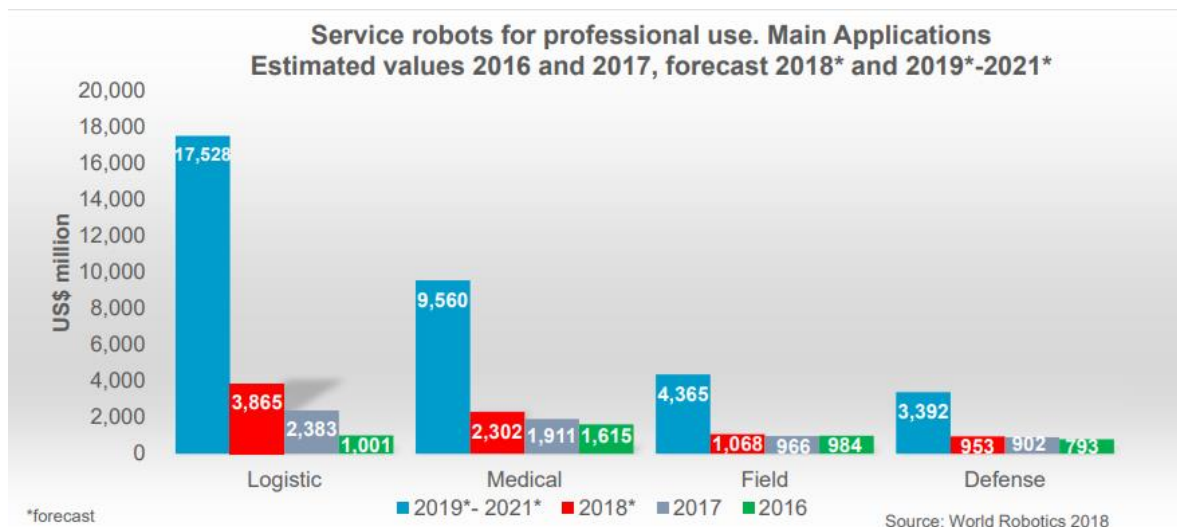


Figura 3 Mercado actual y estimación de los robots de servicio según sector de uso (International Federation of Robotics, 2018a)

En el campo médico, el valor total de robots médicos se incrementó a 1.9 billones, lo que representa un 29% del total de robots de servicio en 2017. Las dos aplicaciones que más destacan en este campo son los robots para asistencia en cirugía y los usados en rehabilitación y terapia. En la Figura 4 se aprecia un robot de este tipo.



Figura 4 Robot de asistencia para la cirugía (International Federation of Robotics, 2018a)

Los referentes a robots de campo, que son acaparados por los usados en el ordeño, representan un 15% de la robótica de servicio. En 2017 se vendieron un total de 6,375 unidades. De estos, para el ordeño se pasó de 5,300 en 2016 a 5,400 en 2017 y para la agricultura de 190 a 520 en el mismo periodo. En la Figura 5 se puede apreciar un ejemplo de estos robots.



Figura 5. Robot de servicio en el campo, para la agricultura (International Federation of Robotics, 2018a)

De forma análoga, en el sector de robots de servicio personal, tal como el caso de Donaxi y otros robots similares en México, su valor de ventas se incrementó en 27% llegando a 2.1 billones de dólares. Se estima que se vendieron cerca de 6.1 millones de robots para tareas domésticas en 2017, que incluyen robots aspiradores, cortadores de césped, limpiadores de ventanas y otros tipos; un impresionante aumento del 31% en comparación con 2016. En la Figura 6 se aprecia un mayor detalle en cómo se distribuyen las ventas de este tipo de robots.

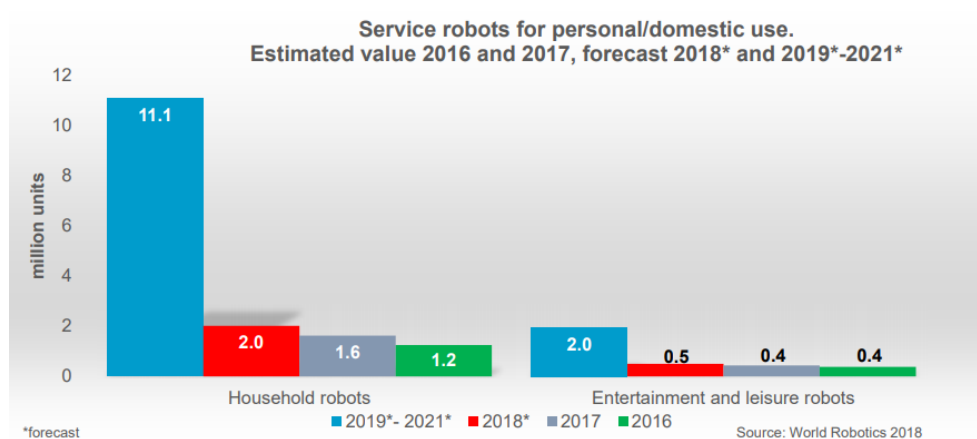


Figura 6 Distribución del volumen de ventas para robots de servicio de asistencia en el hogar y de entretenimiento (International Federation of Robotics, 2018a)

En términos de las compañías dedicadas de este segmento de la robótica, se puede apreciar en la Figura 7 que el continente europeo domina en este sector, seguido por USA y en tercer lugar Asia. De las 700 compañías registradas en este segmento, 300 corresponden a Europa, 250 a Norteamérica y 130 de Asia. De estos, un 30% son considerados “empresa emergente”, lo que indica el mercado joven que esta emergiendo en este sector.

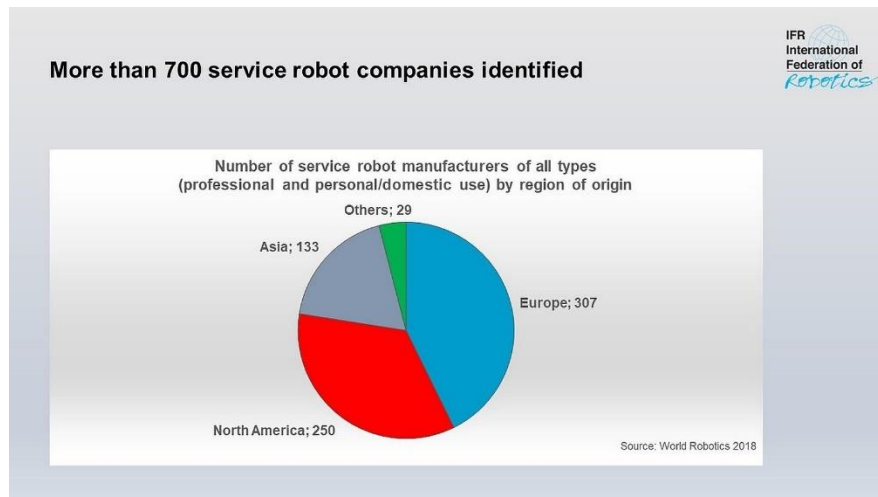


Figura 7 Distribución de compañías constructoras de robots de servicio por continente
(International Federation of Robotics, 2018a)

Interacción humano-robot

Acorde a la ISO 8373:2012, la interacción Humano-Robot (HRI, por sus siglas en ingles), se define como “Información e intercambio de acciones entre humanos y robots para realizar una tarea mediante una interfaz de usuario” (ISO/TC 184, 2012).

En la ISO/IEC 30113-11:2017 se determina la descripción y terminología usada para interacción basados en ademanes con cualquier dispositivo (ISO, 2017)

En la robótica de servicio, uno de los aspectos principales a considerar en el desarrollo de este tipo de robots, es el modo de interactuar con el humano. Si esto no se cuida, el proyecto será muy propenso a fracasar. Es de suma importancia considerar que la mayoría de las personas a las que está destinado este tipo de robot, no tienen conocimientos tecnológicos o científicos para entrenamientos previos del robot o uso de mecanismos muy sofisticados. La tareas domesticas que realizan los robot de servicio personales requieren tres destrezas principales: navegación segura en ambientes internos, manipulación flexible de objetos y comunicación intuitiva con sus dueños (Nieuwenhuisen et al., 2010).

Es por ello que modalidades de interacción muy sofisticados como interfaces basados en la lectura de ondas cerebrales (BCI-Brain Computer Interfaces) (Yan Bian et al., 2016), quedan delegados a laboratorios o personal altamente entrenado.

Kinect

El Kinect, desde su versión 1.0, ha ganado mucha popularidad en los campos científicos y tecnológicos para ser usado en otros fines distintos al original. Es por ello que, en su nueva versión, el Kinect viene acompañado de una serie de paquetes informáticos que permiten el desarrollo de aplicaciones distintas al de video juegos, para usar el dispositivo para otros fines.

Aunque el Kinect 1.0 ya se venía usando en el campo de la robótica y en especial en la robótica de servicio, sus niveles de confianza no habían sido los suficientemente satisfactorios. Es por ello que, con las mejoras en hardware que ofrece la versión 2.0, se espera un mejor aprovechamiento de este dispositivo para distintos fines en la robótica. En la Figura 8 se muestra este dispositivo.



Figura 8 Kinect versión 2.0 de la compañía Microsoft

Visual Gesture Builder

Junto con el Kinect versión 2, se dispone de un conjunto de herramientas de software que permite extender el uso de este dispositivo as allá de los videos juegos en la consola Xbox. Estas herramientas se denominan SDK, y permite crear de modo acelerado nuevas aplicaciones que usen el reconocimiento de ademanes o voz.

El equipo de Microsoft (los propietarios de Kinect) desarrollaron un IDE donde es posible hacer nuevos sistemas con gestos propios. Este IDE se llama Visual Gesture Builder (VGB). Con esto, después de disponer de un buen número de videos de ejemplos, podemos obtener una base de datos con los nuevos gestos entrenados. Con esta base de datos de gestos (DBG), es posible desarrollar un sistema de software donde se utilicen los mismos

En algunas pruebas realizadas, esta técnica demostró ser muy buena para resolver los dos problemas descritos anteriormente: segmentación y reconocimiento. Sin embargo, todo el trabajo anterior fue muy tedioso y largo. Se debe capturar muchos videos de diferentes personas para lograr una buena capacitación de DBG. Después, tenemos que etiquetar cada video con los gestos donde se realizan.

A pesar de lo anterior, es un trabajo que sólo se realiza una vez y que con la DBG ya se puede desarrollar la aplicación sin problemas. Todo cambia si se decide incorporar un nuevo gesto, lo que generaría repetir todo el proceso nuevamente.

Metodología

En lo que se refiere al reconocimiento de ademanes, la metodología de trabajo que se llevó a cabo hasta el momento es el siguiente:

1. Entender el Kinect versión 2: a pesar de disponer de cierta experiencia con el Kinect versión 1, esta versión reciente de este dispositivo presenta nuevas innovaciones que deben ser bien entendidas para su máximo provecho. Uno de las formas más expeditas para lograr esto, es el uso de los ejemplos ya disponibles en el SDK. El ejemplo que más se ajusta a lo que se trabajó en este documento es el de DiscreteGestureBasics. Adicionalmente, se hace necesario el estudio de la documentación y tutoriales disponibles en la web de Microsoft para ahondar en este tema.
2. Determinar calidad de reconocimiento y seguimiento del Kinect: esto se enfoca principalmente en lo referente a personas y manos. Aunque publicitariamente Microsoft asegure una extraordinaria calidad de reconocimiento y seguimiento de personas y manos; es necesario comprobar esto por nuestra parte. Para esto, se plantea diseñar un plan de pruebas que permita determinar si el Kinect puede confundirse y reconocer objetos u otros seres vivos como personas (falsos positivos). Lo mismo aplicaría para las manos, que son la parte del cuerpo que nos interesaría para el reconocimiento de ademanes. Adicionalmente, se plantea la interrogante de la calidad de esta prestación en el caso de niños con capacidad suficiente para instruir un robot de servicio. Esto porque las personas de edad avanzada suelen estar acompañados de nietos.
3. Determinar calidad de reconocimiento de ademanes: se plantea diseñar un experimento para comprobar la calidad del reconocimiento de ademanes que se puede obtener con las tecnologías ofrecidas por el SDK del Kinect. La idea es obtener medidas que permitan asegurar fehacientemente si el reconocimiento de ademanes que se tiene es bueno o no. Esto considerando distintos escenarios no controlados, distintos tipos de usuarios y distintas condiciones de iluminación.

De acuerdo a los resultados que se obtengan en las fases antes descritas, se pueden plantear distintos caminos posibles a seguir para solventar alguna deficiencia que surja

En lo que respecta a este trabajo, para el reconocimiento de ademanes se tiene planteado el modelo mostrado en la Figura 9

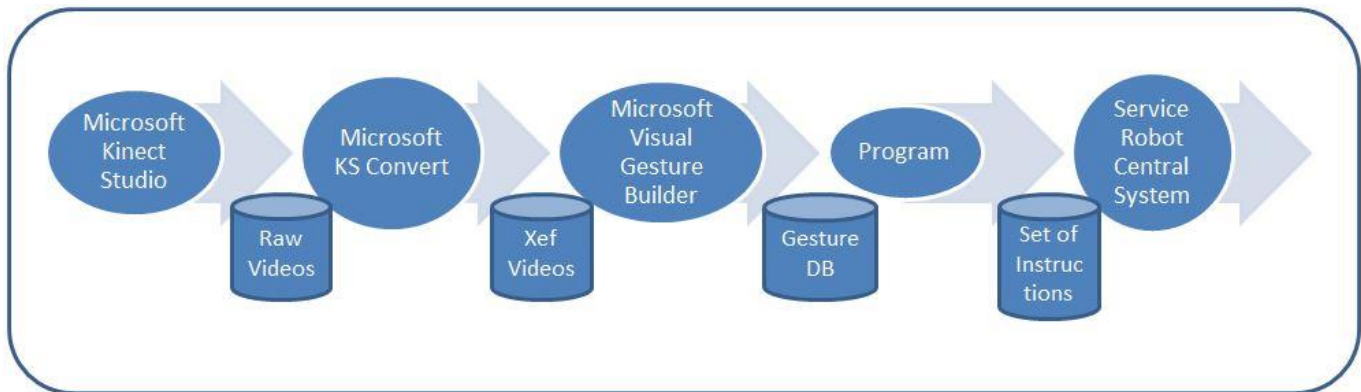


Figura 9 Modelo propuesto para el reconocimiento de ademanes

Como se aprecia, se está sacando provecho de varias herramientas ofrecidas en el SDK que acompaña al Kinect versión 2. Con el Kinect Estudio (KS) se capturan videos en formato Raw, que luego son convertidos con el KS convertidor a formato Xef. Estos últimos son usados en el Constructor de Gestos Visual (VGB, por sus siglas en inglés) para crear una base de datos de gestos, que se usa en una aplicación desarrollada.

Plan de pruebas

Se propone la realización de las siguientes pruebas:

1. Se tienen 14 personas, de cada persona se tienen dos secuencias con m_i repeticiones para cada ademan i , en cada secuencia. En teoría se deberían tener 28 secuencias, pero en la práctica hay 6 secuencias que faltaron por grabar, por lo que en realidad se tienen 22 secuencias. Se pueden tomar 21 secuencias para entrenamiento y la 22 se usa para prueba, al estilo de one

leave off cross validation. En este caso el número de repeticiones del ademan i para entrenamiento y pruebas no es rigurosamente controlable.

2. Se podría repetir el experimento anterior, pero esta vez solo con el grupo de secuencias de distancia máxima, luego lo mismo solo con el grupo de secuencias de distancia mínima y finalmente con los tres grupos juntos.
3. Para la variable de iluminación, solo se tiene dos grandes grupos, por lo que no se espera tener resultados visibles si se separan aplicando la técnica descrita en 1, dado que el número de secuencias para cada grupo sería muy reducida.

Resultados

En esta sección se detalla los resultados obtenidos al desarrollar y probar la aplicación piloto, para la interacción de varias personas con un robot de servicio.

Robot de aplicación

Esta investigación se lleva a cabo en el laboratorio de robótica UPAEP, donde hay un robot de servicio llamado Donaxi. Este robot está completamente construido desde cero por estudiantes de diferentes carreras universitarias como electrónica, mecatrónica, biónica y computación. Donaxi tendrá:

1. Sistema de navegación omnidireccional, con cuatro ruedas, cada una con un motor DC con encoder. Esto permite a Donaxi moverse en casi cualquier dirección.
2. Sistema láser para construir el mapa de navegación. Se compone de dos láser, delantero y trasero.
3. Sistema de movimiento vertical, basado en un riel y un motor, que mueve una plataforma hacia arriba o hacia abajo.
4. Un brazo con cinco grados de libertad y una pinza paralela en el extremo.
5. Un Kinect versión 1 que se mueve con el sistema de movimiento vertical y se usa para el reconocimiento de objetos.
6. Un Kinect versión 2 para el reconocimiento de personas, ya sea todo el cuerpo, la cara o la voz. Esto se fija en la parte superior del robot. Este se utilizará para reconocimiento multimodal con gestos y voz.
7. Dos laptops. Uno con Ubuntu y ROS, usado para algunas de las funciones del robot, y el segundo con Windows, usado para reconocer personas, rostros, gestos y voz. Las computadoras portátiles se comunican entre sí a través de mensajes TCP / IP.

Algunos de estos dispositivos ya están disponibles en el robot, pero otros están en proceso de adaptación. En la Figura 10 se muestra la versión preliminar de Donaxi.



Figura 10 DONAXI, Robot de Servicio de la UPAEP (Foto propia)

Debido a que en abril del 2015 se realizó el Torneo de Robótica Mexicana (TMR2015)¹ y el equipo de Donaxi participó en él, se hizo necesario tener un trabajo de gestos y reconocimiento de voz por separado. Por esta razón, se tuvo algunos avances en estas dos características, que se probaron en un ambiente casi realista en este torneo. En la Figura 11 se muestran los servicios que los robots que compitieron en TMR2015.

¹ <http://www.tmr2015.mx/>



Figura 11 Robots de servicio que compitieron en el TMR2015 (Foto propia)

Así como existe el TMR a nivel nacional con esta sesión especial para robótica de servicio, existen otras competencias a nivel internacional, como Rockin² y el RoboCup³. En todas estas competencias, se construye un ambiente de hogar lo más real posible. Esta competencia se conoce como @Home, donde un robot tiene que comunicarse con el usuario para recibir instrucciones de él; entonces el robot debe realizar ciertas actividades de acuerdo con las instrucciones recibidas. Esta competencia se lleva a cabo comúnmente en escenarios con muchos asistentes, generando tanto ruido que complica el HRI a través de la voz. De este modo se busca garantizar que los robots participantes demuestren si son capaces de desenvolverse en un ambiente como este.

² <http://rockinrobotchallenge.eu/>

³ www.robocup2015.org

Construcción del DataSet

Para el reconocimiento de gestos, se utilizó la versión 2 de Microsoft Kinect. Para lograr esto, fue necesario completar los pasos mostrados en la Figura 12.

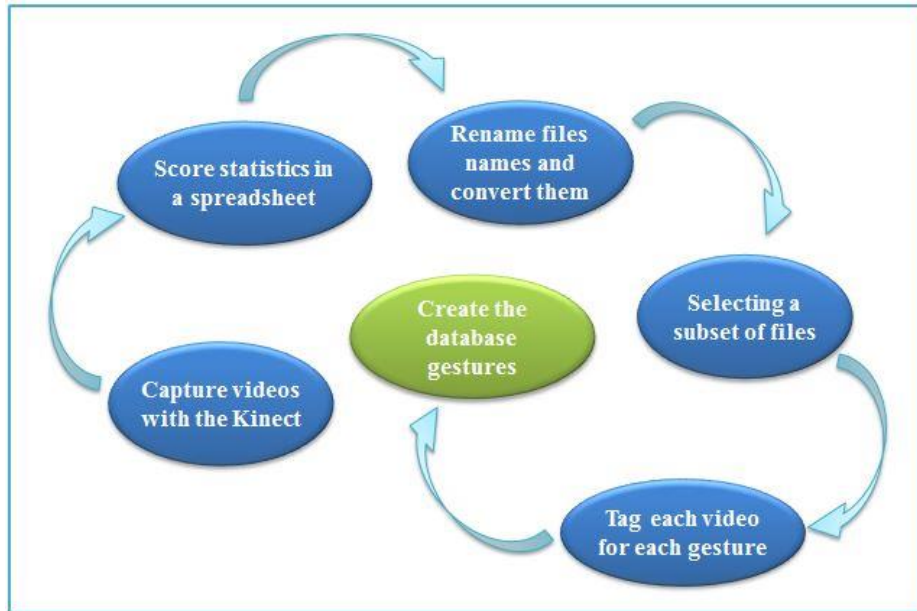


Figura 12 Pasos a seguir para la construcción del DataSet de gestos entrenados (Fuente propia)

Los pasos anteriores obedecen a recomendaciones hechas por los mismos creadores del SDK del Kinect 2. De igual forma, dentro de estos pasos, se siguió lo indicado en la Figura 9.

Primero se realizó varias capturas con el Kinect, usando el Microsoft Kinect Estudio (KS). Para evitar un sobreajuste del algoritmo de reconocimiento (overfitting), se usaron distintos escenarios y personas posibles. Durante la captura se fue registrando algunos datos relevantes en una hoja de cálculo. Estos podrían ser usados en posteriores experimentos y por otros equipos de investigación en robótica de servicio, cuando los datos sean puestos a disposición.

Los archivos resultantes de estas capturas están en formato Raw, que para poder ser usados en VBG, deben transformarse en formatos Xef, usando el convertidor KS. De igual modo, para hacer corresponder los archivos con los datos recabados en la hoja de cálculo, se usó una nomenclatura específica en los nombres de estos. Dado que se aplicó distintos cambios en las capturas realizadas, se decidió usar solo un subconjunto de todos los datos disponibles, para experimentar si con estos era suficiente para el entrenamiento del reconocedor.

Posteriormente se usaron estos videos seleccionados en el VBG, para la tediosa tarea de entrenar cada ademan con todos los videos. Para esto, se debió etiquetar cada video con ejemplos positivos y negativos. Los ejemplos positivos corresponden a las zonas donde las personas están ejecutando el ademán de interés. Los ejemplos negativos corresponden a los demás ademanes.

El resultado de todo esto es una base de datos de los ademanes entrenados en VBG, que se usarán en la aplicación desarrollada con Visual C#.

[DataSet](#)

Existen algunas colecciones de datos (DataSet) útiles para el entrenamiento de base de conocimientos usados en interacciones Humano-Robot (Aggarwal & Ryoo, 2011; Pavlakos et al., 2014; Shotton et al., 2013). Sin embargo, estos no usan el Kinect versión 2 y no son específicos para robot de servicio.

El conjunto de datos presentado aquí se recopiló utilizando la versión más reciente de Kinect: versión 2. Esto le da a este conjunto de datos una mejor calidad de las señales capturadas (Bernstein et al., 2012; Gowing et al., 2014; Zaman, 2011). Por lo tanto, este conjunto de datos se creó para que pueda ser escalable a futuras versiones de Kinect.

La computadora usada para la creación de la base de datos de ademanes cuenta con procesador Intel QuadCore i7-3537U de 2.5 Ghz, 8 Gb de memoria RAM, disco duro de 1Tb, tarjeta de video NVIDIA GeForce 610M y Windows 8.1

De acuerdo a las experiencias en los concursos de robótica de servicio, se decidió recolectar secuencias de videos con los siguientes ademanes:

1. Stop: el usuario extiende cualquiera de sus brazos al frente del robot, mostrando la palma de la mano. Con esto se quiere que el robot se detenga si está en movimiento.
2. Come: moviendo cualquiera de los manos de afuera hacia el hombro, el usuario le pide al robot que se acerque a él.
3. Left: usando solamente el brazo izquierdo, el usuario extiende la misma hacia un lado de él. De este modo el robot debe hacer un giro de 90° hacia su izquierda.
4. Right: usando solamente el brazo derecho, el usuario extiende la misma hacia un lado de él. De este modo el robot debe hacer un giro de 90° hacia su derecha.
5. Attention: con cualquiera de las dos manos, el usuario hace una especie de saludo, moviendo la mano de un lado al otro a la altura de la cabeza.
6. Indication: usando cualquiera de los dos brazos, el usuario le indica al robot un punto en particular. Esta indicación se hace hacia abajo y evitando poner el brazo extendido a la altura del hombro.
7. Turn: con cualquiera de los dos brazos, se hace un movimiento parecido a una palmada hacia abajo, por ejemplo, la mano derecha se mueve de derecha a izquierda, a la altura del torso, con la palma abierta. Esto le indicará al robot que debe hacer un giro de 180° en su eje, en cualquier dirección.

Para una mejor comprensión de cómo son los gestos, en la Figura 13 se muestran algunas secuencias de cada ademan.

Se le solicitó a un grupo de personas que ejecutaran estos ademanes al robot. Los ademanes considerados se seleccionaron para ser altamente intuitivos y útiles para interactuar con un robot de servicio.

Para registrar a las personas que realizan los gestos, se utilizó Microsoft Kinect Studio™. La Figura 14 muestra una captura de pantalla de esta aplicación. Para crear la base de datos de gestos se utilizó Microsoft Visual Gesture Builder™ (VGB) (la Figura 15 muestra una captura de pantalla de esto). Con este software es posible etiquetar los gestos interesados para entrenar la base de datos.

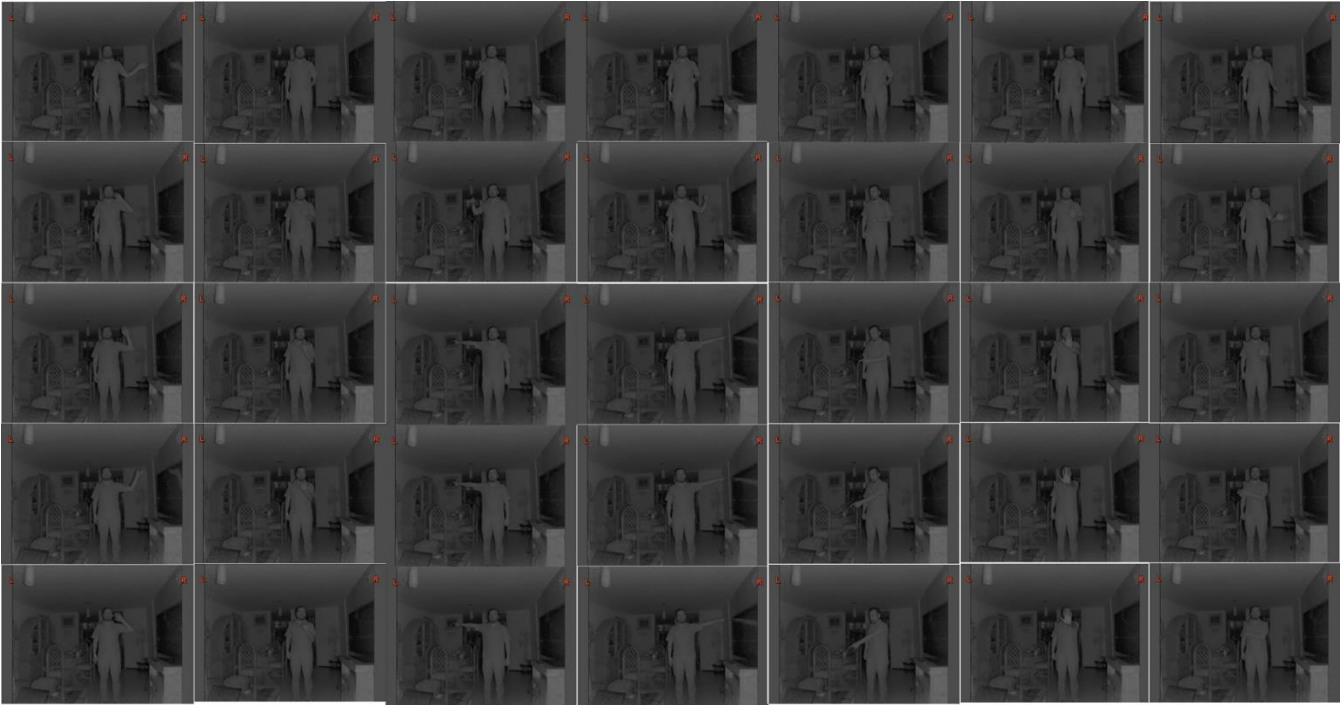


Figura 13 se muestra algunas secuencias (por columna) de los ademanes considerados. Desde la primera a la última columna los ademanes son: attention, come, left, right, pointing, stop y turn.

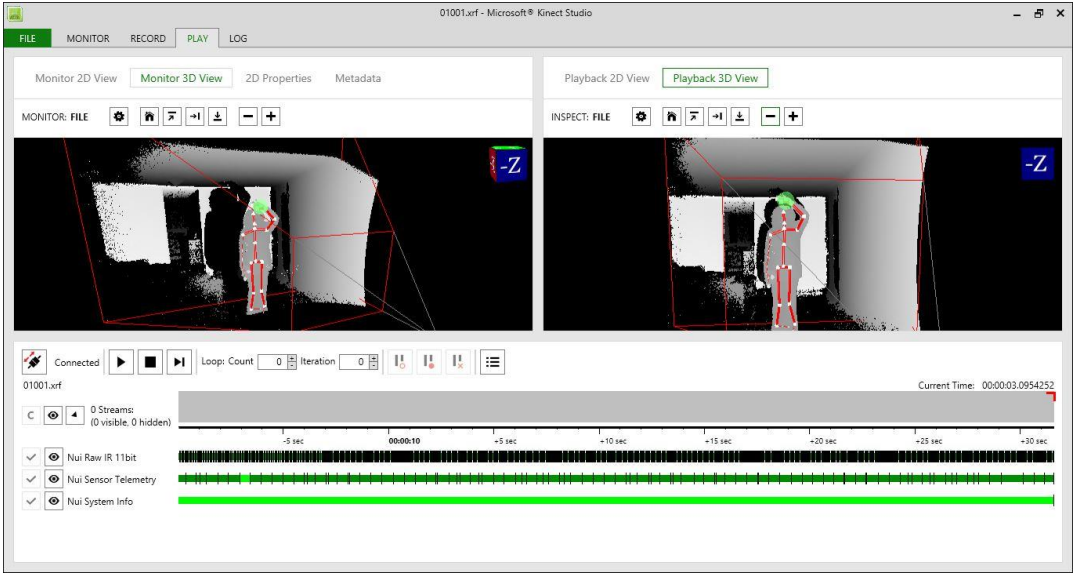


Figura 14 Kinect Estudio de Microsoft (Captura de pantalla)

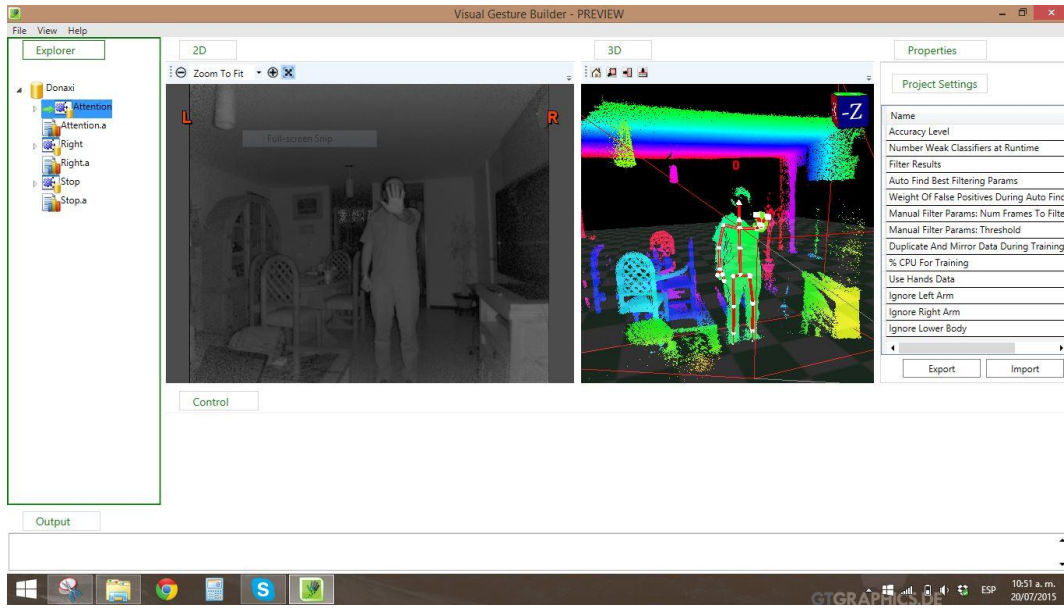


Figura 15 Visual Gesture Builder VGB (Captura de pantalla)

Este conjunto de datos se usará para entrenar una base de datos de ademanes y posteriormente se usará en una aplicación. Es importante que los datos representen las situaciones más realistas donde el robot de servicio será utilizado por su propietario. Por esta razón, la forma de captura del conjunto de datos fue muy cuidadosa. Incluso, se consideraron las recomendaciones dadas por el personal de Microsoft sobre la mejor manera de capturar muestras para gestos de entrenamiento.

Consideraciones de captura de datos

En primer lugar, se consideraron diferentes ambientes, es decir, los datos se capturaron en ambientes domésticos realistas, sin cambiar ningún mueble en las diferentes casas. Además, por razones experimentales, se utilizó el Laboratorio de Robótica de la UPAEP. Por estas razones, debido a que era necesario mover el Kinect en diferentes lugares, el sensor se colocó en un trípode, como se muestra en la Figura 16. La altura de este trípode se mantuvo igual en todas las situaciones, a 1.15 cm, que es la altura donde se instalará el Kinect en Donaxi.

Otro aspecto considerado fue la distancia del usuario. Debido a que el robot de servicio se moverá, la distancia entre él y su propietario puede variar constantemente. Por lo tanto, se utilizaron diferentes distancias entre el usuario y el Kinect. Sin embargo, es imposible considerar todas las diferentes distancias posibles. Por esta razón, se consideraron solo tres distancias diferentes: la más cercana posible, la más alejada posible y la distancia media de las dos anteriores. Cada situación es diferente para cada usuario, por lo tanto, la distancia más cercana se determinó cuando el Kinect perdió al usuario cuando se acercó a él. De la misma manera, la distancia más lejana se determinó cuando el Kinect perdió al usuario cuando se alejó de él. Finalmente, la distancia media es exactamente la mitad entre esas dos distancias antes.

De la misma manera antes descrita, el usuario puede estar frente o detrás del robot de servicio cuando realiza los ademanes. En consecuencia, para cada distancia hay dos sesiones de grabación del usuario: frente y detrás del Kinect.



Figura 16 Kinect y el trípode usado para la captura de los datos (Imagen propia).

También se contemplaba el efecto de la luz sobre el Kinect. Debido a que este sensor usa luz infrarroja, esto puede verse afectado por la luz ambiental alta. Por lo tanto, en algunos casos, existen dos sesiones de registro para cada usuario, de día y de noche. En el día no se utilizaron luces artificiales, frente a donde se encontraba la sesión de grabación por la noche.

Además, es posible que el robot de servicio tenga que ser comandado por un niño. Por ejemplo, cuando el nieto del usuario tiene que ayudar a su abuelo usando el robot. Por lo tanto, entre los usuarios se consideró un niño para este conjunto de datos, sólo si era capaz de realizar todos los ademanes frente al Kinect.

Finalmente, hay que considerar qué mano se usará para hacer los gestos: mano derecha o izquierda. En este caso, todos los gestos se realizaron solo con la mano derecha. Esto se debe a que la herramienta utilizada para entrenar la Base de datos de gestos tiene una funcionalidad para reflejar las acciones realizadas por el usuario, es decir, cuando los usuarios hacen los gestos con la mano derecha, esta herramienta toma estos movimientos y la réplica como si los usuarios realizaran las mismas acciones con la mano izquierda. Obviamente los gestos izquierdo y derecho sólo se pueden realizar con una mano. Además, los gestos de parada, ven y giro sólo se pueden realizar frente al robot.

Para garantizar la sostenibilidad de este conjunto de datos, se siguió la recomendación del personal de Microsoft en cuestión para registrar los datos con una extensión de archivo particular. Esto permite a este conjunto de datos que, cuando aparece una nueva versión de Kinect, estos se pueden transformar en esta nueva versión. Por esta razón, los pasos seguidos para registrar estos datos fueron los mostrados en la Figura 9. Primero, usando Kinect Studio, los movimientos del usuario se registraron en archivos .raw. Estos archivos se deben convertir a la versión de Kinect en la que desea utilizarlo. Esto se puede hacer con KS Convert, con lo que estos archivos se convierten en extensiones .xef. Este tipo de archivos se pueden usar en cualquier estudio de Kinect para visualización o en Visual Gesture Builder para construir la Base de datos de gestos. Esta base de datos de gestos se puede utilizar en cualquier aplicación creada con Visual Studio.

Captura de datos

Se le solicitó al usuario realizar los ademanes, mostrando un video por cada ademan para enseñarle cómo ejecutar los movimientos. Con esto, cada usuario tenía las mismas instrucciones en el momento de entrenarlo. De manera similar, para garantizar ciertos números equilibrados de cada ademan, se le pidió a cada usuario en cada sesión, realizar al menos una vez cada ademan. Después de eso, el usuario podría hacer cualquier repetición de cualquier ademan.

En resumen, a cada usuario, se siguieron los siguientes pasos para registrar sus ademanes:

1. Una sesión de día y otra de noche.
2. En cada sesión, los usuarios realizan seis secuencias: tres frente a Kinect y tres detrás.
3. Cada secuencia está a diferentes distancias: cerca, lejos y en el medio.

Se debe capturar muchos videos de diferentes personas en diferentes lugares. Esto se debe a que, con más ejemplos diferentes proporcionados al algoritmo de entrenamiento, se evita un exceso de ajuste y permitirá detectar a cualquier persona en cualquier entorno. Además, para realizar un seguimiento de la data capturada, se han escrito datos relevantes en una hoja de cálculo. Luego, para conocer el contenido de cada archivo de video sin abrirlo, se cambió el nombre usando una nomenclatura específica, para luego realizar una conversión con la herramienta KSConvert del Kinect SDK. Debido a que se usaron varios tipos de capturas con varios tipos de personas, se decidió usar solo una parte del conjunto completo, para verificar si este subconjunto era suficiente para el entrenamiento y la prueba del reconocedor. Para entrenar cada gesto, se debe mostrar en cada video dónde se encuentran los ejemplos positivos y los negativos, que se denomina etiquetado y se usó Microsoft Visual Gesture Builder. Finalmente, usando el VGB también, se puede crear la base de datos de gestos, para construir la aplicación para el robot de servicio.

Un punto importante es que en cada secuencia hay varias repeticiones de uno o más gestos. Por lo tanto, no es correcto suponer que cada archivo es una repetición de un gesto.

Estadísticas

En la Tabla 1 muestra el número total de gestos capturados hasta ahora para esta investigación.

Tabla 1 Numero totales de gestos en el DataSet

Gesto	Numero
Stop	194
Come	177
Left	395
Right	407
Attention	417
Indication	363
Turn	207

Otros datos importantes recabados son:

1. Se utilizaron 22 personas, 12 eran adultos, 2 jóvenes y 8 niños. Nadie de edad. Para cada persona se consideró el siguiente rango de edad:
 - a. Niño: menos de 15 años.
 - b. Joven: entre 16 y 21 años.
 - c. Adulto: entre 22 y 50 años.
 - d. 3° edad: más de 51 años.
2. Sólo tres personas fueron registradas en el laboratorio. El resto (19 personas) en sus casas (5 casas diferentes).
3. En la Tabla 2 se muestran algunos números sobre el total de gestos entre todas las secuencias, divididos para cada grupo.

Tabla 2 Algunos datos del DataSet. Distancias (Mínima, Media y Máxima), ocasiones o sesiones, de frente o de espalda al Kinect, y los totales por ademan.

Gesture	Session 0						Session 1						Total
	Min		Med		Max		Min		Med		Max		
	Front	back	Front	back	Front	back	Front	back	Front	back	Front	back	
1 Stop	38	0	33	0	32	0	31	0	30	0	30	0	194
2 Come	30	0	28	0	28	0	27	0	34	0	30	0	177
3 Left	30	36	32	42	33	30	23	31	34	36	35	33	395
4 Right	33	36	34	41	37	36	23	27	37	40	36	27	407
5 Attention	30	39	31	46	30	31	29	35	39	43	37	27	417
6 Indication	32	34	27	34	24	28	26	28	34	36	32	28	363
7 Turn	35	0	37	0	34	0	33	0	36	0	32	0	207

Algunos aspectos interesantes que se observaron durante la construcción de este conjunto de datos fueron:

1. A veces el Kinect confunde puertas abiertas y reflejos como personas.
2. Existe una fuerte correlación entre la altura del usuario y la distancia mínima que el Kinect puede seguir rastreando a la persona. Este no es el caso para distancias medias y lejanas.
3. Muchas veces el Kinect pierde al usuario cuando regresa de él.
4. Aunque la diferencia es muy pequeña, las distancias posibles al Kinect cambian cuando se graban las sesiones de día o de noche.

Aplicación

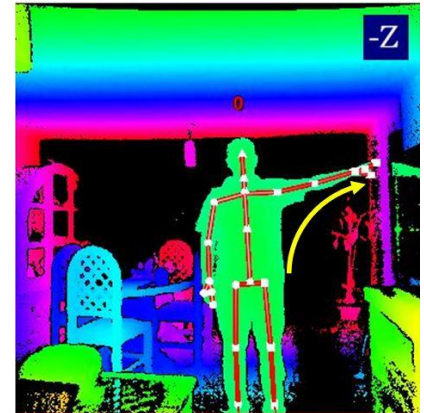
Una vez que esta base de datos de ademanes está disponible, se creó una aplicación para Donaxi, que pudo detectar algunos ademanes realizados por el usuario, a través del Kinect 2. Para este propósito, se utilizó una muestra en C# disponible en el SDK de Microsoft Kinect llamado "DiscreteGestureBasics". Para la competencia TMR, solo se usaron tres ademanes de los siete contemplados. Esto porque no se ha probado la precisión del reconocedor con los siete ademanes juntos en la base de datos. En la Figura 17 Se muestran los tres ademanes utilizados: atención, para que Donaxi sepa dónde está su dueño; a la derecha, para hacer que Donaxi gire a su derecha; y detenerse, por lo que Donaxi se detiene al acercarse. Estos tres ademanes se utilizaron en TMR2015.



(a) Attention



(b) Stop



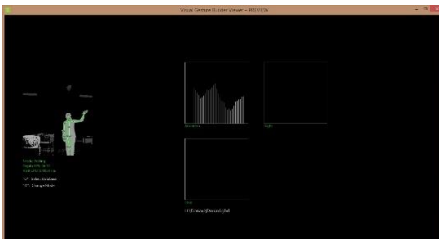
(c) Right

Figura 17 Tres ademanes entrenados para crear la base del sistema prototipo (Foto propia)

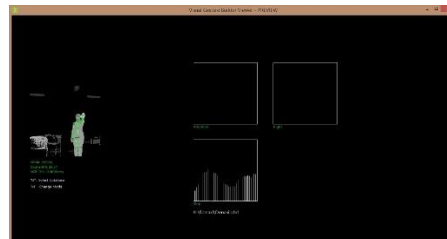
Se puede apreciar en la Figura 17 la flecha que muestra el movimiento en cada caso. Para cada ademan, se programó que el equipo emitiera un audio indicando el correspondiente ademan recién ejecutado.

Ahora bien, la aplicación desarrollada con tres ademanes mostró buenos resultados, lo que hizo que Donaxi hablara sólo cuando el usuario completó cualquiera de los gestos. Sólo el gesto "stop" tuvo algunos problemas de detección (falsos negativos). Por esta razón, es necesario preparar un plan de prueba más riguroso mientras se agregan más gestos y evaluar los resultados.

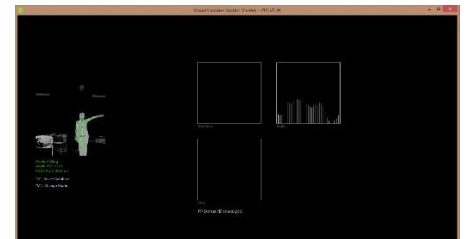
En la Figura 18 se puede apreciar el reconocimiento de gestos que se ejecuta en la herramienta Microsoft VGB LivePreview, donde el usuario realiza cada gesto a la vez.



(a) Attention



(b) Stop



(c) Right

Figura 18 Capturas de pantalla de la herramienta Microsoft VGB Live Preview donde la persona ejecuta cada uno de los tres gestos.

En cada figura se aprecia cómo aumenta considerablemente el nivel de confianza de cada ademan cuando lo ejecuta el usuario, mientras que los demás no muestran ningún valor (cuadros del lado derecho del esqueleto del usuario). Este mismo comportamiento se obtuvo en el robot Donaxi.

Sin embargo, se pudo observar cierta deficiencia en la detección del ademan de Stop, dejando de detectarlos cuando se ejecutaban (falsos negativos). Es por esta razón que es importante primero definir el plan de pruebas antes de ir adicionando más ademanes entrenados a esta base de datos.

Las apreciaciones visuales nos permiten confirmar que los resultados con solo estos tres gestos son muy buenos. Incluso la aplicación funcionó con ambas manos para detener y prestar atención, con los mismos resultados.

Además, como resultado de las experiencias obtenidas de este trabajo, se recomienda:

1. Construya experimentos muy bien planificados para verificar que este conjunto de datos es suficiente para permitir que un robot de servicio entienda estos gestos de su propietario.
2. En caso de que este conjunto de datos no sea suficiente, es necesario agregar más repeticiones de cada gesto. Especialmente de personas mayores, debido a que este conjunto de datos no contiene ejemplos de ellos.
3. Agregue más gestos de acuerdo con los nuevos requisitos del usuario para permitir que el usuario ordene el robot.
4. Permitir con este conjunto de datos un archivo que contenga los gestos etiquetados en VGB.
5. Comunicar la aplicación creada con Visual Studio con ROS (sistema operativo robótico) para operar adecuadamente el robot.

Conclusiones

Es evidente la necesidad de que los robots de servicio interactúen con las personas de la manera más natural posible. Sin embargo, lograr este tipo de interacción no es tan sencillo. Se avecinan muchos desafíos para lograr este tipo de relación robot-humano.

Por ahora, la mayoría de las investigaciones sobre la interacción humano-robot multimodales solo utilizan voz y ademanes, pero obviamente este conjunto puede crecer para alcanzar el objetivo esperado.

Esta investigación tiene como objetivo lograr este HRI multimodal utilizando sólo la versión 2 de Kinect para detectar ademanes y voz, y permitir que Donaxi pueda entender mejor a su propietario.

Con este conjunto de datos es posible dar a un robot de servicio la funcionalidad de reconocimiento de ademanes de cualquier persona, en casi cualquier ambiente. Por el hecho de que se recopiló en diferentes casas, permitió que estos datos sean más útiles cuando se utilizan en una aplicación para el robot.

Gracias a las herramientas proporcionadas por Microsoft Company fue posible recopilar este conjunto de datos y construir una aplicación que permita a un robot de servicio reconocer los ademanes de su propietario.

Esta recopilación es una de las pocas que está pensando especialmente en un robot de servicio. Sin embargo, es fácil ver que se puede usar en cualquier aplicación que requiera una interacción humano-maquina (Human Machine Interaction (HMI)).

Se tiene la hipótesis de que algunas consideraciones para recopilar estos datos eran innecesarias, por ejemplo: distancia y luz ambiente. Sin embargo, solo futuros experimentos pueden confirmar esto.

Referencias bibliográficas

- Aggarwal, J., & Ryoo, M. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43, 16:1-16:43. <https://doi.org/10.1145/1922649.1922653>
- Aracil, R., Balaguer, C., & Armada, M. (2008). Robots de servicio. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 5, 6–13. [https://doi.org/10.1016/S1697-7912\(08\)70140-7](https://doi.org/10.1016/S1697-7912(08)70140-7)
- Bernstein, G., Lotocky, N., & Gallagher, D. (2012). *Robot Recognition of Military Gestures CS 4758 Term Project*. Retrieved from <http://pr.cs.cornell.edu/humanactivities/handgesture/index.php>
- Droeschel, D., Stückler, J., Holz, D., & Behnke, S. (2011). Towards joint attention for a domestic service robot - Person awareness and gesture recognition using time-of-flight cameras. *Proceedings - IEEE International Conference on Robotics and Automation*, (May), 1205–1210. <https://doi.org/10.1109/ICRA.2011.5980067>
- Goodrich, M. a., & Schultz, A. C. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends® in Human-Computer Interaction*, 1(3), 203–275. <https://doi.org/10.1561/11000000005>
- Gowing, M., Ahmadi, A., Destelle, F., Monaghan, D. S., O'Connor, N. E., & Moran, K. (2014). Kinect vs. low-cost inertial sensing for gesture recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8325 LNCS, 484–495. https://doi.org/10.1007/978-3-319-04114-8_41
- International Federation of Robotics. (n.d.). *Introduction into Service Robots*. Retrieved from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?number=55890.
- International Federation of Robotics. (2018a). *IFR Press Conference*. Tokyo. Retrieved from

https://ifr.org/downloads/press2018/WR_Presentation_Industry_and_Service_Robots_18_Oct_2018.pdf

International Federation of Robotics. (2018b). *Service robots-global sales value up 39 percent says IFR*. Tokyo. Retrieved from www.ifr.org

ISO/TC 184, A. systems and integration. (2012). ISO 8373:2012(en), Robots and robotic devices — Vocabulary. Retrieved October 27, 2018, from <https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en>

ISO. (2017). ISO/IEC 30113-11:2017(en), Information technology — Gesture-based interfaces across devices and methods — Part 11: Single-point gestures for common system actions. Retrieved October 23, 2018, from <https://www.iso.org/obp/ui/#iso:std:iso-iec:30113:-11:ed-1:v1:en>

Jaimes, A., & Sebe, N. (2005). Multimodal Human Computer Interaction: A Survey. *Computer Vision in Human-Computer Interaction*, 1–15. https://doi.org/10.1007/11573425_1

Liu, L., Yang, S. H., Wang, Y., & Meng, Q. (2009). Home service robotics. *Measurement and Control*, 42(1), 12–17. <https://doi.org/10.1177/002029400904200104>

Marcel, S. (2002). *Gestures for multi-modal interfaces: A Review*. IDIAP Research Report (Vol. 2). Valais, Switzerland.

Mitra, S., & Acharya, T. (2007). Gesture Recognition : A Survey. *IEEE Transactions On Systems, Man, And Cybernetics - Part C: Applications And Reviews*, 37(3), 311–324.

Moradi, H., Kawamura, K., Prassler, E., Muscato, G., Fiorini, P., Sato, T., & Rusu, R. (2013). Service robotics (the rise and bloom of service robots). *IEEE Robotics and Automation Magazine*, 20(3), 22–24. <https://doi.org/10.1109/MRA.2013.2271580>

Nieuwenhuisen, M., Stuckler, J., & Behnke, S. (2010). Intuitive multimodal interaction for service robots. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 177–178.

<https://doi.org/10.1109/HRI.2010.5453206>

- Pavlakos, G., Theodorakis, S., Pitsikalis, V., Katsamanis, S., & Maragos, P. (2014). KINECT-BASED MULTIMODAL GESTURE RECOGNITION USING A TWO-PASS FUSION SCHEME. In *Proc. Int'l Conf. on Image Processing*. Retrieved from http://cvsp.cs.ntua.gr/publications/confr/PTPKM_MultimodalGestureRecogn2PassFusion_ICIP2014.pdf
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., ... Blake, A. (2013). Real-time human pose recognition in parts from single depth images. *Studies in Computational Intelligence*, 411, 119–135.
- Transparency Market Research. (2018). IT Robotic Automation Market to be worth US\$ 11,417.4 Mn by 2026 - TMR. Retrieved October 20, 2018, from <https://www.transparencymarketresearch.com/it-robotic-automation-market.html>
- Vasquez Chavarria, H., Escalante, H. J., & Sucar, L. E. (2013). Simultaneous segmentation and recognition of gestures for human-machine interaction. In J. Gama, M. May, N. Marques, & P. Cortez (Eds.), *Workshop on Ubiquitous Data Mining* (p. 29). Beijing, China: AAI Press / International Joint Conferences on Artificial Intelligence. Retrieved from <http://www.liaad.up.pt/udm/workshop-proceedings>
- Watson, R. (1993). A Survey of Gesture Recognition Techniques Technical Report TCD-CS-93-11. *Department of Computer Science, Trinity College Dublin*, 2(July), 17. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Survey+of+Gesture+Recognition+Techniques+Technical+Report+TCD-CS-93-11#0>
- Wu, Y., & Huang, T. S. (1999). Vision-Based Gesture Recognition : A Review. *Lecture Notes in Computer Science*, 1739(Lecture Notes in Artificial Intelligence), 103–115. https://doi.org/10.1007/3-540-46616-9_10
- Yan Bian, Xuanfang Wang, Li Zhao, & Shigang Cui. (2016). A new control system

for space robot based on Brain-Computer Interfaces. In *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)* (pp. 232–236). Xian: IEEE.
<https://doi.org/10.1109/IMCEC.2016.7867207>

Zaman, A. (2011). Natural Interaction Interfaces using the Microsoft Kinect. *Virtual Reality*, (March). Retrieved from
http://ahmedzaman.co.cc/downloads/20305028_Dissertation.pdf

Zielinska, T. (2016). Professional and Personal Service Robots. *International Journal of Robotics Applications and Technologies*, 4(1), 63–82.
<https://doi.org/10.4018/IJRAT.2016010104>