Universidad Autónoma del Estado de Morelos

Facultad de Ciencias

Percepción de Distancia en un Agente Artificial a través de la Simulación de Esquemas Sensorimotrices

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
Doctor en Ciencias

(Modelación Computacional y Cómputo Científico)

PRESENTA:
Wilmer Gaona Romero

DIRECTOR DE TESIS:
Dr. Bruno Lara Guzmán

3 de junio de 2014

Alcanzó a cerrar otra vez los párpados, aunque ahora sabía que no iba a despertarse, que estaba despierto, que el sueño maravilloso había sido el otro, absurdo como todos los sueños...

La noche boca arriba, Julio Cortázar.

RESUMEN

Percepción de Distancia en un Agente Artificial a través de la Simulación de Esquemas Sensorimotrices

Wilmer Gaona Romero

de ultimo...

Índice general

Capítulo 1
Introducción

La percepción del espacio y en particular la percepción de distancia es un proceso que ha sido estudiado de forma exhaustiva dentro de las ciencias cognitivas y que aún hoy sigue siendo un tema de investigación (Turvey (2004)). De acuerdo a la teoría de la percepción ecológica (Gibson (1979)), la percepción de distancia no es un proceso geométrico sino una propiedad emergente basada en la asociación de la información multi-modal de un agente, es decir, en su información motriz y sensorial (Braund (2007)). Esta concepción para la percepción de distancia constituye una fuente de inspiración para proponer nuevos modelos en agentes artificiales que den cuenta de esta capacidad.

Además de la asociación de información multi-modal, se han investigado otros aspectos que intervienen en nuestra percepción del espacio, tales como el uso de unidades relativas escaladas a nuestras dimensiones corporales (Proffitt (2006)), el contexto donde interactuamos (Gibson (1979)) y la preparación de cada una de nuestras acciones (Wexler and Boxtel (2005)). Estos factores también deben ser tenidos en cuenta al modelar de forma computacional la capacidad para percibir la distancia.

Tradicionalmente, en la robótica móvil que hace uso de la visión artificial, la estimación de la distancia a los objetos del entorno se basa en el uso un modelo de cámara. Esta estimación consiste en el cálculo de un valor numérico a partir de relaciones geométricas, de un procedimiento de calibración de cámara y de la realización de diversos cálculos matemáticos (Moons (1998)). Sin embargo, aunque el valor calculado es exacto, este necesita ser interpretado por el diseñador del robot o de algún otro agente externo.

Nosotros los humanos de acuerdo a nuestras capacidades sensori-motrices, tenemos la capacidad para interpretar adecuadamente la información acerca de la distancia a los objetos que nos rodean, incluyendo además de la información visual a la información motriz (Proffitt (2006)). Esta capacidad nos permite desenvolvernos eficazmente en nuestro entorno. Por lo tanto, un conocimiento espacial adecuado para que un agente artificial sea capaz de interactuar en una forma congruente en su entorno, debería ser un conocimiento cimentado que incluya las propias capacidades sensori-motrices del agente artificial en cuestión.

La inteligencia artificial clásica concibe a la cognición como un procesamiento de la información basado en la manipulación secuencial de símbolos abstractos e inspirado por el trabajo de (Newell and Simon (1976)). En dicho trabajo se propuso la hipótesis: "The Physical Symbol System Hyphotesis "[1] la cual sostiene que la inteligencia proviene de la creación y manipulación de representaciones de la realidad. Esta propuesta estuvo presente en la mayoría de los laboratorios de inteligencia artificial, hasta la década de los 90's en la que trabajos como el del roboticista Rodney Brooks empezaron a cuestionar esta concepción (Brooks (1991)).

Trabajos como el de Brooks y críticas como la del filósofo John Searle (Searle (1980)), abrieron la posibilidad para el surgimiento de una nueva postura entorno al estudio de la inteligencia artificial, conocida como la nueva inteligencia artificial o la robótica cognitiva (Pfeifer and Scheier (1999)).

[1] Traducción libre por el autor: sistema de símbolos físicos

Esta nueva disciplina propone estudiar a la cognición a través de agentes artificiales dotados de un cuerpo que les permite interactuar en un entorno determinado. Además se inspira de los estudios provenientes de la cognición humana para estudiar los procesos que subyacen a la cognición a través de la implementación y prueba de modelos cognitivos en agentes artificiales (Asada et al. (2009)).

La nueva inteligencia artificial es una disciplina que está sustentada en el paradigma de la *cognición cimentada* (Barsalou (2008)). Este paradigma se caracteriza por considerar a la cognición como un fenómeno emergente a partir de la interacción de un agente con su entorno.

Dentro de la cognición cimentada existe un área de investigación que se ha dedicado a estudiar las *simulaciones internas*, entendidas como la recreación de situaciones sensoriales y motrices. En el trabajo de (Jeannerod (1995)) se hace referencia a estudios experimentales para mostrar como el sistema motriz interviene en procesos como la imaginación de acciones, el reconocimiento de herramientas, el aprendizaje por observación y el reconocimiento de las acciones del otro. Esto le permitió a (Jeannerod (1995)) proponer y desarrollar la hipótesis de que el sistema motriz es parte de una red neuronal en el sistema nervioso central humano dedicada a la realización de simulaciones internas.

Un mecanismo, que aunque muy debatido en sus inicios, y conocido como *imaginería mental* (Kosslyn (1994)) actualmente constituye el ejemplo mas ilustrativo de la realización de estas simulaciones internas en humanos (Barsalou (2008)). La imaginería mental es una experiencia interna que se asemeja a una experiencia perceptual real y que ocurre en la ausencia de un estímulo físico y sin la necesidad de la ejecución de forma explícita de una acción motriz.

Para que un agente artificial pueda navegar a través de un pasaje con obstáculos es necesario que ejecute una serie de acciones evasivas, en las cuales está involucrada la predicción de una consecuencia sensorial eventualmente nociva y la realización de una acción motriz que permita evitar dicha consecuencia. Este acople entre una predicción sensorial y una acción motriz se ha propuesto como uno de los mecanismos utilizados por el sistema nervioso central para el control motriz en humanos (Wolpert et al. (1995)).

Uno de los modelos que se han formulado en un intento por contextualizar los aspectos funcionales de las simulaciones internas son los *modelos internos*. Estos modelos se han propuesto como esquemas que pueden dar cuenta de diversos procesos relacionados con el cerebelo como la predicción de estados sensoriales,la cancelación de los efectos sensoriales del movimiento, la transformación de un error en coordenadas sensoriales a coordenadas motrices (Wolpert et al. (1998)), e incluso la imaginería mental (Grush (2004)).

Los modelos internos se pueden dividir en dos clases: el modelo inverso y el modelo directo y aunque se han trasladado hacia las neurociencias, su origen se encuentra en la teoría del control (Jordan and Rumelhart (1992)). El modelo inverso es un controlador que dado un estado sensorial actual y uno deseado provee el comando motriz necesario para llevar al sistema al estado sensorial deseado.

Por otro lado, el modelo directo es un modelo predictor, capaz de proveer las consecuencias sensoriales de llevar a cabo una acción, es decir, capaz de realizar una predicción sensorial en base a la situación actual y a una acción motriz que se pretende ejecutar. Esta capacidad de realizar predicciones sensori-motrices es considerada una pieza clave para la ejecución adecuada de movimientos coordinados en nosotros los humanos (Sporns and Edelman (1993)).

Esta tesis propone la implementación computacional de un *modelo directo* sobre un agente artificial. Este modelo le otorga la posibilidad al agente de disponer de un repertorio de posibles consecuencias sensoriales dado un conjunto específico de comandos motrices. Esta capacidad para realizar predicciones sensoriales puede ser utilizada para evaluar posibles escenarios futuros donde

se presenten situaciones nocivas para el agente y en base a estas determinar una acción correctiva adecuada.

El principal interrogante que surge es como la implementación computacional del modelo directo, podría dar cuenta en un agente artificial de una noción de la distancia a los objetos que le rodean en una forma semejante a como sucede en los humanos. Esta noción de distancia tendría que estar basada en la asociación de la información multi-modal (motriz y sensorial) del propio agente, tal como (Braund (2007)) lo propone y estar escalada a sus capacidades físicas como lo sugiere (Proffitt (2006)) y sin la necesidad de hacer uso de un modelo geométrico como el utilizado tradicionalmente en el área de la visión artificial (Moons (1998)).

Para logra una noción de distancia en un agente artificial cimentada en sus capacidades sensori-motrices, en primer lugar se estudia la manera en que el modelo directo de cuenta de una asociación sensorial multi-modal dentro de un proceso artificial de imaginería mental. Esto constituye la primera pieza clave en este trabajo, en el que las capacidades sensori-motrices de un agente artificial están representadas por las modalidades motriz, visual y táctil.

El primer objetivo de esta tesis es lograr lograr una asociación sensorial multi-modal en un agente artificial enmarcado dentro de la imaginería mental y en base a las predicciones provistas por el modelo directo. Para tal efecto se dispone de un modelo directo implementado sobre un robot móvil y codificado a través de un sistema de redes neuronales artificiales. Este se encuentra reportado previamente en el trabajo de (Escobar et al. (2012)).

Como modalidad visual se calculó un vector de disparidad -imagen en escala de grises de una sola fila de píxeles y que indica la distancia a los objetos del entorno en base al valor de intensidad de cada píxel- a partir de las imágenes izquierda y derecha captadas por el par de cámaras estéreo que dispone el robot y de un modelo geométrico cada una de estas.

La modalidad táctil se representó a través de una señal binaria que codificó un estado de colisión o no colisión a partir de las señales de los parachoques del robot. Estos datos fueron recolectados mientras el robot se desplazaba hacia adelante en intervalos constantes de 15 cm en un ambiente con obstáculos rodeando al robot.

Aunque en el sistema propuesto por (Escobar et al. (2012)) se utilizó un modelo geométrico de cámara (Moons (1998)) para representar en la modalidad visual los datos que codifican la distancia a los objetos, es suficiente para lograr la asociación multi-modal buscada.

Al ubicar al robot móvil en diferentes orientaciones y en el centro de una arena de obstáculos, en lugar de ejecutar de forma explícita movimientos hacia enfrente, se llevan a cabo una serie de predicciones sensori-motrices a través del modelo directo implementado. De esta forma se dispone de las consecuencias visuales y táctiles, que eventualmente producirían la ejecución de dichos movimientos.

A la par que se realizan las predicciones sensori-motrices, se va lleva a cabo un proceso de aprendizaje artificial tipo Hebbiano (Hebb (1949)). Este proceso logra decantar en una matriz de pesos sinápticos la asociación entre estas las modalidades visual y táctil. Inicialmente, los pesos de esta matriz fueron inicializados con valores aleatorios, pero a medida que las simulaciones internas se llevaban a cabo comenzó a emerger una estructura en esta matriz con la forma de una diagonal.

Una vez que se obtuvo una diagonal definida en esta matriz, se detuvo la realización de las predicciones sensori-motrices. La forma final de esta matriz es una prueba clara de la factibilidad de lograr una estructuración en la información sensorial de entrada correspondiente los datos de las modalidades visual y táctil.

Cabe resaltar que este procedimiento fué llevado a cabo en una forma introspectiva, es decir, se obtuvo una asociación multi-modal a través de un proceso de imaginería mental mediante el uso de

un modelo directo. Esto abre la puerta hacia un nuevo paradigma en el que el hecho de disponer de predicciones sensori-motrices puede ser aprovechado para el estudio de la imaginería mental en la cognición cimentada a través de la modelación computacional, tal metodología se ha descrito en el trabajo de (Pezzulo et al. (2012)).

Aunque se obtuvo una asociación multi-modal para las modalidades visuales y táctiles en un robot móvil, la codificación de los datos todavía descansa sobre una base geométrica, y según lo expuesto anteriormente la percepción de distancia en humanos no radica en un modelo geométrico (Gibson (1979); Braund (2007)). Por lo tanto, para dotar a un agente artificial con una noción de la distancia cimentada en sus capacidades sensori-motrices, requiere que en la información visual no esté codificada de forma explícita la información relacionada con las distancia.

En segundo objetivo de esta tesis es desarrollar un modelo capaz de dotar a un agente artificial con la capacidad de adquirir una noción de distancia sin recurrir a la geometría y a los modelos de cámara (Tsai (1987); Moons (1998)). Para esto se propone hacer uso únicamente de las imágenes originales que provee el par de cámaras estéreo para representar la información visual. Esto permite ubicar al presente trabajo en un nivel de abstracción mas alto en el que se reemplaza una estructura de información como un vector de disparidad que explícitamente codifica la distancia a los objetos, por otra en la que únicamente se dispone de las dos imágenes crudas que solo contienen información pictórica de la escena.

La característica especial es que esta noción de distancia a los objetos de su entorno está cimentada en las capacidades sensori-motrices propias del agente dando cuenta de un significado corporizado e intrínseco para el agente en sí. Para realizar esto, se implementó de forma computacional un modelo directo codificado a través de un sistema de redes neuronales y donde el estado sensorial fué representado a través de las modalidades visual y táctil, mientras que el espacio motriz por tres movimientos diferentes, un movimiento hacia adelante y giros hacia la izquierda y derecha.

Como resultado se obtuvo que al colocar un obstáculo a diferentes distancias, las predicciones para la modalidad táctil producidas por el modelo directo mostraron la aparición de un valor umbral constante para dichas predicciones táctiles. Este valor umbral, es un aspecto que va acorde con uno de los principios de diseño de agentes artificiales completos que se propone en (Pfeifer and Scheier (1999)) y concerniente al principio del valor, el cual establece la existencia de alguna medida intrínseca que le indique al agente la conveniencia de experimentar una determinada situación sensorial.

Las simulaciones internas provistas por este modelo directo fueron usadas por el agente artificial para expresar la distancia a los obstáculos presentes en el entorno. Esta expresión estuvo basada en el número de movimientos requeridos y/o permitidos antes de que se colisione con un obstáculo situado enfrente del agente. De esta manera se cimentó el concepto de *distancia a colisión* en una forma tal en que las modalidades sensoriales y motrices estuvieron interrelacionadas, acorde a lo propuesto en la robótica cognitiva (Pfeifer and Scheier (1999)).

El tercer objetivo de esta tesis es estudiar como comportamientos complejos pueden emerger en base a habilidades sensori-motrices previamente aprendidas. Esto se enmarca dentro de la robótica del desarrollo, la cual es un área de investigación dentro de la robotica cognitiva que sustenta que el surgimiento de habilidades y capacidades sensori-motrices en los agentes artificiales siguen una línea secuencial, en la que primero aparecen habilidades básicas las cuales posteriormente permiten que emerjan comportamientos mas complejos, tal como sucede en los humanos (Asada et al. (2009)).

Tomando como referencia este tipo de desarrollo secuencial de capacidades sensorimotrices, se propone que un comportamiento como el de elegir atravesar entre dos aperturas o el de navegar a través de un pasaje de obstáculos; puede ser obtenido en un agente artificial en base a la capacidad adquirida previamente para determinar la distancia a los objetos del entorno.

Para ilustrar esto se llevaron a cabo dos experimentos diferentes. En primera instancia se situó a un agente artificial frente a dos pasajes o entradas con el propósito de determinar por cual de ellos podría pasar sin colisionar. En este caso se cimentó el concepto *pasabilidad*. Por medio de la realización de un proceso de simulaciones internas, el agente fué capaz de encontrar el pasaje mas amplio de manera segura, sin la necesidad de ejecutar algún movimiento de forma explícita. Este experimento se relaciona al reportado en humanos por (Warren and Whang (1987)).

En segunda instancia al situar a un agente artificial en un laberinto de obstáculos, este fué capaz de navegar sin colisionar hasta encontrar la salida utilizando el mismo principio de simulaciones internas que en el caso anterior. La característica principal estuvo en el hecho de que las predicciones sensoriales producidas por el modelo directo le permitieron anticipar una futura colisión y a su vez elegir la dirección con menor probabilidad de colisión como una acción correctiva en el curso de su trayectoria. Por consiguiente se ejecutaron únicamente los comandos motrices que le permitieron al agente encontrar la salida al laberinto anticipándose a situaciones potencialmente nocivas como una colisión. Este forma de control puede considerarse como una implementación computacional de uno de los mecanismos a los que recurre el sistema nervioso central para realizar un adecuado control motriz (Wolpert et al. (1995)).

Las preguntas de investigación que originaron este trabajo son las siguientes:

1. ¿En que forma el modelo directo podría cuenta de un proceso artificial de imaginería mental?

2. ¿Cómo se podría dotar a un agente artificial con una noción de distancia cimentada en sus capacidades sensori-motrices?

3. ¿Podría esta noción de distancia ser obtenida a partir de una asociación multi-modal en el agente artificial, tal como se ha propuesto que ocurre para los humanos?

4. ¿Podría este modelo dar cuenta de la exhibición de comportamientos mas complejos en el agente, como la navegación en su entorno?

1.1. Hipótesis

La hipótesis de trabajo para esta investigación radica en considerar el modelo directo como un mecanismo básico para implementar sobre los agentes artificiales un proceso de imaginería mental y a su vez dotar a un agente artificial con la capacidad de una noción o juicio de distancia basado en un proceso de simulación sensorimotriz generado mediante el uso del modelo directo.

1.2. Publicaciones

- **Gaona, W.**, Hermosillo, J., and Lara, B. (2012). Distance perception in mobile robots as an emergent consequence of visuo-motor cycles using forward models. In Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2012 IEEE Ninth, pages 42–47.

- **Gaona, W.**, Escobar, E., Hermosillo, J., and Lara, B. (2014). Anticipation by multi-modal association through an artificial mental imagery process. In Connection Science, 2014. En impresión.

- **Gaona, W.**, Escobar, E., Hermosillo, J., and Lara, B (2014). Adquisición de conceptos espaciales en un agente autónomo artificial a través de simulaciones internas. In Revista Nova Scientia. En revisión.

Capítulo 2

Percepción de distancia

La percepción de distancia es la habilidad que poseemos los seres humanos de percibir el espacio existente entre los objetos que nos rodean y nuestro cuerpo. Si se tiene en mente que una adecuada percepción de distancia es el requisito primordial para llevar a cabo comportamientos mas complejos como la navegación, la orientación, la manipulación de objetos y la elaboración de un mapa del entorno, esta habilidad adquiere un carácter esencial para que podamos desenvolvernos adecuadamente en nuestro mundo.

Esta capacidad es un tema de investigación actual debido a que no se ha podido comprender en su totalidad (Turvey (2004)). A partir de diversos estudios provenientes de las ciencias cognitivas, se ha retomado la concepción de que la percepción de distancia en los seres humanos es una capacidad que emerge a través de la intervención de diversos factores, entre los que se encuentran:

1. Lo que sucede es una asociación de información multimodal, es decir, que la percepción de distancia es el resultado de la asociación de la información que aportan dos modalidades sensoriales, específicamente la modalidad visual y la táctil (Braund (2007)).

2. Las características físicas de nuestro cuerpo determinan la forma en que percibimos la distancia, según el cual, no expresamos las relaciones espaciales que existen entre los objetos en términos de unidades absolutas, sino que usamos términos relativos, basados en ciertas unidades que están escaladas a nuestras propias dimensiones corporales (Proffitt (2006)).

3. Retomando lo propuesto por Gibson en (Gibson (1950)), es en la propia interacción agente-ambiente donde se definen los usos potenciales que adquieren los objetos para el observador y que son estos usos los que definen el concepto de espacio y no simplemente entidades como puntos, lineas o planos pertenecientes a una geometría abstracta.

4. La acción y el entorno están vinculadas con nuestra percepción de distancia de tal forma que la preparación o la ejecución de una acción motriz sobre el entorno es suficiente para modificar la percepción y representación del espacio que nos formamos como observadores (Wexler and Boxtel (2005); Lappin et al. (2006)).

Con el propósito de comprender de una forma mas completa los demás aspectos que intervienen en la percepción de distancia se describe a continuación como surge esta capacidad en base al desarrollo del sistema visual humano desde el momento de gestación.

El desarrollo de la percepción de distancia comienza desde la etapa de la gestación del infante. En el artículo de (Law et al. (2011)) se puede encontrar una revisión exhaustiva de los principales aspectos de la maduración de los sistemas biológicos del bebé desde la gestación y hasta el primer año de vida. Dentro de estos, se pueden encontrar los cambios en el sistema visual concernientes a la percepción de profundidad, los cuales se describen en seguida.

El desarrollo del sistema visual comienza entre las semanas 8 y 16 de gestación. Las células fotorreceptoras (conos y bastones), al igual que las neuronas bipolares de la retina se pueden encontrar

hacia el final de este periodo, aunque su desarrollo no se completa antes del nacimiento. Hacia final de este periodo surgen los primeros movimientos del globo ocular.

Hacia la semana 23, los ojos son capaces de ejecutar movimientos suaves y rápidos en patrones complejos las cuales incluyen rotaciones. Para la semana 26 los ojos se abren y el sistema visual se encuentra activo. Desde la semana 30 se incrementa el número de neuronas en la corteza visual del cerebro y ya se pueden observar movimientos oscilatorios rápidos en el globo ocular.

Desde del nacimiento y hasta el mes 1, los bebés tienen una percepción burda del color a través de una imagen retinal difusa. Son atraídos por estímulos visuales simples como luces tenues, estímulos de movimiento, y el rostro de la madre. La región periférica de la retina está casi madura, aunque no así la región donde se encuentra la fóvea - una pequeña región en la retina - donde se percibe con mayor detalle gracias a la alta concentración de conos y a que la luz que refleja un objeto observado es concentrada en esta región. Aunque con poca frecuencia, ya se pueden observar movimientos sacádicos.

Durante los meses 1 y 2 de vida, la discriminación del color es similar a la de un adulto y con movimientos sacádicos capaces de centrar su atención en objetos estáticos y en movimiento.

Entre los meses 2 y 3 aparece la convergencia de los ojos sobre objetos acercados al bebé e inicia la aparición de la *estereopsis* (impresión de la profundidad). Al final del mes 4 la percepción del color es comparable a la de un adulto, mientras que la visión binocular y la apreciación de la profundidad comienzan a desarrollarse en esta etapa.

Entre los meses 5 y 6 los ojos se mueven al simultáneamente y pueden seguir un objeto que cae hasta el suelo. Surge el reflejo de la acomodación del cristalino junto con la percepción de profundidad. Se mueve de manera coordinada la cabeza y los ojos para enfocar la atención en cualquier estímulo novedoso y se intenta alcanzarlo con las manos.

Finalmente, desde el mes 11 y hasta el 12 la estereopsis es cercana a la de un adulto y puede percibir con claridad en la zona de la fóvea.

En la edad adulta, el ser humano hace uso de diversas pistas visuales a través de las cuales es posible conocer la distancia a la cual están ubicados los objetos en nuestro entorno. Según (Goldstein (2010)) estas son obtenidas a partir de la información contenida en cada una de las retinas y a partir de la información del sistema motriz del globo ocular. Se clasificar según la fuente de información en oculo-motrices, monoculares y binoculares.

Un esquema de la clasificación de las pistas visuales se puede observar en la figura 2.1.

Las pistas oculo-motrices se basan en el grado de tensión que adquieren los músculos que mueven el globo ocular y los que controlan el cristalino cuando se enfoca un objeto. La primera pista oculo-motriz es la convergencia, la cual consiste en el movimiento de ambos ojos cuando se enfoca un objeto cercano, y la segunda es la acomodación del cristalino, es decir, la propiedad que tiene dicha estructura de engrosarse o alargarse para enfocar objetos a diferentes distancias.

Las pistas monoculares proveen información relevante para la percepción de la distancia a partir de una sola imagen. Estas pistas, a su vez se dividen en pistas pictóricas las cuales contienen información implícita acerca de la distancia a los objetos. Entre estas se encuentran:

- Tamaño relativo: Los objetos mas pequeños tienden a percibirse mas distantes.
- Brillantez: Los objetos mas brillantes parecen estar mas cercanos.
- Oclusión: Un objeto que superpone u ocluye a otro es percibido como mas cercano
- Perspectiva: Las líneas paralelas parecen converger a medida que se incrementa la distancia desde el observador

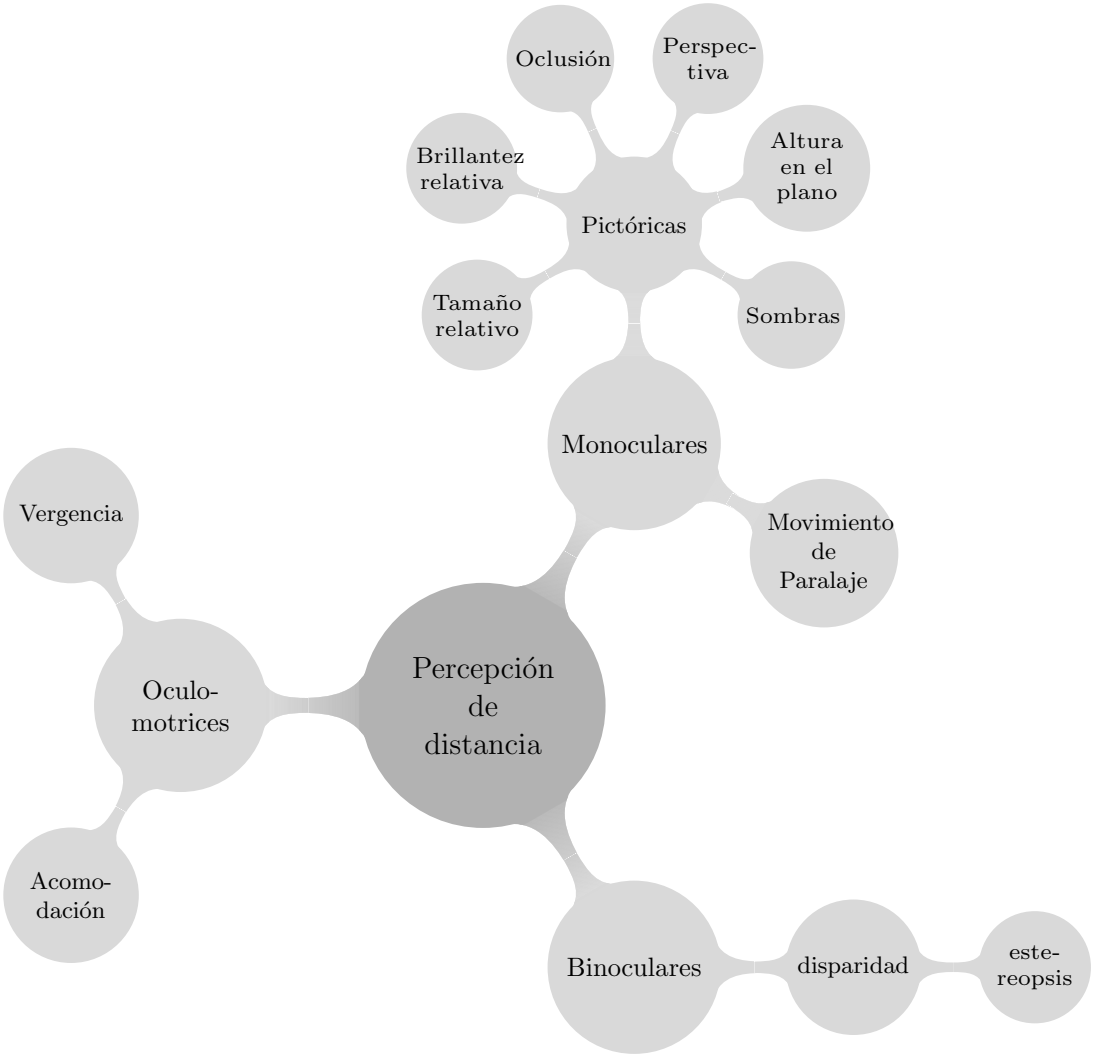


Figura 2.1: Pistas visuales para la percepción de distancia.

- Altura en el plano horizontal: Los objetos que se ubican por encima de otros parecen estar mas lejos que los que se ubican debajo
- Sombras: Las sombras que producen los objetos brindan información acerca de la ubicación de los mismos

Existe otra categoría, dentro de las pistas monoculares que se basan en el movimiento relativo entre el observador y la escena visual, la mas importante es el movimiento de paralaje, es decir, el flujo óptico en el cual el movimiento de cada punto depende de la distancia a la que se encuentra el objeto del observador (Wexler and Boxtel (2005)).

El ejemplo clásico del flujo óptico es el observar el paisaje mientras se viaja en coche o en tren. Los objetos mas distantes como las montañas y los parajes parecen moverse muy lentamente, mientras que las vías del tren o de la carretera adyacentes a nuestra ventanilla parecen desplazarse a gran velocidad.

Por otro lado, además de las pistas oculo-motrices y monoculares, existen las pistas concernientes a la información binocular. Esta es una de las fuentes principales para percibir la distancia y el espacio en una forma tridimensional. La principal razón es la separación existente entre nuestros ojos lo que ocasiona una diferencia entre las dos imágenes captadas por ambas retinas, permitiéndonos percibir el mundo desde dos perspectivas distintas.

En un la escena visual cuando un objeto es observado, sus representaciones retinales caen sobre la fóvea de cada una de las retinas. Los objetos que se encuentren ubicados a la misma distancia que el observado, tendrán puntos correspondientes en la misma región de ambas retinas, aunque ya no sobre la fóvea. Por otro lado, los objetos que se encuentren a una mayor o menor distancia que el observado, no tendrán puntos correspondientes en la retina.

Esta diferencia en los puntos correspondientes que se produce en ambas retinas, se conoce con el nombre de *disparidad binocular*, y es la responsable de originar el fenómeno denominado *estereopsis* es decir, la impresión de profundidad que se crea en el cerebro cuando una escena es observada (Goldstein (2010)).

(Julesz (1971)) fué quién demostró la existencia de la estereopsis al crear un estímulo visual denominado estereogramas de puntos aleatorios. Estos consistieron pares de imágenes casi idénticos, de puntos negros y blancos distribuidos de forma aleatoria, con la característica especial que en cada par de estas imágenes una pequeña región de forma cuadrada se encontraba desplazada ligeramente en una sola de las imágenes.

A través de estos estereogramas de puntos aleatorios, Julesz mostró que los observadores podían percibir la profundidad en este par de imágenes haciendo uso de la disparidad binocular como el único recurso para percibir la profundidad.

Una vez percibido y localizado un objeto en el espacio, es preciso reconocerlo sin que variaciones en la información sensorial acerca de su tamaño, forma, localización, brillantez, color, entre otras, afecten la capacidad para determinar que se trata de un único objeto. Esta capacidad se conoce como *constancia perceptual*.

Con respecto a la constancia en el tamaño de un objeto, si este es familiar para un observador, lo podrá reconocer sin importar la distancia relativa entre observador-objeto, a pesar de que el tamaño de la imagen retinal varíe según la ubicación de este último. Esta modificación de la percepción de la distancia para un objeto obedece al hecho de que se mantiene constante la percepción del tamaño de dicho objeto y por ende se denomina constancia de tamaño (Gregory (1966)).

Esta modificación en la percepción de distancia a la que se encuentra un objeto de tamaño conocido, ocasiona que frente a un aumento en la imagen retinal, este se perciba mas cercano. Por el contrario, una disminución en el tamaño de la imagen retinal ocasiona que el objeto se perciba mas lejano.

En el estudio de (Bower (1966)) con bebés de 2 meses de edad se mostró que estos tienen una capacidad innata para discriminar el tamaño de los objetos. En el experimento se condicionó a los bebés a responder a un estimulo visual que consistió de un cubo de 30 cms. y ubicado a una distancia de 1 mt. Una vez que se obtuvo la respuesta condicionada, el estímulo se modificó de acuerdo a tres situaciones diferentes:

- El mismo cubo de 30 cms. pero esta vez ubicado a 3 mts. de distancia (la imagen retinal producida era de un $\frac{1}{3}$ de la original).

- Un cubo de 90 cms. a 1 mt (la imagen retinal era 3 veces mas grande que la original)
- Un cubo de 90 cms. a 3 mts (la imagen retinal era igual que la original).

Después de registrar el número de veces que cada estímulo producía la respuesta condicionada, se encontró que el último estímulo, a pesar de producir la misma imagen retinal que el original, generó el mínimo número de respuestas condicionadas entre los cuatro estímulos, indicando que los bebés estaban respondiendo al tamaño real del cubo.

Posteriormente (Slater et al. (1990)) complementaron lo encontrado por (Bower (1966)) al mostrar que aunque los bebés responden al cambio en el tamaño de la imagen retinal ocasionado por un estímulo visual, si existe una etapa de aprendizaje y desarrollo de la percepción del tamaño real del objeto.

La constancia en la percepción de la forma de los objetos no altera su percepción a pesar de que estos sean presentados al observador desde diferentes orientaciones. Los bebés recién nacidos son capaces de reconocer un objeto independientemente de que este les sea presentado en diferentes orientaciones (Bower (1966); Slater (1989); Bremner (2003)), indicando así que esta capacidad es innata.

Las dos grandes teorías de la percepción visual se pueden clasificar en dos escuelas de pensamiento según la forma en que este proceso es guiado.

La percepción directa o ecológica considera que la percepción es guiada por la información captada por los sentidos fluyendo de abajo hacia arriba y cuyo principal exponente fué Gibson (Gibson (1979)), quien propuso que la función primaria de la percepción visual es ser el medio para facilitar las interacciones del individuo con su entorno, dejando de lado la actividad simbólica y representacionalista.

Por otro lado, la percepción indirecta o constructivista sostiene que la percepción es guiada conceptualmente fluyendo de arriba hacia abajo y cuyo propósito es crear una hipótesis que permita interpretar la información sensada a partir de la selección e inferencia de dicha información (Gregory (1966)).

La teoría ecológica (Gibson (1979)) toma como punto de partida el considerar la información de entrada como un patrón de luz en el tiempo y el espacio teniendo tres aspectos principales: patrones de flujo óptico, gradientes de textura y el concepto de "posibilitadores" .

Los patrones de flujo óptico proporcionan información sobre la dirección y velocidad del movimiento relativo entre el observador y el objeto. Los gradientes de textura brindan información clave acerca de la profundidad de los objetos sin necesidad de realizar alguna inferencia. Los posibilitadores describen los usos potenciales y directamente perceptibles de los objetos, es decir, constituyen sus significados comportamentales para el observador.

Para Gibson las propiedades que definen el concepto de espacio y por ende el de distancia, se encuentran en la interacción del agente con el ambiente. Gibson propone que la compresión del espacio está en los mecanismos que permiten la realización de acciones. Es dentro de esta concepción que se surgen los posibilitadores, ya que se define al espacio en función de las posibilidades que este ofrece para la acción. Por ejemplo una persona que se encuentra frente a un precipicio determinará si lo puede saltar no solo en función del ancho del mismo sino también en función de la velocidad que puede alcanzar antes del salto.

Uno de los aspectos que la teoría de la percepción directa no tiene en cuenta es el procesamiento que necesariamente se realiza a la información visual para detectar variables físicas contenidas en esta, como el reconocimiento e inferencia de la superficie en una imagen y la complejidad de este procesamiento.

Ambos teorías explican diferentes fenómenos relacionados con la percepción visual, pero ninguna abarca la totalidad de los mismos debido al hecho de que la teoría constructivista no considera la influencia del medio y la teoría ecológica no toma en cuenta que la percepción puede estar a su vez, influenciada por el conocimiento previo acerca del contexto en que se sitúa la escena visual.

Por otro lado aunque ambas teorías parecen ser contradictorias, es posible considerarlas interrelacionadas según la forma en que la percepción es guiada, ya que cada una de estas teorías concibe el flujo de procesamiento de la información durante la percepción desde dos puntos de vista complementarios en el que cada teoría compensa las deficiencias de la otra para producir una adecuada percepción de la escena visual (Goldstein (2008)).

Existen estudios muy interesantes que se enmarcan dentro de la teoría ecológica propuesta por (Gibson (1979)) y resaltan la necesidad de considerar a la percepción y a la acción como parte de una misma entidad, modelo o esquema.

(Lee and Lishman (1975)) utilizaron una habitación oscilante en la que podrían controlar la velocidad del flujo óptico percibido por los sujetos de prueba. A medida que oscilaba la habitación, los adultos pudieron mantener su postura realizando ligeras adaptaciones, mientras que los niños tendían a caerse. Esto les permitió proponer que la estimación del tiempo para hacer contacto con los objetos es proporcional a la relación entre el tamaño de la imagen retinal y tasa de expansión de dicha imagen.

Posteriormente, el trabajo de (Regan and Gray (2000)) analiza el tipo de información de la que se hace uso para evitar o lograr una colisión con un objeto. Se propone que los humanos nos basamos en el tiempo requerido para colisionar con un objeto. Para sostener esta hipótesis, se muestra evidencia empírica correspondiente con diferentes ecuaciones derivadas de forma teórica para la estimación de la dirección de un objeto en movimiento y el tiempo requerido para colisionar con este.

Esta correspondencia entre la evidencia empírica y las expresiones matemáticas están en función de la tasa de expansión de la imagen retinal y de la disparidad relativa, a semejanza de la relación propuesta por (Lee and Lishman (1975)) en su experimento de la variación den la velocidad del flujo óptico en una habitación oscilante.

(Gibson and Walk (1960)) en su experimento del precipio visual estudiaron la percepción de profundidad en los bebés a través de un dispositivo basado en una plataforma elevada del suelo a 1.20 mts. Esta plataforma estaba compuesta de dos secciones que descansaban sobre una lámina de acrílico transparente. En una de las secciones se colocó un patrón de ajedrez inmediatamente debajo de esta lámina, mientras que en la sección adjunta un patrón idéntico se ubicó sobre el suelo, dando la apariencia de un precipicio.

La conjetura a investigar en dicho experimento (Gibson and Walk (1960)) era si el bebé, frente al llamado de llamado de su madre, se desplazaría hacia la zona del precipicio aparente. Sin embargo, lo que se encontró fué que los bebés entre 6 y 14 meses no gateaban hacia esa zona, indicando así que a esta edad ya se contaba con una percepción de la profundidad.

En la fecha en que Gibson realizó su experimento del precipicio visual no se disponía de la misma información que hoy se tiene acerca del desarrollo de la percepción de profundidad en los bebés. Tal como se describió anteriormente, la estereopsis se empieza a desarrollar entre los 2 y 3 meses de edad, y entre los 5 y 6 meses ya se cuenta con una incipiente percepción de profundidad (Law et al. (2011)), por lo tanto era de esperarse que los bebés de prueba en el experimento de Gibson no se desplazaran hacia la zona del precipicio visual.

Capítulo 3

Visión artificial

En este capítulo se muestra el modelo geométrico mas reconocido en la visión artificial para modelar los parámetros de una cámara, este es el modelo pin-hole. Se describe el desarrollo de las expresiones matemáticas y de los fundamentos geométricos en los que se sustenta este modelo. Este procedimiento tiene como finalidad calcular una medida de distancia en unidades absolutas y en un sistema métrico a partir de la imagen provista por una cámara.

Este modelo se ha aplicado en agentes artificiales equipados con una o más cámaras. Para ilustrar esto, se revisan importantes trabajos en el área en los que se ha calculado la distancia a los objetos con el propósito de lograr una navegación segura en distintos tipos de entornos: estructurados, no estructurados, interiores y exteriores.

3.0.1. Modelo de cámara pin-hole

Tradicionalmente la estimación de la distancia en el área de robótica y visión artificial ha sido estudiada dentro del marco de la geometría y las matemáticas. El propósito fundamental es formular un modelo que a partir de las coordenadas en píxeles de un objeto representado en una imagen, permita obtener la distancia desde el centro de la cámara hasta dicho objeto.

El modelo geométrico de cámara mas conocido en visión artificial es el modelo pin-hole (Tsai (1987); Moons (1998)), este se basa en la realización de una transformación proyectiva en base a un sistema de coordenadas euclidiano.

La formulación de este modelo es simple y comienza con la figura 3.1. Esta figura muestra como un punto en el espacio se proyecta en el plano de una imagen de una cámara centrada en el origen del sistema de coordenadas tridimensionales.

Mediante una relación de triángulos semejantes, es posible expresar matemáticamente el mapeo de un punto en el espacio $\vec{X}$ con coordenadas $[X, Y, Z]^T$ a un punto de la imagen $\vec{x}$ con coordenadas $[x, y]^T$. Una vista descriptiva de este modelo se puede ver en la figura , a partir de la cual se puede deducir que $[x, y]^T = [fX/Z, fY/Z]^T$, f representa la distancia focal de la cámara, es decir, la distancia desde el plano de la imagen hasta el centro de la cámara.

Mediante la introducción de una constante ρ este mapeo puede ser expresado de forma matricial (Ver ecuación 3.1).

$$\rho \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.1}$$

Cuando se trabaja con imágenes digitales, es conveniente referenciar las coordenadas a un punto distinto al centro de la imagen, el cual generalmente se elige en la esquina superior izquierda. Además resulta útil, representar la distancia focal f en términos de unidades dadas en píxeles y considerar la inclinación de estos. Para tal efecto, al introducir en la matriz los denominados *parámetros*

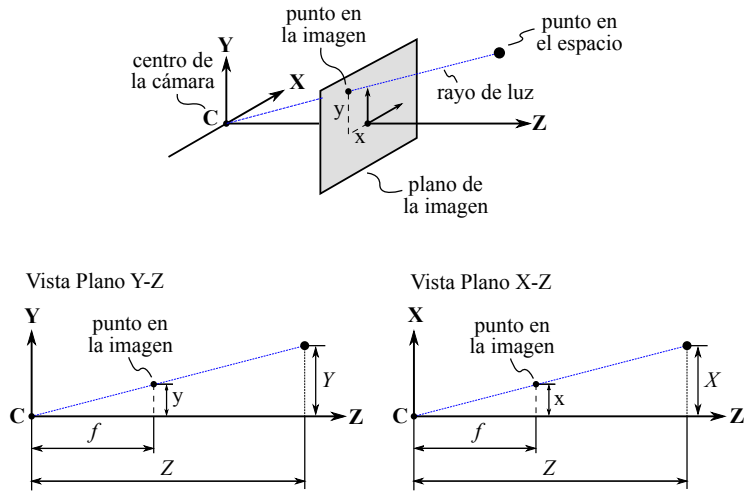


Figura 3.1: Geometría del modelo Pinhole

intrínsecos de cámara, $p_x, p_y, \alpha_x, \alpha_y, s$, se origina la *matriz de calibración* $[K]$, la cual es única y caracteriza a cada cámara. La ecuaciones 3.2 y 3.3 describen estas relaciones en forma abreviada y explícita, respectivamente.

$$\rho\vec{x} = [\ K\ ]\ \vec{X} \tag{3.2}$$

$$\rho \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.3}$$

Hay que tener presente que tanto la cámara como cada punto en el espacio están referenciados respecto a algún marco de referencia externo. Por lo tanto, se hace necesario conocer la posición y orientación de la cámara respecto a este marco. La matriz de rotación $[R]$ y el vector de traslación dado por las coordenadas $\vec{C} = [C_x, C_y, C_z]^T$ realizan respectivamente, la alineación y el traslado del centro de la cámara entre el marco de referencia exterior y el sistema de coordenadas de la cámara (Ver ecuación 3.4). Los parámetros contenidos en $[R]$ y $\vec{C}$ son conocidos como los *parámetros extrínsecos* de la cámara.

$$[\ \vec{X}\ ] = [\ R\ ] \left[\ \vec{X}_{ext}\ -\ \vec{C}_{ext}\right] \tag{3.4}$$

Un punto en el espacio $\vec{X}_{ext}$ referenciado según un marco de referencia externo y con coordenadas $[X_{ext}, Y_{ext}, Z_{ext}]^T$ puede ser trasladado y rotado a un punto $\vec{X}$ con referencia a la cámara y en seguida ser proyectado hacia la imagen para obtener las coordenadas $\vec{x} = [x, y]$ de su proyección. Al fusionar las ecuaciones 3.2 y 3.4, se puede desarrollar la expresión matricial que permita calcular esta transformación (Ver ecuación 3.5).

$$\rho\,[\vec{x}] = [\ K\ ][\ R\ ][\ \vec{X}_{ext}\ -\ \vec{C}_{ext}] \tag{3.5}$$

Al representar el vector $\vec{X}_{ext}$ mediante la introducción de coordenadas homogéneas, las cuales tienen en cuenta que los puntos en un mismo rayo de luz se proyectarán en las mismas coordenadas de la imagen, se pueden reacomodar los términos de la ecuación 3.5 para obtener una expresión general que describe formalmente el mapeo desde un punto en el espacio y referenciado según un marco de referencia externo a sus correspondientes coordenadas en píxeles en la imagen en la que se proyecta. La expresión resultante se puede observar en la ecuación 3.6.

$$\rho[\vec{x}] = [\ K\ ][\ R\ ][\ I\ |\ -\vec{C}_{ext}\ ][\vec{X}_{ext}] \tag{3.6}$$

La ecuación 3.6 puede escribirse de forma abreviada al denominar a una matriz $[\ P\ ]$ como el producto de todas las matrices del lado derecho de esta expresión, según se muestra en las ecuaciones 3.7 y 3.8. Esta matriz es conocida como la *matriz de proyección* y es la que finalmente contiene los parámetros intrínsecos y extrínsecos de la cámara para realizar este mapeo, de este modo, si se conocen todos ellos se puede decir que la cámara está calibrada.

$$[\ P\ ] = [\ K\ ][\ R\ ][\ I\ |\ -\vec{C}_{ext}\ ] \tag{3.7}$$

$$\rho\,[\vec{x}] = [\ P\ ][\vec{X}_{ext}] \tag{3.8}$$

3.0.2. Calibración de cámara

Inicialmente, no se dispone de los parámetros de la matriz de cámara $[\ P\ ]$, pero estos se pueden obtener a partir de un procedimiento de calibración. Este hace uso de un patrón tridimensional en el que se puedan determinar las coordenadas tridimensionales en un sistema métrico de por lo menos 6 puntos y que cada uno de estos puntos pueda ser ubicado en la imagen obtenida por la cámara por medio de coordenadas en píxeles .

Un patrón de calibración típico es como el que se muestra en la figura 3.2, el cual corresponde a una cámara web. Este patrón consiste de un arreglo tridimensional de dos cuadrículas a semejanza de un tablero de ajedrez y con 2 cms. de lado en cada cuadro. En este se pueden ver los 6 puntos elegidos para realizar el procedimiento de calibración los cuales están marcados en color rojo, y en color azul se indican 2 puntos elegidos que son utilizados para la prueba del procedimiento de calibración.

El número de puntos necesario para efectuar la calibración, 6 en este modelo, hace referencia al número de parámetros de la matriz $[\ P\ ]$. Cada punto en la imagen del patrón de calibración aporta dos coordenadas: x y y, con las cuales se pueden formular dos ecuaciones por cada punto, si se utilizan los 6 puntos, se obtiene un total de un sistema de 12 ecuaciones, las cuales corresponden al número de parámetros de la matriz $[\ P\ ]$.

Al desarrollar la expresión matricial indicada en la ecuación 3.8, se obtienen las ecuaciones 3.9. Al desarrollar de forma algebraica las estas expresiones, dividiendo las dos primeras ecuaciones

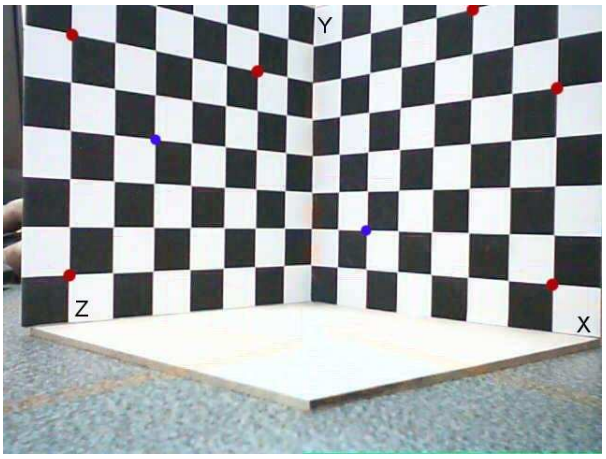


Figura 3.2: Patrón de calibración para el modelo Pinhole

entre la tercera y ordenando términos se obtienes las ecuaciones 3.10, las cuales corresponden a uno solo de los puntos de calibración.

$$\begin{aligned} \rho x &= P_{11}X_{ext} + P_{12}Y_{ext} + P_{13}Z_{ext} + P_{14} \\ \rho y &= P_{21}X_{ext} + P_{22}Y_{ext} + P_{23}Z_{ext} + P_{24} \\ \rho &= P_{31}X_{ext} + P_{32}Y_{ext} + P_{33}Z_{ext} + P_{34} \end{aligned} \tag{3.9a}$$

$$P_{11}X_{ext} + P_{12}Y_{ext} + P_{13}Z_{ext} + P_{14} - P_{31}X_{ext}x - P_{32}Y_{ext}x - P_{33}Z_{ext}x - P_{34}x = 0 \tag{3.10a}$$

$$P_{21}X_{ext} + P_{22}Y_{ext} + P_{23}Z_{ext} + P_{24} - P_{31}X_{ext}y - P_{32}Y_{ext}y - P_{33}Z_{ext}y - P_{34}y = 0 \tag{3.10b}$$

Si se reemplazan las coordenadas para cada uno de los 6 puntos elegidos para calibrar la cámara, se obtiene un sistema de 12 ecuaciones lineales que se puede resolver por medio de la descomposición en valores singulares y así obtener los parámetros de la matriz [P].

En este caso, se especifican las coordenadas en píxeles x, y de cada uno de los 6 puntos en la tabla 3.1. Los cuales permiten obtener los parámetros que componen a la matriz de cámara [P]. Estos se muestran en la tabla 3.2

Con los parámetros de la matriz de cámara [P] es posible realizar el mapeo desde las coordenadas tridimensionales de un punto en el espacio $\vec{X}$ hasta las coordenadas bidimensionales de la representación $\vec{x}$ del mismo punto en la imagen captada por la cámara.

3.0.3. Visión estéreo

Como analogía a la visión binocular, la visión artificial ha trabajado con un par de cámaras estéreo al ubicar dos cámaras sobre el mismo plano, separadas una cierta distancia y cuyos ejes ópticos son paralelos. Un par de cámaras estéreo se puede observar en la figura 3.3

Cuadro 3.1: Coordenadas de los puntos de calibración mostrados en la figura 3.2

X_{ext}	Y_{ext}	Z_{ext}	x	y
100	140	0	502	5
140	100	0	592	86
140	20	0	587	300
0	120	140	72	30
0	140	40	275	30
0	20	140	70	290

Cuadro 3.2: Parámetros obtenidos para la matriz de calibración [P]

0.001470	0.000031	-0.004368	0.723889
-0.000936	-0.004212	-0.000974	0.689883
-0.000004	-0.000000	-0.000004	0.002190

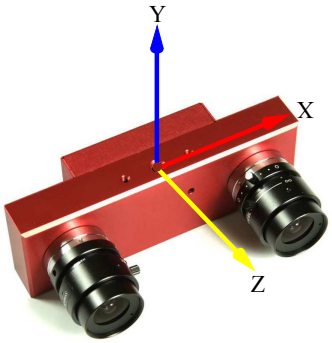


Figura 3.3: Par estéreo de cámaras

Con esta configuración es posible calcular la ubicación de cada uno de los objetos de la escena que se está observando mediante la relación de las matrices de proyección de ambas cámaras (Hartley and Zisserman (2004)).

Con el modelo de cámara pin-hole aplicado a cada una de las dos cámaras es posible permite realizar el mapeo inverso, es decir, a partir de las proyecciones $\vec{x}_i = [x, y]_i^T$ y $\vec{x}_i = [x, y]_d^T$ del mismo punto en la imagen izquierda y derecha y de las matrices de proyección $[P]_i$ y $[P]_d$ de ambas cámaras, es posible determinar las coordenadas tridimensionales $\vec{X} = [X_{ext}, Y_{ext}, Z_{ext}]^T$ del punto en cuestión.

En imágenes como la del patrón de calibración es relativamente sencillo encontrar puntos correspondientes. Sin embargo, en imágenes reales este proceso se torna mas complejo y es un problema fundamental en visión artificial conocido como el problema del emparejamiento estereoscópico (**?**).

Para afrontar con este problema de emparejamiento, se puede hacer uso de la geometría mediante la perspectiva desde dos vistas diferentes de la escena que se observa. Esta se conoce con el nombre de geometría epipolar según la cual un punto en una imagen define en el plano de la otra imagen una línea sobre la cual debe estar punto correspondiente (Hartley and Zisserman (2004)). Esta línea

se conoce con el nombre de línea epipolar.

La geometría epipolar es independiente de la estructura de la escena, y solo depende de los parámetros internos de las cámaras y de su posición relativa. En el caso de un par estéreo de cámaras como el mostrado en la figura 3.3 donde las cámaras son paralelas y están sobre el mismo plano, las líneas epipolares son horizontales. Esta característica implica que la búsqueda de puntos correspondientes se restringe a un búsqueda sobre una fila en la imagen.

La geometría epipolar es representada por una matriz $[F]$ de 3×3 denominada *matriz fundamental*. En el caso de que proyecciones de un punto en el espacio sean correspondientes en la imagen izquierda $\vec{x}_i$ y en la imagen derecha $\vec{x}_d$, el producto matricial entre estos y $[F]$ es igual a 0, ver ecuación 3.11. Esta condición se denomina restricción epipolar y permite determinar de forma mas eficiente y precisa la correspondencia de las proyecciones en ambas imágenes.

$$\vec{x}_d{}^T[\ F\ ]\vec{x}_i = 0 \tag{3.11}$$

Actualmente además de utilizar la geometría epipolar, se han propuesto diversos métodos para tratar con el problema del emparejamiento estereoscópico. Existen métodos y algoritmos que trabajan al buscar características relevantes, tales como bordes o segmentos de líneas, correlación entre regiones de píxeles, transformaciones de Fourier y métodos de minimización de energía (Alvarez et al. (2002)).

Un estudio comparativo acerca de la efectividad de los principales métodos de correspondencia se puede encontrar en la tesis de (Martínez (2010)). Los métodos mas reconocidos son los métodos basados en correlación entre regiones de píxeles y la correspondencia de características relevantes.

Las técnicas para determinar la correspondencia de puntos que se basan en las características utilizan principalmente los bordes de los objetos en las imágenes para realizar el emparejamiento en las imágenes del par estéreo. Entre las ventajas se destacan: una mayor tolerancia a diferencias de iluminación en la escena y un menor tiempo de cómputo. La principal desventaja de estos métodos es que la cantidad de correspondencias realizadas no es la suficiente para representar la profundidad de la escena de forma completa. Además dependen en gran medida que los objetos presentes tengan bordes claramente definidos.

Por otro lado, las técnicas que se basan en la correlación de píxeles utilizan ventanas de búsqueda para determinar el grado de similitud entre dos píxeles en ambas imágenes. El objetivo es encontrar el par de ventanas de mayor similaridad dado un criterio de similitud como la suma de diferencias absolutas, la suma de diferencias al cuadrado, entre otros. Una comparación del funcionamiento de estos métodos se puede encontrar en el trabajo de (Graffigna et al. (2005)).

La ventaja de las técnicas basadas en el área, es la producción de mayor cantidad de puntos correspondientes ya que la correspondencia se realiza para cada uno de los píxeles que componen ambas imágenes. Por otro lado, entre las desventajas se encuentran la sensibilidad a variaciones de iluminación en la escena, la distorsión en la correspondencia cuando la ventana de búsqueda encuentra el borde de un objeto o una discontinuidad, y la dependencia de la restricción epipolar para reducir el tiempo de cómputo, esta restricción es propia de la geometría de los sistemas de visión estéreo con cámaras calibradas y fijas.

En la figura 3.4 se puede observar un sistema simplificado de un par de cámaras estéreo. Se define como b la distancia entre los centros C_i y C_d de ambas cámaras, conocida como linea base y la distancia focal f de cada una de las cámaras.

Al disponer del resultado del aparejamiento a través del uso de alguno de los métodos mencionados anteriormente, se pueden determinar las coordenadas horizontales x_i y x_d en la imagen

izquierda I_i y derecha I_d, respectivamente, de las proyecciones resultantes de un punto $\vec{X}_{ext}$ en el espacio tridimensional.

La diferencia $d = x_i - x_d$ es lo que se conoce como la *disparidad* entre ambas imágenes. El valor de la disparidad d es inversamente proporcional al de la distancia Z_{ext}, es decir, que entre mayor sea la disparidad mas cercano está el objeto y viceversa.

Mediante la semejanza de los triángulos formados por los puntos $(C_i, \vec{X}_{ext}, C_d)$ y $(x_i, \vec{X}_{ext}, x_d)$ de la misma figura (3.4), se puede establecer la relación mostrada en la ecuación 3.12. Al despejar el valor de Z_{ext} se obtiene la expresión que nos permite obtener el valor de la profundidad para cada uno de los puntos correspondientes (ver ecuación 3.13).

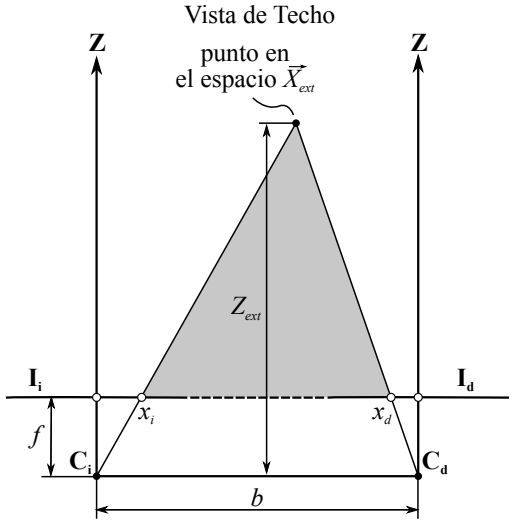


Figura 3.4: Cálculo de la disparidad

$$\frac{b}{Z_{ext}} = \frac{b - (x_i - x_d)}{Z - f} \tag{3.12}$$

$$Z_{ext} = \frac{bf}{d} \tag{3.13}$$

Recapitulando, la estimación de la distancia usando la visión estéreo (dos cámaras paralelas y separadas una cierta distancia) involucra tres pasos:

1. Establecer las correspondencias entre la imagen izquierda y derecha.
2. Proceder con el cálculo de la disparidad.
3. Determinar la distancia usando las relaciones geométricas del modelo de cámara

Diversos estudios han utilizado este modelo para resolver problemas relacionados con la navegación de robots.

En el artículo (Murray and Little (2000)) se calculó un mapa de disparidad (una imagen en escala de intensidades que indica el valor de la disparidad para cada par de puntos correspondientes) haciendo uso de un sistema de visión estéreo para construir una malla de ocupación del entorno y de esta manera realizar la planeación de rutas para una navegación segura.

En (Collins and Kornhauser (2006)) se detectaron obstáculos para la navegación de un vehículo autónomo a través de la búsqueda del mismo valor de disparidad para una columna de pixeles sobre un intervalo de filas contiguas, si este intervalo era mayor que un determinado umbral, los píxeles eran declarados como un obstáculo.

En (Saxena et al. (2007)) se implementó un algoritmo de aprendizaje haciendo uso de la información provista por un par de cámaras estéreo. Esto le permitió a un robot móvil obtener una estimación de profundidad y navegar a alta velocidad en un ambiente exterior.

En (Hasan et al. (2010)) se extrajo una región del mapa de disparidad para obtener la distancia a los obstáculos del entorno dentro de esa región. Enseguida se implementó un algoritmo sencillo para navegar en forma segura.

(DeSouza and Kak (2002)) realizaron una revisión extensa de los trabajos relacionados con la navegación de robots móviles basada en visión artificial. Se consideraron navegación en interiores y exteriores así como también en ambientes estructurados y no estructurados.

La conclusión principal a la que (DeSouza and Kak (2002)) llegaron fué que aunque los procedimientos utilizados son muy precisos y pueden llevar a un robot de un punto inicial a uno de destino, son susceptibles a ser afectados por iluminación variable y las condiciones propias del entorno. Sin embargo afirman, que si la meta final es llevar a cabo una navegación en la que los objetos y las propiedades del entorno adquieran un significado intrínseco para el robot, todavía hace falta mucho trabajo que realizar en este sentido.

La capacidad para realizar una estimación de la distancia constituye un ejemplo de lo anterior, la cual debería estar cimentada en las capacidades sensorimotrices del robot con un significado para este, en lugar del cálculo de una cantidad dimensional a partir del valor determinado para la disparidad, que de manera forzosa necesita ser interpretada por un operador o agente externo.

Esta interpretación externa surge a partir del modelo de cámara pin-hole y la geometría epipolar. Ya que este modelo y este marco de trabajo es independiente de la estructura de la escena visual al producir un valor numérico para la distancia a los objetos que no representa mas que una cantidad dimensional.

Según esta nueva postura para la nueva inteligencia artificial (Pfeifer and Scheier (1999)) el ambiente y su contenido son parte del fenómeno cognitivo, lo conforman y lo dirigen. En este caso, el valor numérico para la distancia a un objeto determinado por el modelo pin-hole y la geometría epipolar tendrá un significado diferente para agente artificial del doble del tamaño que otro, ya para la misma distancia el primero solo podría avanzar la mitad de pasos que el segundo.

Es a este tipo de conocimiento intrínseco y con un significado comportamental para el agente en sí, que esta tesis se decida a investigar. Específicamente, se estudia la adquisición de un conocimiento espacial cimentado para un agente artificial.

Capítulo 4

Robótica Cognitiva

4.1. Inteligencia artificial

Hasta principios de los años 90, los trabajos realizados en el área de la inteligencia artificial estuvieron soportados por la hipótesis de un sistema simbólico (Newell and Simon (1976)). La idea central postulaba que el símbolo, definido como las entidades por medio de las cuales el conocimiento es representado, constituía la raíz de toda acción inteligente.

De acuerdo a esta concepción, la concepción de la inteligencia fué influenciada por el procesamiento simbólico realizado por las computadoras de la época. Esto ocasionó que la visión de la inteligencia se limitara a ser únicamente a ser un procesador en forma secuencial de información abstracta. Donde la información fluía en una sola dirección, partiendo desde la recepción de una entrada, realizar un procesamiento para finalmente producir una salida o respuesta.

Esta forma de entender a la cognición predominó en la mayoría de los laboratorios de inteligencia artificial. Por lo tanto, en los agentes artificiales de la época, se implementaron algoritmos computacionales que procesaban la información de manera secuencial, recibiendo una entrada, realizando un proceso y expresando el resultado como una salida al entorno.

Esta concepción representacionalista y simbólica de la cognición fué criticada por filósofos como John Searle y su ejercicio mental del "cuarto chino"(Searle (1980)). En este, Searle describe a los sistemas computacionales como manipuladores de información abstracta, capaces de manejar sintaxis pero no semántica.

Para esto, recurre a una analogía que cuestiona la naturaleza de un sistema que da cuenta de capacidades que se puedan denominar inteligentes. Serla comienza describiendo una situación en la que supone la existencia de una persona que se encuentra en un cuarto asilado del mundo exterior, a excepción por la presencia de dos ranuras en una da las paredes del cuarto. La característica de esta persona imaginaria es que no entiende el idioma chino, y sin embargo tiene asignada la tarea de emitir una respuesta en este idioma frente a preguntas escrita que le son introducidas por una de las ranuras del cuarto.

Por supuesto que la tarea que Serle le propone a esta persona - la de responder a preguntas en un idioma que no entiende - no puede ser llevada a cabo. Para subsanar esto, situándonos nuevamente el ejercicio mental, Searle le permite a la persona que se encuentra dentro del cuarto asilado, disponer de un diccionario de reglas sintácticas del idioma chino. Por lo tanto, cuando la persona toma las preguntas introducidas a través de la ranura de entrada, esta es capaz de utilizar este diccionario para producir la combinación correcta de símbolos en el idioma chino y emitir una respuesta escrita por la ranura de salida.

El punto que Searle cuestiona si esta persona imaginaria - quien en el fondo representa la forma de operar de un sistema computacional - realmente entiende a un nivel semántico el proceso que lleva a cabo. Para una persona ubicada afuera de este cuarto imaginario y que entienda el idioma chino, podría decir que el sistema que emite las respuestas, realmente entiende chino, aunque en realidad no sea así.

Posteriormente, Stevan Harnad describió el problema de la cimentación de los símbolos (Harnad (1990)). Este trata acerca de como la información semántica puede llegar a adquirir un significado intrínseco a un sistema. En el caso de los agentes artificiales, este problema se traslada a cuestionar como conectarlos con su ambiente de tal forma que la información sensorimotriz así como también los comportamientos exhibidos, adquieran un significado intrínseco para el agente, evitando la necesidad de ser interpretados por un observador externo.

Mas tarde, el filósofo Daniel Dennett expuso el problema del marco de referencia. Este plantea que comprender la dinámica de cambio del entorno es fundamental y a la vez una tarea compleja, si se desea que un agente exhiba comportamientos congruentes con este. En el caso del mundo real, este problema obliga a que el agente siempre esté en sincronía con el medio, ya que siempre está en constante cambio (Dennett (1993)).

De forma paralela, el roboticista Rodney Brooks, entonces investigador del laboratorio de inteligencia artificial del instituto tecnológico de Massashusetts, publicó un artículo en el cual criticó la visión puramente representacionalista y simbólica de la cognición para el desarrollo de sistemas artificiales y propuso una nueva metodología en la que la interacción con el ambiente debería ser el primer factor a considerar (Brooks (1990)). En este trabajo Brooks propone que los organismos biológicos encuentran un perfecto balance entre sus capacidades y los recursos que el entorno les ofrece, logrando así una estrecha relación que les permite sobrevivir y adaptarse a su medio ambiente.

Posteriormente, Brooks diseñó y construyó cuatro robots móviles, en base a una arquitectura organizada en niveles jerárquicos, denominada arquitectura basada en actividades (Brooks (1991)). En cada uno de estos niveles se asentaban distintos comportamientos como *leer la información sensorial* en el nivel mas bajo, hasta *exploración del entorno* en el nivel mas alto. La idea central era que sin ningún tipo de representación central explícita, cada módulo de esta arquitectura tenía acceso al entorno y a su vez se conectaba con otros modúlos. El resultado fué que con esta arquitectura emergieron comportamientos complejos de una forma incremental. Aunque conciso pero muy importante, en este mismo trabajo Brooks propuso que el mejor modelo del mundo es el mundo mismo, y de forma congruente con esta afirmación, el ambiente de prueba para sus experimentes fué su propio laboratorio.

Estos estudios fueron decisivos e hicieron eco en la robótica al replantear la dirección de la investigación realizada hasta ese momento, hacia una nueva escuela de pensamiento denominada *cognición cimentada*. Esta disciplina agrupa las diferentes posturas que sostienen que los procesos cognitivos están íntimamente vinculados a la interacción del agente con el entorno y resaltan además, el papel que desempeña el cuerpo, los procesos motrices y el entorno mismo para la cognición.

La psicóloga Margaret Wilson propuso diferentes vistas o concepciones acerca de la cognición cimentada (Wilson (2002)). Entre ellas que se encuentran: *es situada, es presionada por el tiempo, utiliza el ambiente y lo considera como parte del sistema cognitivo, tiene como propósito la acción, y su carácter introspectivo es basado en el cuerpo.* [1]

La cognición es una actividad situada que tiene lugar en un entorno físico y un agente que haga parte de ella tiene que tener un cuerpo capaz de percibir y actuar sobre su entorno. Es presionada por el tiempo, ya que la dinámica de la interacción agente-entorno ocurre en un tiempo real con duración finita y por lo tanto, las respuestas motrices de un agente deberán ser correspondientes y estar sincronizadas frente a los cambios percibidos en el ambiente.

Una de las razones que nos ha permitido adaptarnos y sobrevivir en nuestro ambiente reside en nuestra capacidad para hacer uso de nuestro de los recursos que este nos ofrece para completar

[1] Traducción libre por el autor.

diversas tareas. De forma correspondiente, la cognición cimentada considera al agente y al entorno como un sistema unificado, en el cual la actividad cognitiva no reside únicamente en la mente del agente sino que se encuentra distribuida entre este y la situación experimentada.

La cognición para la acción se relaciona con la necesidad del surgimiento de comportamientos que promuevan la adaptación del agente a su entorno. Por lo tanto, la actividad cognitiva debe ser entendida a luz de su contribución para la aparición de dichos comportamientos.

El carácter introspectivo de la cognición hace referencia a procesos llevados a cabo por nuestro cerebro en los cuales recreamos o simulamos situaciones perceptuales y motoras. La parte esencial es que estas simulaciones ocurren en la ausencia de un estímulo externo y sin la necesidad de ejecutar una respuesta motriz o de experimentar las consecuencias sensoriales de forma explícita. El proceso conocido como imaginería mental (Kosslyn (1994)), constituye el ejemplo mas representativo de estos procesos de simulación interna.

A consecuencia de todas estas consideraciones, el término inteligencia comenzó a adquirir un nuevo significado dentro de las disciplinas que buscaban el desarrollo de robots completos y autónomos para interactuar, sobrevivir y adaptarse a un entorno en constante cambio (Pfeifer (1996)).

Pfeifer y Scheier definen una serie de principios básicos para el diseño de sistemas inteligentes, entre los que se encuentran: *agentes completos, procesos realizados en forma paralela e íntimamente acoplados, diseño económico, balance ecológico, principio del valor, redundancia y coordinación sensorimotriz* (Pfeifer and Scheier (1999)) [2].

Un agente completo es autónomo, autosuficiente, corporizado y situado. Implicando así que debería operar sin intervención humana, teniendo un cuerpo sujeto a las leyes físicas del ambiente y adquiriendo la información del entorno a través de su propio sistema sensorial. La realización de procesos paralelos establece que un comportamiento puede ser generado a partir de procesos básicos llevados a cabo de forma simultánea en el agente, tal como ocurre en la arquitectura basada en actividades propuesta por Brooks.

El diseño económico establece que los buenos diseños son aquellos que aprovechan la dinámica de la interacción con el entorno y las restricciones físicas que este impone, es decir, debe existir un balance ecológico que de cuenta de un equilibrio entre las condiciones del ambiente y las capacidades físicas del agente, tal como el proceso evolutivo lo determinó en los agentes naturales a lo largo de millones de años de evolución.

El principio del valor define la existencia de una medida intrínseca al agente que le informe acerca de la conveniencia de las consecuencias sensoriales de un determinado comportamiento para una interacción eficaz y coherente con su entorno.

La redundancia sostiene que la disposición de los sistemas sensoriales en un agente deberían permitir la existencia de cierto traslape y correlación de la información sensorial adquirida, con el objeto de lograr agentes artificiales robustos. La coordinación sensorimotriz es un aspecto fundamental ya que propone un cambio de paradigma al sostener que la información sensorial y motriz generada por el agente constituye dos procesos interdependientes y cuyo propósito fundamental es la estructuración de dicha información.

Algunos de estos principios de diseños reflejan la esencia de las características de la cognición propuestas por Wilson y descritas anteriormente. La cognición situada y presionada por el tiempo se relaciona directamente con el principio de un agente autónomo, corporizado y situado. Hacer uso del ambiente corresponde a lo propuesto en el principio de diseño económico y del balance ecológico, descritos como un acople coherente entre las capacidades físicas de un agente y las restricciones del ambiente.

[2] Traducción libre por el autor.

(Barsalou (2008)) revisa de forma detallada la cognición cimentada para mostrar que se basada en el papel que desempeñan la acción situada, los estados corporizados y las simulaciones internas, al considerar evidencia empírica en torno procesos como la percepción, el lenguaje y la memoria. De forma especial, resalta notablemente las simulaciones internas como instancia de este carácter introspectivo de la cognición y basadas en términos de las modalidades sensoriales y motrices del agente. El mejor caso conocido de este mecanismo es la *imaginería mental* (Kosslyn (1994)).

4.2. Implementaciones en Robots

Todos estos planteamientos son los que han llevado a la investigación en inteligencia artificial al uso de agentes artificiales autónomos como plataformas de prueba y experimentación, dando origen a una nueva metodología para el estudio de la cognición centrada en la modelación computacional (Pezzulo et al. (2012)).

(Lungarella et al. (2003)) mostraron numerosos casos de estudio en los que se implementaron sobre agentes artificiales modelos cognitivos de habilidades como: categorización, auto-exploración, auto-organización, interacción social y plasticidad neuronal. Estos trabajos permitieron observar la ventaja de aplicar la metodología del diseño sintético (Pfeifer and Scheier (1999)), en la cual se estudiaron diversas hipótesis y modelos provenientes de las ciencias cognitivas mientras que de manera simultánea se obtuvieron comportamientos complejos en agentes artificiales.

En un caso concreto, la categorización es una habilidad que desarrollamos los humanos a pesar de la gran variabilidad en la representación sensorial que origina un objeto. Este proceso es observado durante el desarrollo de comportamientos exploratorios en bebés (Rochat (1989)), en el cual la percepción y la acción están involucradas en la recolección de información de una forma activa y simultánea a través de las diferentes modalidades sensoriales, con el propósito de reconocer y categorizar a los objetos en función de los usos potenciales que estos pueden llegar a ofrecer para el agente que interactúa con ellos (Gibson (1950)).

En la robótica cognitiva, los trabajos de (Scheier and Lambrinos (1996)) y (Pfeifer and Scheier (1997)) muestran precisamente como esta habilidad de categorizar y reconocer objetos puede ser lograda en un agente artificial a través de un proceso de coordinación sensorimotriz. La contribución principal de estos trabajos consistió en mostrar como se reduce de manera significativa la complejidad del proceso de categorización al considerarla desde esta perspectiva, en lugar de intentar predefinir las categorías o de incluir un sensor específico para cada tipo de objeto.

El desarrollo del sistema visual también ha sido estudiado bajo esta metodología. En el trabajo de (Held and Hein (1963)) se mostró como el movimiento auto-inducido influye en el desarrollo del sistema visual. Ellos utilizaron una góndola en la que colgaron dos gatos con edades desde 2 semanas de nacidos. En un extremo, uno de los gatos podría hacer girar la góndola a libertad mientras que en el otro extremo, el otro gato solo estaba suspendido. Aunque ambos gatos recibieron el mismo patrón de estimulación visual, solo el gato que se movió libremente desarrolló un comportamiento normal frente a tareas visualmente guiadas como el parpadear frente a un objeto puesto frente a los ojos o el de la locomoción.

En el trabajo de (Suzuki et al. (2005)) se recreó el experimento de (Held and Hein (1963)) pero esta vez haciendo uso de robots y robótica evolutiva. Los autores encontraron, que un robot evolucionado bajo una condición en la que podía controlar libremente sus comandos motrices, se desempeñó mejor que cuatro robots restantes a los que les fué restringido el control de sus movimientos. Sugiriendo, al igual que Held & Hein, que la correlación entre el movimiento auto-inducido y la estimulación visual constituye un factor necesario para un adecuado desarrollo de

comportamientos visualmente guiados.

Por último, la interacción social entre humanos-robot y robot-robot es considerado un tema de suma importancia dentro de las ciencias cognitivas para comprender como surge el desarrollo de comportamientos sociales. En la revisión que realizan (Breazeal and Scassellati (2002)) se tratan aspectos relevantes para la interacción social con robots a través del estudio de dos cuestiones básicas: ¿Qué imitar? y ¿Cómo mapear la tarea percibida a los comandos motrices adecuados para replicarla?. Según Breazal y Scassellati, la metodología para dar respuesta a estas preguntas se ha originado a partir del estudio de la percepción del movimiento, de la atención visual y de la transformación del espacio sensorimotriz en agentes artificiales, con el interés de que estos, puedan servir de plataformas para estudiar y evaluar modelos del aprendizaje social en animales y humanos.

4.3. Imaginería mental

Las simulaciones internas, definidas como la recreación de situaciones perceptuales, motrices e introspectivas, se han propuesto como las funcionalidades que le confieren el carácter introspectivo a la cognición. Es precisamente la última de las características que (Wilson (2002)) propone.

El fenómeno denominado imaginería mental (Kosslyn (1994)) definido como una capacidad cognitiva básica cuya función principal es permitirnos generar predicciones específicas basadas en nuestra experiencia pasada y soportado por evidencia neurológica (Kosslyn et al. (2006)) constituye el ejemplo mas representativo del carácter introspectivo de la cognición.

Desde esta perspectiva, diversos estudios han mostrado que utilizamos las mismas estructuras cerebrales para recordar eventos pasados y para imaginar el futuro. Esto les permitió a (Daniel L. Schacter and Buckner (2007)) proponer la teoría del cerebro prospectivo, según la cual una de las funciones cruciales del cerebro es usar la información almacenada para imaginar, simular y predecir posibles eventos futuros.

(Driskell et al. (1994)) observó que la práctica mental, definida como el ejercicio introspectivo de simular la ejecución de una tarea, logró una mejoría en el desempeño de atletas. En el trabajo de (Smith et al. (2001)) se mostró que el desempeño a la hora de anotar un punto, fue mejor para un grupo jugadores novatos de hockey que incluyeron a la práctica mental durante su fase de entrenamiento, comparado con otro grupo de jugadores novatos que no incluyeron en su rutina sesiones de práctica mental.

En la resolución de problemas relacionados con el razonamiento mecánico como por ejemplo estimar la dirección de rotación de dos engranajes acoplados, la rotación de figuras geométricas de manera tridimensional, o el ejercicio de la torre de Londres, se ha mostrado a la simulación mental como una estrategia empleada para encontrar la solución Shepard and Metzler (1971); Hegarty (2004).

4.3.1. El Modelo Directo

Antes de definir un modelo directo es necesario traer a la mente la importancia de conocer las consecuencias sensoriales de nuestras acciones, desde diferentes perspectivas, ya que este aspecto tiene diferentes implicaciones que se encuentran interrelacionadas.

Desde la filosofía, (Kiverstein (2007)) aborda el tema de la conciencia en los agentes artificiales desde la perspectiva de la dinámica sensorimotriz. Según este enfoque, una actividad consciente es aquella exploración perceptual del entorno a través de un conocimiento sensorimotriz, es decir, aquellos esquemas en un agente que dan cuenta de un dominio de la dinámica que gobierna su

comportamiento. De acuerdo a esto, una vez experimentadas las consecuencias de ejecutar una acción, el modelo directo es un mecanismo que le confiere la capacidad al agente de generar posibles expectativas sensoriales para posteriormente compararlas con que se experimentaron. Kiverstein propone que para lograr subjetividad en un agente artificial es necesario tener la capacidad de notar una diferencia en esta comparación.

Por su parte (Goodman and Holland (2003)) proponen que los modelos internos, entre ellos el modelo directo, podrían ser utilizados por agentes artificiales o robots para procesar datos nuevos o incompletos, detectar perturbaciones en el ambiente, mejorar el control motriz y guiar la selección de acciones futuras. Estos características son esenciales para el desarrollo de agentes artificiales completos, en el sentido descrito por (Pfeifer (1996)) los cuales dan cuenta de comportamientos congruentes en su entorno.

Desde la fisiología y psicología, el conocer las consecuencias sensoriales de acciones propias es una pieza fundamental para la ejecución de movimientos coordinados. (Sporns and Edelman (1993)) propusieron, que en lugar de resolver ecuaciones de cinemática inversa, el proceso que el cerebro realiza es coordinar los movimientos de las extremidades a través de la selección entre un repertorio de esquemas sensorimotrices. Para tal efecto, se deberían cumplir con tres requisitos básicos:

- La existencia de un repertorio amplio y variado de esquemas sensorimotrices.
- Un impacto diferenciable en los efectos sobre el entorno de los movimientos ejecutados.
- La existencia de mecanismos en el sistema nervioso que provean las consecuencias sensoriales de dicha ejecución.

Los *modelos internos*, propuestos desde la teoría clásica de control (Jordan and Rumelhart (1992)) se han propuesto como esquemas computacionales que dan cuenta de esquemas sensorimotrices. Estos modelos se dividen en dos clases: el modelo directo y el modelo inverso.

El modelo directo es un modelo predictor que capaz de proveer las consecuencias sensoriales de llevar a cabo una acción, es decir, capaz de realizar una predicción sensorial en base a la situación actual y una acción motriz que se pretende ejecutar. Por otro lado, el modelo inverso es un modelo controlador, el cual produce el comando motriz necesario para llevar al sistema desde un estado sensorial dado a uno deseado.

Es inevitable observar que una entidad abstracta y computacional como el modelo directo, por su carácter predictor, refleje el aspecto funcional de los mecanismos que según (Sporns and Edelman (1993)) deberían existir en el sistema nervioso central para proveer de forma anticipada las consecuencias sensoriales de las acciones.

Poniendo los elementos sobre la mesa, ya se consideró la importancia de disponer de predicciones sensoriales desde la filosofía (Kiverstein (2007)), desde la fisiología (Sporns and Edelman (1993)) y desde la robótica cognitiva (Goodman and Holland (2003)). La parte que ahora hace falta, es la evidencia neurológica de que el sistema nervioso central disponga de unos mecanismos que reflejen la funcionalidad de los modelos internos.

(Jeannerod (1995)) desarrolló una teoría en base a estudios neurológicos para mostrar que el sistema motriz es parte de una red de simulación en el sistema nervioso central humano. A su vez, esta red neurológica interviene en procesos como la imaginación de acciones, el reconocimiento de herramientas, el aprendizaje por observación y el reconocimiento de las acciones del otro.

Posteriormente, (Wolpert et al. (1998)) mostró como el cerebelo interviene en la realización de predicciones sensoriales, indicando que este es el órgano idóneo si se quisieran situar circuitos neuronales que den cuenta de los aspectos funcionales que poseen los modelos internos.

Los trabajos de (Wolpert et al. (1995); Wolpert and Flanagan (2001)) y (Grush (2004)) describen procesos que son llevados a cabo por el sistema nervioso central y que pueden ser explicados por un esquema abstracto como el de los modelos internos, entre estos se encuentran:

- La predicción de estados sensoriales.
- La compensación por el retraso en las señales sensoriales.
- La cancelación de los efectos sensoriales debidos al movimiento propio.
- La transformación de un error en coordenadas sensoriales a coordenadas motrices.
- La evaluación y desarrollo de planes motrices.
- La imaginería mental.

Las señales eléctricas del SN requieren de cierto tiempo para la adecuada transducción y transporte desde los órganos sensoriales hasta el cerebro (Wolpert et al. (2003); Miall and Wolpert (1996)). Este retraso que llega a ser del orden de cientos de milisegundos dependiendo de la señal puede producir un comportamiento motriz incongruente. Una alternativa para compensar esto es la estimación de estados sensoriales, ya que estas estimaciones presentan un menor tiempo para estar disponibles que el requerido para que la señal sensorial viaje desde los órganos sensoriales hasta el cerebro.

Las predicciones sensoriales pueden ser usadas para filtrar la información que contenida en la señal real que es captada por los sentidos, al resaltar la información crítica que se contenga. Esta información puede provenir de dos fuentes: aferente, debida a un cambio en el entorno, o re-aferente cuando es producto de un movimiento propio y que modifica la percepción sensorial. Esto explica el hecho de experimentar en menor intensidad un estímulo táctil cuando es auto-generado, ya que si se conocen de antemano los efectos de una acción propia, estos se atenuarán cuando se experimenten. Por ejemplo, el hecho de no experimentar en igual grado, la sensación que producen las cosquillas cuando uno mismo es el causante de ellas (Blakemore et al. (1999)).

El hecho de disponer de expectativas o predicciones sensoriales, nos da la facultad de comparar y evaluar las situaciones reales se experimentan en cada instante. El punto es que, al emparejar la experiencia sensorial real con las predicciones sensoriales generalmente coinciden, es decir, conocemos que los cambios en nuestro entorno se deben a nuestras propias acciones. En dado caso, que no lleguen a coincidir, se puede afirmar que el cambio en el ambiente no es debido a nuestra acción sobre este. (Frith et al. (2000)) ha propuesto que una falla en este mecanismo de comparación podría ser la causante de la esquizofrenia.

El hecho de generar las respuestas motrices adecuadas o congruentes con nuestro entorno puede estar basado en el uso de un amplio espectro de esquemas sensorimotrices que son seleccionados de acuerdo al contexto donde se da la interacción. Una propuesta desde la perspectiva computacional que puede dar cuenta de esto es la arquitectura planteada por (Wolpert and Kawato (1998)). En esta arquitectura se toma como unidad básica el acople entre un modelo directo y uno inverso para formar un par predictor-controlador.

Al interactuar con un ambiente dinámico y un cuerpo en constante cambio, hace necesario recurrir a modelos que sean flexibles y que se puedan actualizar a través de la experiencia. Por lo tanto, extrapolando esta consideración al modelo directo, este no es una entidad estática sino que debe poder ser refinado al comparar el error entre las predicciones realizadas y las consecuencias reales de llevar a cabo una acción específica (Wolpert et al. (2001)).

Bajo estas consideraciones, el modelo directo constituye una abstracción de aspectos funcionales del sistema nervioso central involucrados en la realización de predicciones sensoriales. Este carácter predictor es el que ha permitido considerar al modelo directo como uno de los mecanismos esenciales sobre el cual se basa el desarrollo sensorimotriz humano (Wolpert et al. (1995)), constituir una instancia del último requisito para la selección y evaluación de planes motrices que propusieron Sporns y Edelman (Sporns and Edelman (1993)).

Formalmente se puede definir al modelo directo, como un esquema que recibe como entrada una situación sensorial a un tiempo t (S_t), y un comando motriz a un tiempo t (M_t) y proporciona como salida la situación sensorial resultante al tiempo t_{t+1} (S^*_{t+1}). Vale la pena aclarar que el comando motriz que se usa es una copia eferente, es decir, es un comando motriz que no necesariamente se llega a ejecutar. De forma esquemática, el modelo directo se puede representar de acuerdo a la figura 4.1

Figura 4.1: Esquema del modelo directo.

Dentro de las implementaciones de estos modelos en robots, en la literatura se encuentra que se han usado para navegación segura en un agente autónomo artificial (Escobar et al. (2012)), implementación de sistemas de neuronas espejo (Arceo et al. (2013)).

(Hoffmann and Möller (2004)) codificaron un modelo directo en un agente artificial equipado con una cámara omnidireccional para tomar imágenes de 360^o del entorno, el cual consistió de obstáculos dispuestos en forma circular y alrededor del agente. A través del procesamiento de las imágenes obtenidas, se extrajo la distancia en píxeles para cada uno de los 10 obstáculos.

Cada una de estas distancia fueron introducidas al modelo directo para formar la entrada S_t junto con las velocidades de la rueda izquierda y derecha correspondientes al comando motriz M_t. La salida S_{t+1} estuvo representada por las distancia en píxeles resultantes si el comando motriz fuera ejecutado.

Al formar una red de modelos directos mediante el encadenamiento en serie de la salida de uno para proveer la entrada del siguiente, le permitió a Hoffmann realizar un proceso de simulación para dotar al agente con la capacidad de conocer su posición relativa al centro del círculo de obstáculos.

Posteriormente, (Hoffmann (2007)) implementó un modelo directo para realizar predicciones de imágenes provenientes de la cámara omnidirecional de un agente artificial en un entorno de obstáculos dispuestos alrededor de este (similar a su anterior trabajo (Hoffmann and Möller (2004)).

Con el modelo directo implementado y las imágenes predichas resultantes, Hoffmann pudo establecer un criterio para dotar a un agente con una noción de distancia, basado en el número de píxeles que representaban a los obstáculos. Una vez cimentada esta capacidad, y en base a un proceso de simulación interna, se implementaron algoritmos basados en reglas sencillas para encontrar a la salida a un laberinto de obstáculos o por el contrario determinar que era un laberinto cerrado.

En su trabajo (Möller and Schenck (2008)) diseñaron e implementaron un esquema computacional basado en el uso de 3 modelos directos en un agente artificial en un ambiente simulado. Cada modelo directo codificó un tipo de movimiento diferente: adelante, derecha o izquierda. La entrada sensorial consistió de las posiciones del centro de masa de los obstáculos del entorno, mientras que

la salida las posiciones predichas. Con esta implementación y en base a un proceso de simulación sensorimotriz, a través de una cadena de modelos directos, Möller logró obtener un agente en un ambiente de simulación capaz de identificar si un laberinto de obstáculos conformaba un camino cerrado o abierto.

Capítulo 5

Asociación multi-modal a través de un proceso de imaginería mental

En este capítulo se describe el agente artificial que se usa en toda la tesis. Posteriormente se retoma el modelo directo implementado sobre dicho agente y propuesto el trabajo de (Escobar et al. (2012)) con el objetivo de lograr una asociación multi-modal a través de un proceso de imaginería mental. En seguida se presenta el modelo computacional para lograr dicha asociación. En la sección final se muestran y discuten los experimentos que dan cuenta de esta asociación.

5.1. Agente artificial

Los experimentos reportados en esta tesis fueron implementados en un robot móvil Pioneer 3D-X (figura 5.1). Este agente artificial tiene dos ruedas laterales y una rueda que gira libremente ubicada en la parte posterior para darle estabilidad mientras se desplaza. Cada una de estas ruedas puede ser controlada de forma independiente en dirección hacia adelante o hacia atrás, otorgándole la posibilidad de desplazarse en cualquier dirección.

El agente cuenta con un arreglo de sonares dispuestos en forma de anillo capaces de reportar la distancia a los objetos en un rango de 0 a 5000 mm. Las medidas provistas por cada uno de los sonares pueden ser utilizadas para simular un sensor de parachoques en el agente, es decir, si se fija un umbral de distancia mínima, se puede considerar que el agente colisionó con un obstáculo cuando la distancia agente-obstáculo es un valor menor a este umbral.

El sistema de visión está compuesto por una cámara estéreo que provee dos imágenes a color de 320×240 píxeles. En la figura 5.2 se observa un ejemplo de estas imágenes correspondientes a un entorno de obstáculos de color rojo. Este sistema cuenta principalmente, con una librería de funciones para el procesamiento de imágenes entre las que se incluyen: calibrar ambas cámaras y calcular un mapa de disparidad el cual es una imagen de intensidades o niveles de gris, que contiene de forma explícita la información de la distancia a los objetos del entorno en base al valor de intensidad de cada píxel.

5.2. El modelo directo como mecanismo para predicciones sensori-motrices

Como se ha mencionado anteriormente, el modelo directo ha sido caracterizado como un mecanismo básico que puede dar cuenta de un proceso de simulaciones internas (Grush (2004); Wolpert et al. (2001)).

En un trabajo previo (Escobar et al. (2012)) reportaron la implementación de un modelo directo sobre el agente artificial de la figura 5.1, capaz de predecir la posición de los obstáculos en su am-

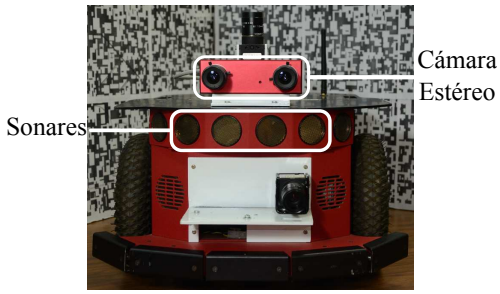


Figura 5.1: Agente artificial: Robot Pioneer 3-DX con un arreglo de sonares y una cámara estéreo en el frente.

(a) Imagen izquierda (320 × 240)

(b) Imagen derecha (320 × 240)

Figura 5.2: Imágenes obtenidas del sistema estéreo.

biente. Esto fué posible gracias a la re-enactuación de ciclos visuo-motrices para detectar colisiones a partir de la información visual y táctil.

A partir de la diferencia en la imagen izquierda y derecha provenientes de la cámara estéreo, se calculó un mapa de disparidad para extraer una región de interés (ROI) de 228 × 6 localizada en la fila de píxeles 152 de la imagen. Este valor corresponde a una distancia máxima de detección de obstáculos de 2.15 m.

Una vez obtenida esta ROI, se seleccionó para cada una de 228 columnas de píxeles, el mayor valor de disparidad entre los 6 valores disponibles. El resultado de este procedimiento fué un *vector de máxima disparidad* (VMD) el cual representa a los obstáculos mas cercanos en el campo visual de la cámara y su vez constituye la modalidad visual. El proceso para obtener el VMD se puede observar en la figura 5.3

La modalidad táctil está formada un *vector táctil* (VT) de 228 valores con el objeto de mantener una correspondencia dimensional con el VMD. Este vector codifica una señal binaria de un solo parachoques que se activa cuando se presenta un estado de colisión, sin importar la zona con respecto al cuerpo del agente donde esta ocurrió.

El objetivo es relacionar la información del VMD con un estado de colisión reportado por el VT. Este vector depende de los valores reportados únicamente por los 2 sonares frontales del robot, ya que los comandos motrices del agente, se restringieron a movimientos hacia adelante de 15 cm.

La representación esquemática del modelo directo implementado se puede ver en la figura 5.4. Este recibe como entrada el VMD_t de 228 valores y el comando motriz M_t correspondientes al tiempo t. La salida está formada por las predicciones VMD^*_{t+1} y el VT^*_{t+1} correspondientes esta vez, al tiempo $t+1$ para los estados visual y táctil respectivamente.

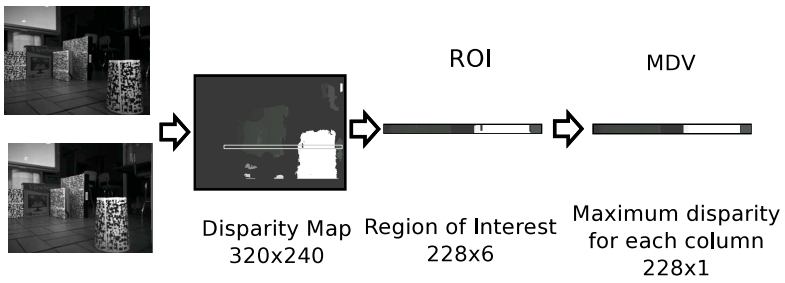


Figura 5.3: Procesamiento visual para obtener VMD. La descripción detallada del sistema se encuentra en (Escobar et al. (2012))

Figura 5.4: Modelo directo implementado. Se muestran la información de entrada y las predicciones de salida. Al tiempo t la entrada sensorial visual junto con el comando motriz permiten realizar una predicción sensorial de las modalidades visuales y táctiles al tiempo $t + 1$.

La codificación del modelo se realizó en base a un sistema de redes neuronales que actúan como predictores locales en un forma de imitar la distribución y el procesamiento efectuado por las diversas clases de células presentes en la retina (Kolb (2003)).

Cada uno de estos predictores locales toma como entrada una ventana de 14 valores del VMD_t y predice los 4 valores centrales para el VMD^*_{t+1} y para el VT^*_t. Esto origina que sea necesario un sistema de 57 redes neuronales tipo perceptrón multicapa. Para el entrenamiento se eligió el algoritmo de la retro-propagación fuerte del error (*resilient back-propagation*) (Riedmiller and Braun (1993)).

En la implementación computacional del modelo directo no se codificó de forma explícita el comando motriz, ya que este se encuentra codificado en la estructura misma de los datos sensoriales. La predicción sensorial para el tiempo $t + 1$ es una consecuencia directa de la situación sensorial y la acción a ejecutar al tiempo t y dado que en este caso la acción siempre es constante (movimientos hacia adelante), la predicción solo dependerá de la situación sensorial anterior.

El sistema de redes neuronales fué entrenado con datos recolectados mientras el agente se desplazaba hacia adelante en un ambiente rodeado de obstáculos. A cada uno de estos obstáculos se les añadió un patrón texturizado para que el mapa disparidad pudiera ser calculado mas fácilmente. Cada patrón de datos consistió en el mapa de disparidad antes y después de llevar a cabo el movimiento. El número total de patrones recolectados fué de 2259 dejando el 20 % para prueba.

5.3. Simulaciones internas corporizadas

En este trabajo se refiere a proceso de simulación interna como la retroalimentación de la salida al modelo directo, en donde las predicciones sensoriales al tiempo $t+1$ sirven como entrada al modelo

para producir las predicciones al tiempo $t + 2$ y así sucesivamente. Este proceso se denomina una *predicción de largo plazo* (PLP).

Una PLP es un proceso de simulación interna que representa las consecuencias sensoriales de llevar a cabo una serie de comandos motrices, en este caso 5 movimientos hacia adelante por el agente. En otra palabras una PLP indica "que pasaría si" el agente ejecutara dicha serie de comandos motrices.

Respecto a las predicciones para la modalidad visual, la figura 5.5 muestra la comparación entre el VMD de 5 movimientos hacia enfrente y la PLP tomando como situación inicial el vector de la parte superior de ambas imágenes. Regiones obscuras indican la ausencia de obstáculos mientras que regiones brillantes codifican a un obstáculo cercano al agente. Se aprecia como el agente se va aproximando a dos obstáculos sobre su lado derecho y como los VMD^* predichos se asemejan a los VMD reales.

(a) Salida esperada. (b) Predicción de largo plazo (PLP).

Figura 5.5: Comparación entre el VMD real y la predicción de largo plazo (PLP) del VMD^* en un horizonte de 5 instantes. Figura (a) La evolución del VMD real cada vez que el robot se mueve hacia adelante comenzando con el vector de la parte superior de la imagen. Figura (b) La predicción para el VMD^* para cada uno de los 5 instantes.

Por otro lado, las predicciones para la modalidad táctil, representadas por el VT^* presentan características especiales. Es importante notar que la codificación de los datos de entrenamiento para esta modalidad fué del tipo binaria, es decir, se codificó una señal con un valor de 1 o 0, para indicar un estado de colisión o no colisión respectivamente. Sin embargo, durante la ejecución de la PLP, se observó que la predicción táctil adquirió valores continuos en un rango de $[0, 1]$ y que los valores con mayor activación correspondían topológicamente a la posición de los obstáculos en la predicción VMD^*.

En la figura 5.6 se muestran las predicciones para el estado táctil para la misma PLP de las predicciones para el estado visual mostradas en la 5.5. Se puede observar como en el instante $t+5$ se produce la mayor activación y en la zona del lado derecho del VT^* indicando una correspondencia entre ambas modalidades sensoriales.

5.4. Asociación multi-modal en la imaginería mental

5.4.1. Aprendizaje de carácter introspectivo

El modelo directo implementado constituye un mecanismo que le provee al agente de forma anticipada con las consecuencias visuales y táctiles de moverse hacia adelante en pasos de 15 cm. Pero el agente aún no hace uso de estas capacidades motrices para navegar en el entorno, ya que carece de una noción de distancia o colisión que esté basada en la asociación de información visual, táctil y motriz.

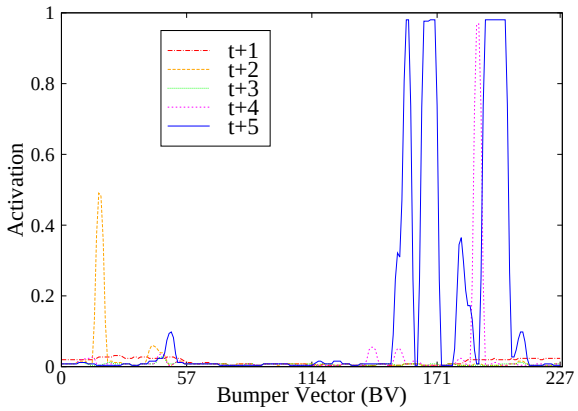


Figura 5.6: PLP de 5 instantes para el estado táctil.

En esta tesis se propone que a través de un mecanismo de asociación multi-modal llevado a cabo de manera instrospectiva y basado en el modelo directo, que pueda proveer al agente con una estrategia de control motriz de carácter anticipatorio. Esto significa, que en lugar de evitar un obstáculo en base a una simple reacción frente al estado táctil predicho, el agente sea capaz de realizar esta misma tarea en base al estado visual actual.

Para lograr una estrategia de control corporizada se hace uso de una técnica de aprendizaje fuera de línea (*Off-line*), basada en la asociación de las modalidades sensoriales visual y táctil durante la ejecución de una PLP. Se hace uso de una configuración de una red neuronal artificial en la que los pesos sinápticos sean sujetos a un *proceso de aprendizaje Hebbiano* (Hebb (1949)).

Lo que distingue a esta propuesta es que en contraste con la literatura relacionada con esta técnica de aprendizaje donde se ha utilizado durante la ejecución explicita de movimientos (Verschure et al. (1992); Salomon (1998); Wang et al. (2009)), aquí se propone implementarla de tal forma que la actualización en el valor de los pesos sinápticos de la red ocurra en el marco de un proceso de simulaciones internas.

Este proceso de aprendizaje puede ser visto como una *asociación multi-modal fuera de línea* que además de recrear situaciones perceptuales y motrices, estructura la información contenida en estas modalidades con el objetivo de decantar en una representación con un significado corporizado e intrínseco para el agente. Este proceso se asemeja a las características funcionales de la imaginería mental, nombrada a su vez como el ejemplo mas representativo del carácter introspectivo de la cognición cimentada (Barsalou (2008)).

5.4.2. Formalización del modelo computacional

5.5. Prueba del modelo a través de una tarea de navegación

5.6. Robustez del modelo

A la par que se realizan las predicciones sensori-motrices, se va lleva a cabo un proceso de aprendizaje artificial tipo Hebbiano (Hebb (1949)). Este proceso logra decantar en una matriz de

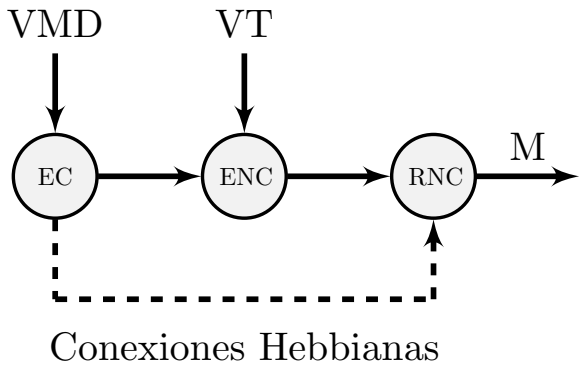


Figura 5.7: Conceptualización del mecanismo de asociación propuesto.

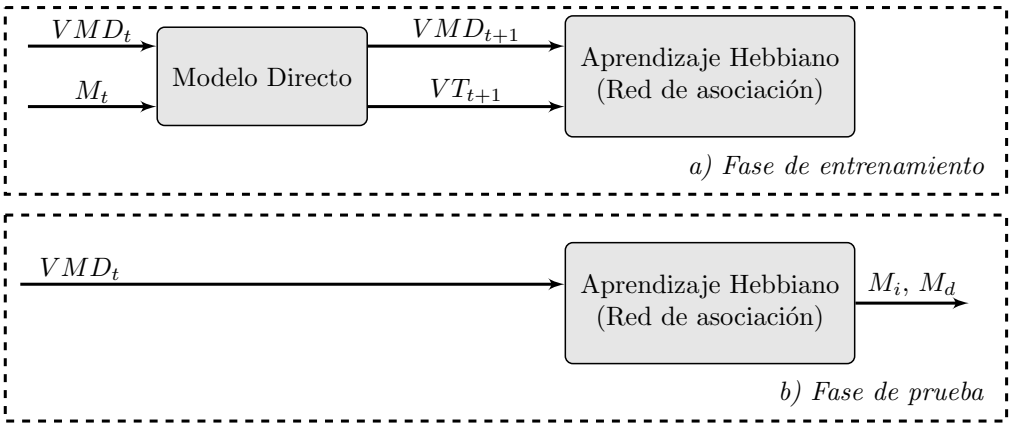


Figura 5.8: Esquema de la arquitectura computacional dividida en las fases de: (a) aprendizaje Off-line y (b) prueba del sistema.

pesos sinápticos la asociación entre estas las modalidades visual y táctil. Inicialmente, los pesos de esta matriz fueron inicializados con valores aleatorios, pero a medida que las simulaciones internas se llevaban a cabo comenzó a emerger una estructura en esta matriz con la forma de una diagonal.

Una vez que se obtuvo una diagonal definida en esta matriz, se detuvo la realización de las predicciones sensori-motrices. La forma final de esta matriz es una prueba clara de la factibilidad de lograr una estructuración en la información sensorial de entrada correspondiente los datos de las modalidades visual y táctil.

Cabe resaltar que este procedimiento fué llevado a cabo en una forma introspectiva, es decir, se obtuvo una asociación multi-modal a través de un proceso de imaginería mental mediante el uso de un modelo directo. Esto abre la puerta hacia un nuevo paradigma en el que el hecho de disponer de predicciones sensori-motrices puede ser aprovechado para el estudio de la imaginería mental en la cognición cimentada a través de la modelación computacional, tal metodología se ha descrito en el trabajo de (Pezzulo et al. (2012)).

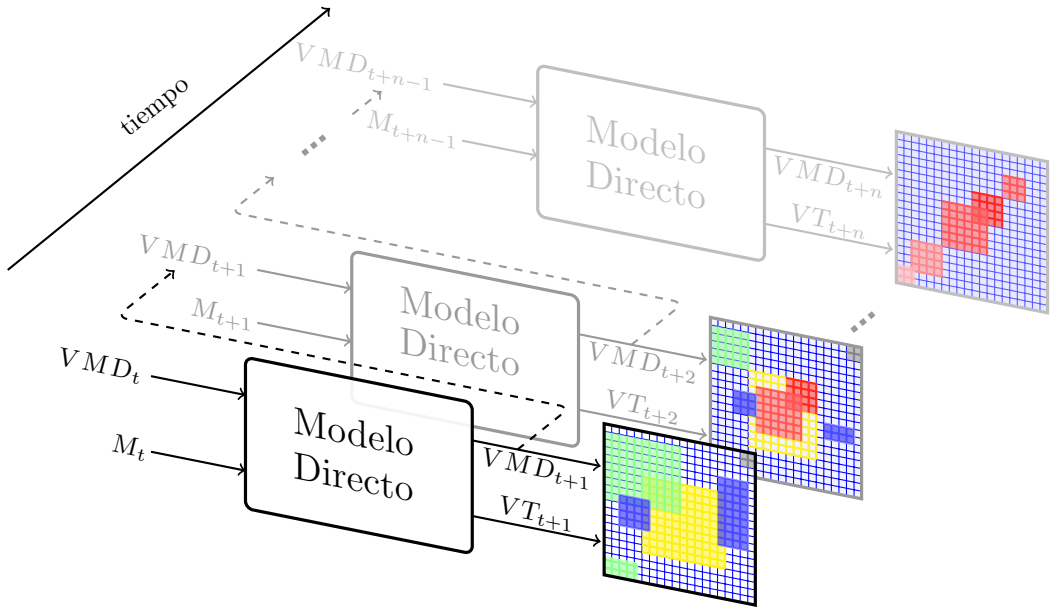


Figura 5.9: Proceso de asociación durante el aprendizaje Off-line. La salida del modelo directo es retro-alimentada como una entrada en lo que se denomina una predicción de largo plazo (PLP). En cada ciclo de la PLP, los pesos en la red de asociación (cuadrados de colores) son modificados de acuerdo al aprendizaje hebbiano.

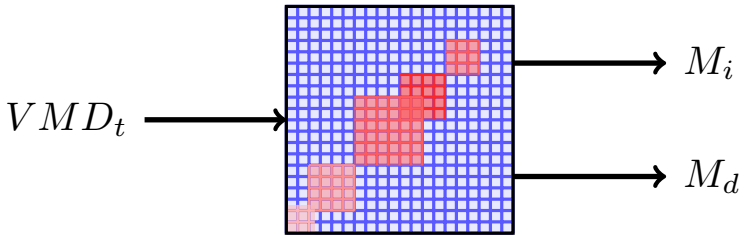


Figura 5.10: Matriz de asociación una vez efectuado el proceso de aprendizaje hebbiano. A través de esta matriz, el sistema es controlado únicamente por la entrada visual. M_i and M_d son las activaciones motrices de los motores izquierdo y derecho, respectivamente.

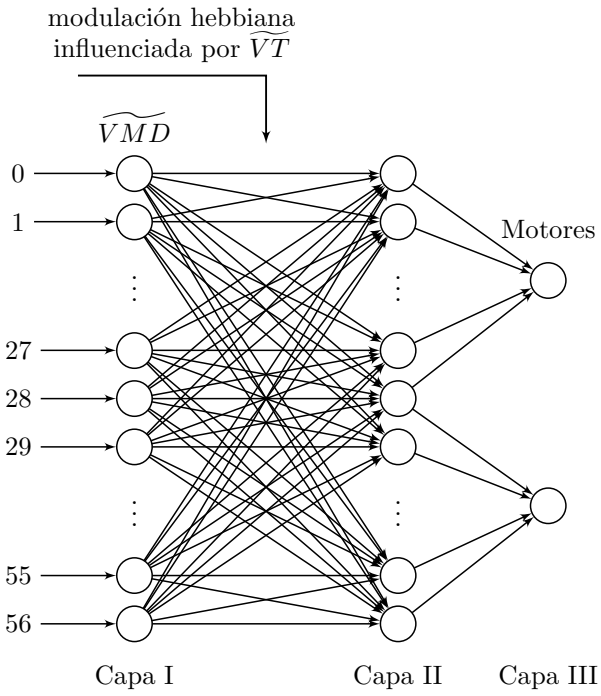
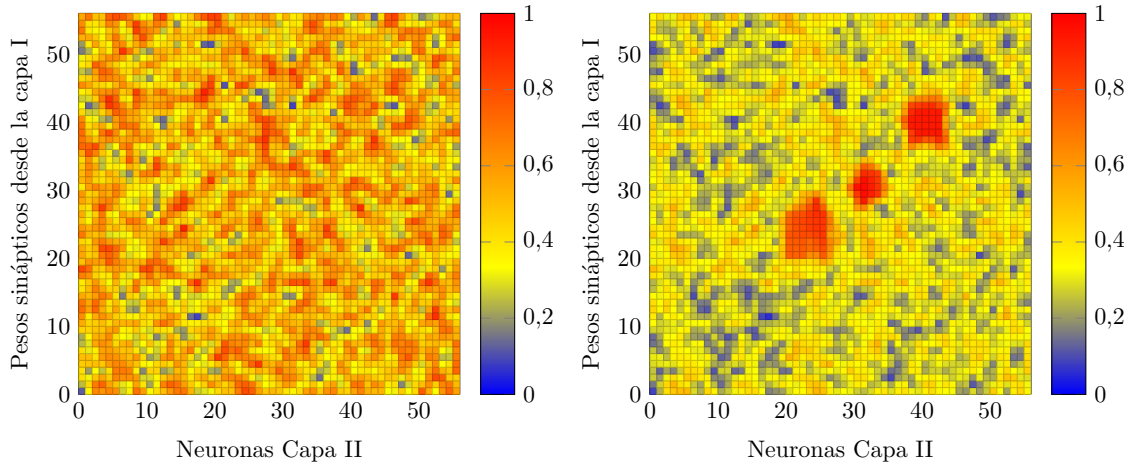


Figura 5.11: Topología de la red de asociación.

Figura 5.12: Ambiente de prueba para la tarea 1.



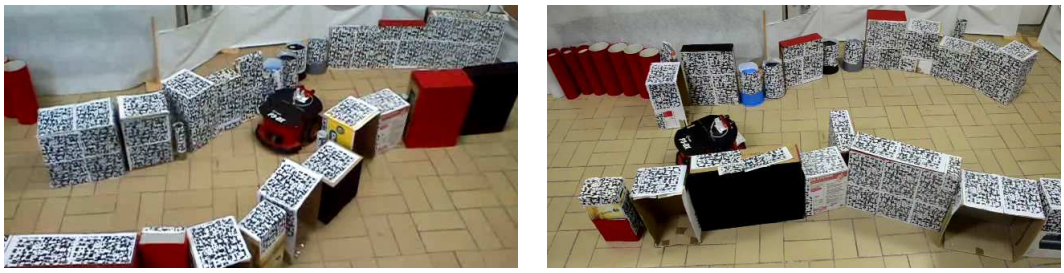
(a) Pesos inicializados de forma aleatoria.

(b) Pesos después de 40 PLP.

(c) Pesos después de 80 PLP.

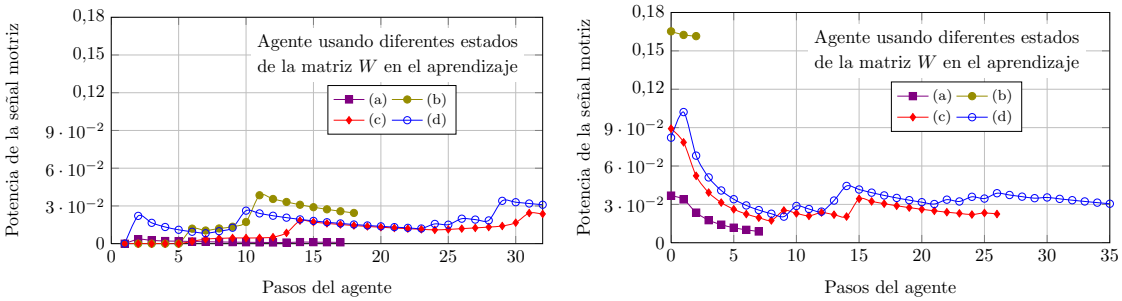
(d) Pesos al final del aprendizaje (160 PLP).

Figura 5.13: Pesos sinápticos (matriz W) de la red de asociación durante la fase de aprendizaje hebbiano.



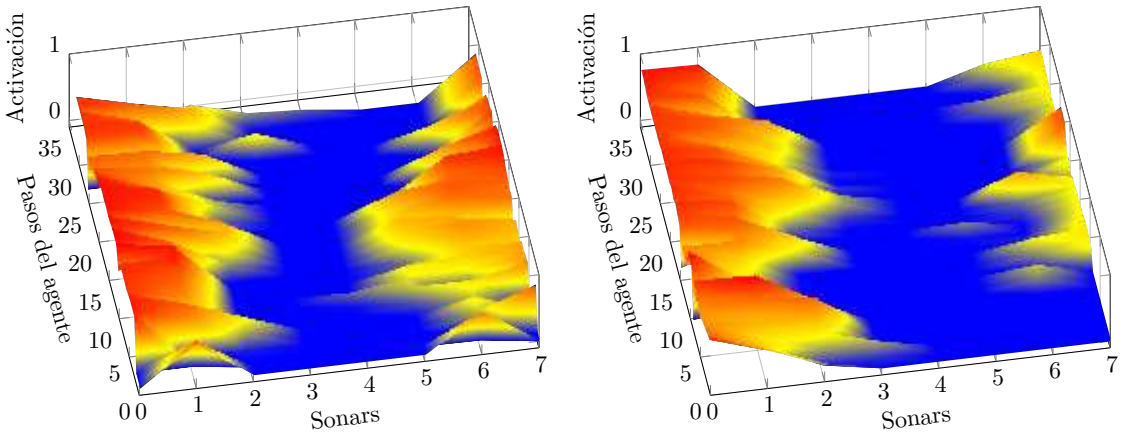
(a) Pasaje 1: Corredor estrecho en forma de 'S'. (b) Pasaje 2: Corredor con obstáculos en ambos lados.

Figura 5.14: Condiciones experimentales mostrando dos ambientes de prueba diferentes.

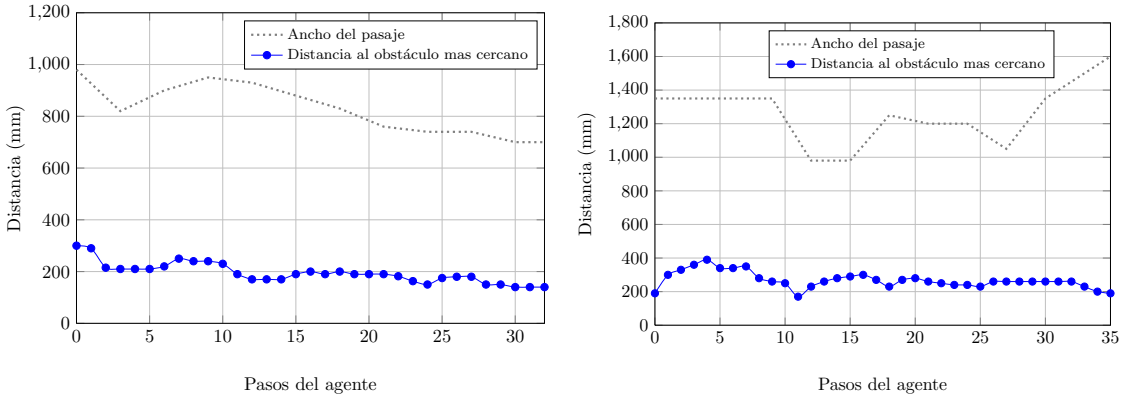


(a) Potencia de la señal motriz para el maze 1. (b) Potencia de la señal motriz para el maze 2.

Figura 5.15: Potencia de las señales motrices indicando la magnitud del cambio de dirección en el agente. La figura (a) corresponde al pasaje 1, mientras que la figura (b) al pasaje 2. Las curvas etiquetadas en cada subfigura como (a), (b), (c) y (d) corresponden respectivamente, al estado de la matriz W durante la fase de aprendizaje y mostrado en las figuras 5.13a, 5.13b,5.13c y 5.13d.



(a) Mapa topográfico de los sonares para el pasaje 1 (b) Mapa topográfico de los sonares para el pasaje 2



(c) Desempeño en el pasaje 1 (d) Desempeño en el pasaje 2

Figura 5.16: Datos experimentales para un individuo usando la matriz de asociación 5.13d. Las figuras a y b muestran respectivamente, el mapa topográfico construido a partir de la lectura de los sonares del agente para los pasajes 5.14a y 5.14b. Las figuras c y d describen la posición relativa del agente con respecto al ancho del pasaje.

Capítulo 6

Adquisición del concepto distancia

El trabajo presentado en el capítulo anterior 5 se obtuvo una estrategia de control motriz basada en la asociación multi-modal, específicamente en las modalidades visuales y táctiles de un agente artificial. Sin embargo, la codificación de los datos todavía descansa sobre una base geométrica al calcular un mapa de disparidad a partir de las imágenes provistas por la cámara estéreo.

Según estudios recientes en las ciencias cognitivas, la percepción de distancia en humanos no radica en un modelo geométrico sino en una asociación de información multi-modal (Gibson (1979); Braund (2007)). Por lo tanto, para dotar a un agente artificial con una noción de la distancia cimentada en sus capacidades sensori-motrices, se requiere que la información visual no codifique de forma explícita la información relacionada con la distancia a los objetos del entorno.

En este capítulo se describe la formulación de un modelo directo que en vez de recibir información a partir de un mapa de disparidad, reciba información visual a partir de las imágenes izquierda y derecha que provee el par de cámaras estéreo y que además tenga un repertorio de comandos motrices mas amplio que incluya giros a derecha e izquierda y movimientos hacia enfrente.

la pregunta a responder es si el modelo directo podrá dar cuenta de una noción de distancia, en base a información visual que no contiene ninguna información explícita de la distancia a los objetos y que a su vez esté cimentada en las capacidades sensori-motrices del agente.

6.1. Modelo directo propuesto

La representación esquemática del modelo propuesto se observa en la figura 6.1. Este, recibe como entrada una situación sensorial al tiempo t compuesta por las imágenes izquierda (I_i) y derecha (I_d) provenientes del par estéreo. Así mismo recibe un comando motriz a ejecutar M_t, este puede ser un movimiento hacia adelante o giros hacia la izquierda o derecha.

El modelo directo proporciona una predicción de lo que sucedería si el comando motriz fuese ejecutado. En la salida ($t = t+1$), la situación sensorial predicha esta compuesta por dos modalidades. La modalidad visual está dada por las imágenes izquierda y derecha $(I_i, I_d)^*_{t+1}$, mientras que la modalidad táctil T^*_{t+1} está dada por un estado binario de los parachoques del robot, indicando si hay o no colisión con un obstáculo. Aunque el robot cuenta con sensores de choque, por razones prácticas estos valores se obtuvieron de umbralizar los valores de los sonares.

Figura 6.1: Esquema del modelo directo propuesto.

6.2. Preparación de los datos de entrada

La asociación de las imágenes y el comando motriz implica un problema de alta dimensionalidad debido a que cada una de las imágenes provenientes del par estéreo tiene un tamaño de 320×240 píxeles, resultando en un total de $153,600$ datos. Para reducir el tamaño de este espacio de entrada se utilizaron dos estrategias: la elección de una región de interés (RDI) y la implementación de un algoritmo de "fovealización".

Posteriormente, y con base en la propuesta anterior para la obtención de una imagen fovealizada, este procesamiento se mejoró notablemente con el propósito de realizar PLP's de mas larga duración y con un grado de exactitud mayor.

con este propósito en mente, se situó al agente en un entorno conteniendo obstáculos únicamente de color rojo y a una mayor distancia que en el caso anterior. Debido a que agente de trabajo (ver sección 5.1) tiene una altura de 30 cms. por lo que objetos mas altos de esto no afectan su navegación segura, la RDI se eligió tomando como referencia el origen en la parte superior de la imagen, comenzando en la coordenada vertical del píxel 95 y terminando en el píxel 239. La RDI elegida se muestra en la figura 6.2b.

A esta RDI se le realizó un proceso de segmentación para resaltar la información relevante para el agente, esto es, obstáculos en su camino en una tarea de navegación. En primer lugar, se obtuvo una imagen de la RDI en valores de intensidad al seleccionar, en base a su histograma, el canal de color que representa con mayor contraste a los obstáculos sobre el fondo del entorno, el canal elegido es el correspondiente a la tonalidad del color verde. En segundo lugar, se le realizó un recorte con umbral utilizando el método de Otsu (Otsu (1975)) El resultado es una imagen en la que los obstáculos de la escena aparecen en niveles de gris mientras que el fondo aparece de color negro, ver figura 6.2c.

La segunda estrategia consistió en implementar un algoritmo de fovealización estableciendo una analogía con el ojo humano, este presenta una mayor densidad de células fotoreceptoras en la región conocida como fóvea. El resultado de esta operación es que las partes centrales de la imagen son representadas con un mayor número de píxeles mientras que las periféricas con un número menor. Este proceso es análogo a tener un sensor de cámara con diferentes resoluciones para diferentes áreas de la escena, mayor en el centro y menor en la periferia.

El trabajo Traver and Bernardino (2010) revisa diferentes métodos de fovealización de imágenes basados en diferentes transformaciones paramétricas. Entre estos se eligió el método de mapeo exponencial dimensionalmente independiente[1](Peters and Sowmya (1998)) consistente en un muestreo exponencial en las direcciones vertical y horizontal.

El muestreo producido es tal que se realiza una selección de puntos mas densa hacia el centro de la imagen que en su periferia. El resultado de aplicar esta fovealización a la RDI se puede apreciar en la figura 6.2d cuyas dimensiones son ahora de 32×10 píxeles logrando así una reducción de la información visual de $153,600$ a 320 píxeles.

Este método realiza el muestreo de diversos puntos de la RDI, espaciados entre sí en la dirección vertical y horizontal según las ecuaciones 6.1 y 6.2.

La ecuación 6.1 x y H representan, respectivamente, las filas seleccionadas y el alto de la RDI, mientras que i y h corresponden a las filas y el alto de la imagen fovealizada. De forma análoga, la ecuación 6.2, contiene los parámetros para el muestreo en la dirección horizontal, donde y y W representan las columnas seleccionadas y el ancho de la RDI, mientras que j y w las columnas y el ancho de la imagen fovealizada.

[1]Traducción libre del autor: Dimensionally-Independent Exponential Mapping

Los parámetros γ_f y γ_c controlan la densidad de muestreo en las filas y columnas respectivamente. Estos pueden tener valores entre 0 y ∞. Si se desea que la mayor densidad de muestreo en las imágenes resultantes se encuentre en el centro de la RDI, estos deben tener un valor dentro del rango $[0, 1]$. De forma específica, se eligieron $\gamma_f = 0{,}4$ y $\gamma_c = 0{,}6$ para lograr un balance entre el número de puntos muestreados en la región central y en la periferia de la RDI.

$$x = \frac{H-1}{2}\left(\frac{2i}{h-1}\right)^{\gamma_f} \tag{6.1}$$

$$y = \frac{W-1}{2}\left(\frac{2j}{w-1}\right)^{\gamma_c} \tag{6.2}$$

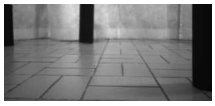
Una vez elegidos los píxeles muestreados, el valor final de intensidad en cada uno de estos se calculó mediante la convolución con un kernel gaussiano lineal, es decir, como el promedio ponderado del valor de intensidad con los píxeles vecinos y no elegidos en el muestreo.

Según la posición del píxel, el tamaño del kernel se reducía a medida que se efectuaba la convolución desde la periferia hasta la región central de la RDI.

Al implementar este proceso, se obtuvo una imagen fovealizada de 10 píxeles de alto $\times$ 32 píxeles de ancho con una región central fovealizada. Finalmente, se aplicó un filtrado gaussiano con un kernel de 3×3 para suavizar las transiciones de intensidad entre los píxeles vecinos.

El tamaño y la forma rectangular de la imagen fovealizada se eligió con el propósito de reducir el tamaño de los datos y de dar prioridad a la dimensión donde se reflejan los mayores cambios en las imágenes adquiridas por el par estéreo cuando al agente se desplaza a través de su entorno.

(a) Imagen original (320 $\times$ 240)



(b) Región de interés (RDI) (310 $\times$ 145)

(c) RDI umbralizada

(d) RDI fovealizada (32 $\times$ 10) (Esc 8:1)

Figura 6.2: Procesamiento de las imágenes de entrada.

6.3. Codificación del comando motriz

El espacio motriz del agente consiste en un conjunto de 3 acciones, girar 5° a la derecha, moverse hacia adelante 15 cms. o girar 5° a la izquierda. Cada uno de estos comandos motrices se transformó a un vector de 15 valores mediante el uso de 3 funciones gaussianas con igual desviación estándar

$\sigma = 1$ pero con distinta media μ, $\mu = 3$ para el giro hacia la izquierda, $\mu = 7$ para el movimiento hacia adelante y finalmente $\mu = 11$ para el giro hacia la derecha. Las curvas de la codificación de los comandos motrices se pueden observar en la figura 6.3. La codificación de los comandos motrices por un vector de 15 valores esta basada en buscar un equilibrio en los efectos que las tres modalidades (visual, táctil y motriz) deberán tener en el aprendizaje para cada una de las redes (ver Sección **??**).

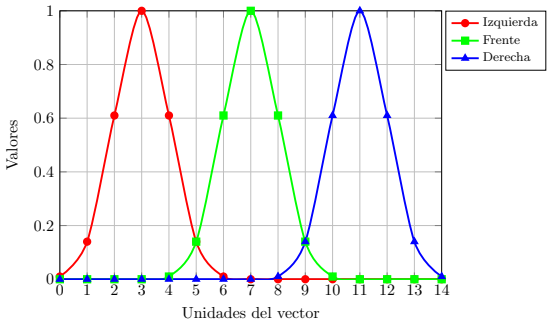


Figura 6.3: Codificación del comando motriz.

6.4. Sistema de Redes Neuronales Artificiales

El modelo directo (ver figura. 6.1) fue codificado a través de un sistema de redes neuronales artificiales (RNA) tipo perceptrón multicapa para realizar una predicción simétrica local. Esto significa que cada red recibe de cada imagen de entrada una región de 12×2 píxeles y un vector de 15 datos que codifica al comando motriz y produce regiones de 2×2 píxeles para cada una de las imágenes de salida y para el estado táctil, a excepción de las regiones predichas para la primera y última columna, las cuales son de 2×1 píxeles. El objetivo de este ajuste es lograr que las regiones de salida correspondan a la parte central de las regiones de entrada. En la figura 6.4 se observa en color azul la conexión típica de una de las redes que predicen una región de 2×2 píxeles y en color rojo la conexión de una de las redes para una región de 2×1 píxeles.

Dado que el tamaño de las regiones de entrada es mayor que el de las regiones de salida, existe un traslape, en la dirección horizontal, en todas las regiones de entrada a excepción de los bordes izquierdo y derecho en donde 4 redes reciben la misma información de entrada. Como resultado, se obtuvo un sistema compuesto por 85 redes neuronales artificiales distribuidas en una malla de 17 redes en el sentido horizontal y 5 redes en el sentido vertical.

El protocolo de recolección de patrones para el entrenamiento consistió en la preparación de una arena con el robot ubicado en el centro y obstáculos cilíndricos de diámetros 6.5, 11.5 y 16.5 cms., dispuestos alrededor de este. Se recolectaron 5727 patrones durante la ejecución de 150 trayectorias de movimientos en el entorno. Cada trayectoria se llevó a cabo seleccionando un comando motriz (adelante, izquierda, derecha) de forma aleatoria en cada paso. La trayectoria terminaba si se reportaba una señal de colisión por alguno de los 4 sonares frontales o si se alcanzaba la ejecución de 50 pasos sin presentarse una colisión. La señal de colisión se representó mediante una variable binaria que tomaba el valor de 1 cuando se detectaba un obstáculo a menos de 500 mm. (medida que equivale al tamaño del robot) o de 0 en caso contrario.

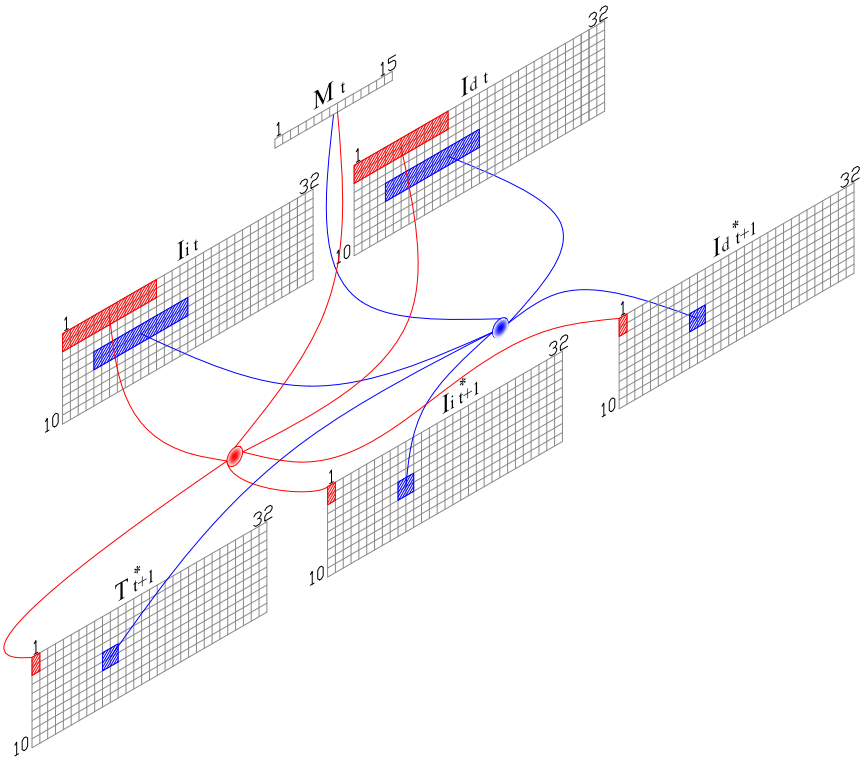


Figura 6.4: Distribución esquemática de la conexión de 2 redes neuronales artificiales que codifican el modelo directo. En azul se muestra la conexión típica de las redes del sistema y en rojo se aprecia la conexión para las redes de la primera y última columna

En cada paso de las trayectorias se almacenó el par de imágenes proveniente de la cámara estéreo antes y después de ser ejecutado el comando motriz, el propio comando motriz y el estado de colisión. El entrenamiento de las redes fue realizado por medio de una variación del algoritmo estándar de retro-propagación del error conocido como retro-propagación fuerte (*resilient back-propagation*) (Riedmiller and Braun (1993)), el cual es comparativamente mas rápido y eficiente.

La capacidad del agente para estimar distancia como un concepto cimentado esta íntimamente relacionado a la capacidad de este para realizar PLPs. Esto es, en una escena, el agente lleva a cabo una PLP la cual incluye la predicción de los estados táctiles. La evaluación de estas predicciones sensorimotrices puede considerarse como un concepto "*en X pasos colisiono* ". Aquí *X pasos* es una distancia en el sentido de que cada paso está relacionado a un comando motriz de desplazamiento en términos de las capacidades del agente.

6.5. Análisis de una PLP

En primera instancia para observar el resultado de realizar una PLP se sitúa un obstáculo frente al agente a una distancia de 75 cms. tal como se observa en la figura 6.5, dado que el modelo directo

codifica desplazamientos de 15 cms., esta distancia es equivalente a la ejecución de 5 movimientos hacia adelante para colisionar con el obstáculo.

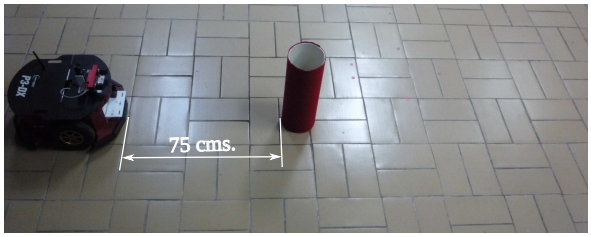


Figura 6.5: Obstáculo ubicado a 75 cms. del agente

(a) Imagen izquierda

(b) Imagen derecha

(c) Imagen fovealizada I_i

(d) Imagen fovealizada I_d

Figura 6.6: Imágenes originales y fovealizadas de la situación inicial con el obstáculo ubicado a 75 cms., distancia equivalente a la ejecución de 5 movimientos hacia adelante para colisionar con este.

Las imágenes capturadas por el par estéreo y sus correspondientes fovealizaciones para la situación mostrada en la figura 6.5 se pueden observar en la figura 6.6. Se aprecia como el obstáculo está desplazado con respecto al centro en la imagen derecha y en menor grado en la imagen izquierda. Las imágenes fovealizadas repiten esta característica siendo el desplazamiento de mayor magnitud en la imagen izquierda donde la representación del obstáculo se magnificó por el efecto de fovealización.

6.5.1. Predicciones para el estado visual

Las predicciones visuales resultantes se observan en la figura 6.7. Durante la primera predicción $t = t + 1$ y hasta la predicción a $t = t + 5$ se puede observar como el obstáculo va ocupando cada vez una mayor área en el extremo derecho para la imagen izquierda y de forma inversa para la imagen derecha. Implicando así, que conforme avanzan los instantes de la PLP, la representación del obstáculo en las predicciones se va desplazando hacia el centro del agente. A partir del instante $t = t+5$, se puede ver que la representación del obstáculo alcanzó los extremos en ambas imágenes. A partir de $t = t + 7$, surge de manera significativa ruido y falsas regiones en cada una de las predicciones.

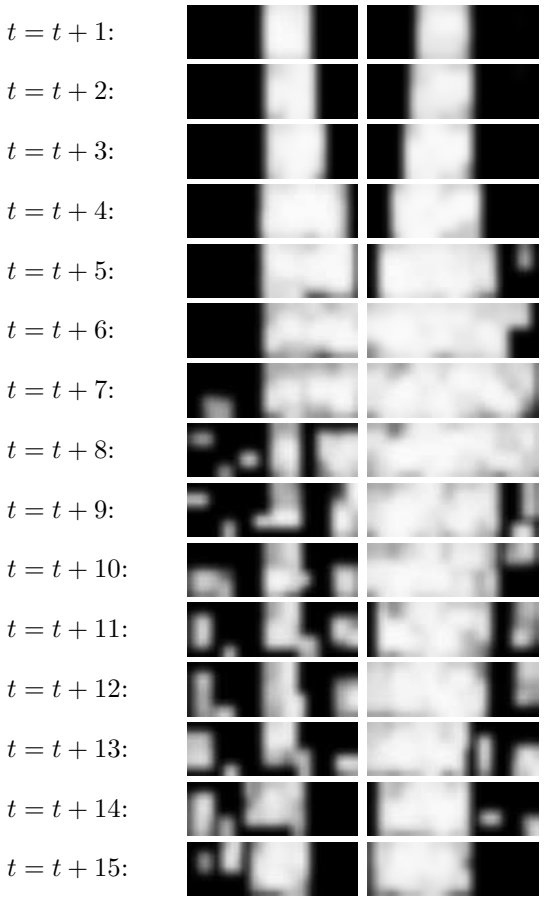


Figura 6.7: Imágenes resultantes en la *predicción de largo plazo* (PLP) de 15 movimientos hacia adelante con el obstáculo ubicado a 75 cms. La columna de la izquierda corresponde a las predicciones I_i^* mientras que la de la derecha a I_d^*. Escala 8:1.

6.5.2. Predicciones para el estado táctil

De forma correspondiente con las predicciones visuales mostradas en la figura 6.7, las predicciones del estado táctil para la misma situación se muestran en la figura 6.8. Cada una de estas gráficas fueron codificadas usando un mapa de calor, donde la mínima activación corresponde al valor de 0 y está indicado por las zonas de color azul, mientras que el máximo valor corresponde a 1 indicado por las zonas de color rojo.

El primer aspecto a resaltar es el hecho de tener valores continuos en un rango de [0-1], contrario a los valores binarios codificados durante el entrenamiento, donde 1 representó una colisión y 0 cualquier otro caso (Ver sección 6.4). Esta es una propiedad emergente del sistema y constituye la manifestación de una representación multi-modal que se codifica en el sistema de redes neuronales.

Al analizar estas gráficas se observa que desde el instante $t = t + 1$ hasta el $t = t + 3$ existe una activación en la región central llegando a un máximo en el instante $t = t + 5$ donde la gran mayoría

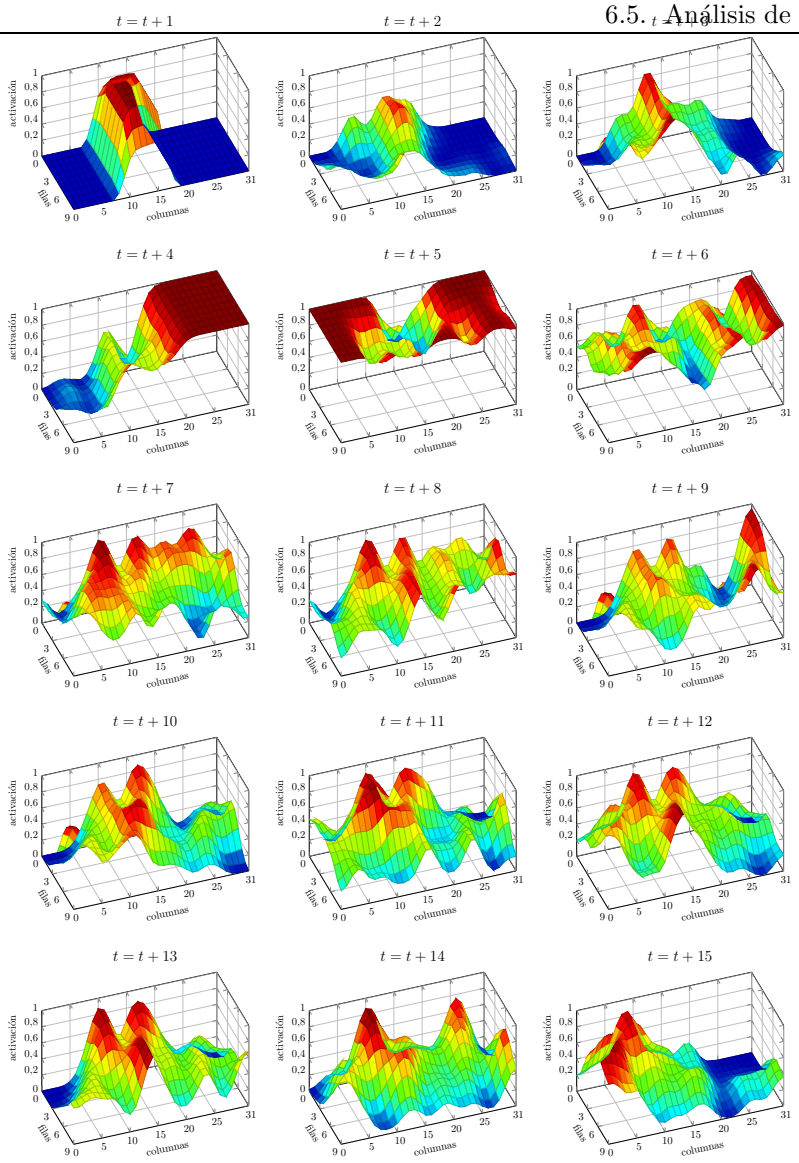


Figura 6.8: Representación tridimensional de las predicciones del estado táctil para la PLP de 15 movimientos hacia adelante con el obstáculo ubicado a 75 cms.

de las activaciones se muestran en color rojo.

Esto se relaciona con las predicciones visuales debido a que en este mismo instante $t = t + 5$, se observa que el obstáculo alcanzó el extremo interior para la predicción de la imagen izquierda I_i^* y se encuentra a punto de no ser percibido en su totalidad.

A partir del instante $t = t + 6$ la forma de las gráficas en estos mapas de calor ya no describen una estructura definida y se mantienen fluctuando hasta el final de la PLP.

Cabe recordar que las figuras 6.7 y 6.8 muestran resultados de PLP cuando el obstáculo esta localizado a $t = t+5$. Este hecho se puede observar en la predicción táctil del sistema, donde a partir

de una activación promedio máxima (0,83) esta fluctúa sin necesariamente mostrar una correlación con las predicciones visuales.

6.6. Análisis de PLP's para diferentes distancias

Para evaluar el modelo y su capacidad de predicción se llevaron a cabo PLP's para obstáculos situados a diferentes distancias del agente. El primer obstáculo se colocó a 15 cms. y se llevo a cabo una predicción de 15 movimientos hacia adelante. En seguida el obstáculo se colocó a 30 cms. del agente, llevándose a cabo la misma PLP. Esto esto se repitió incrementalmente cada 15 cms. hasta que el obstáculo se ubicó a 225 cms. del agente. En las figuras 6.9a y 6.9b se muestran las ubicaciones del obstáculo correspondientes a las distancias mínima y máxima para esta prueba. En cada movimiento simulado y para cada PLP se registraron las predicciones visuales y táctiles.

(a) Obstáculo a 15 cms.

(b) Obstáculo a 225 cms.

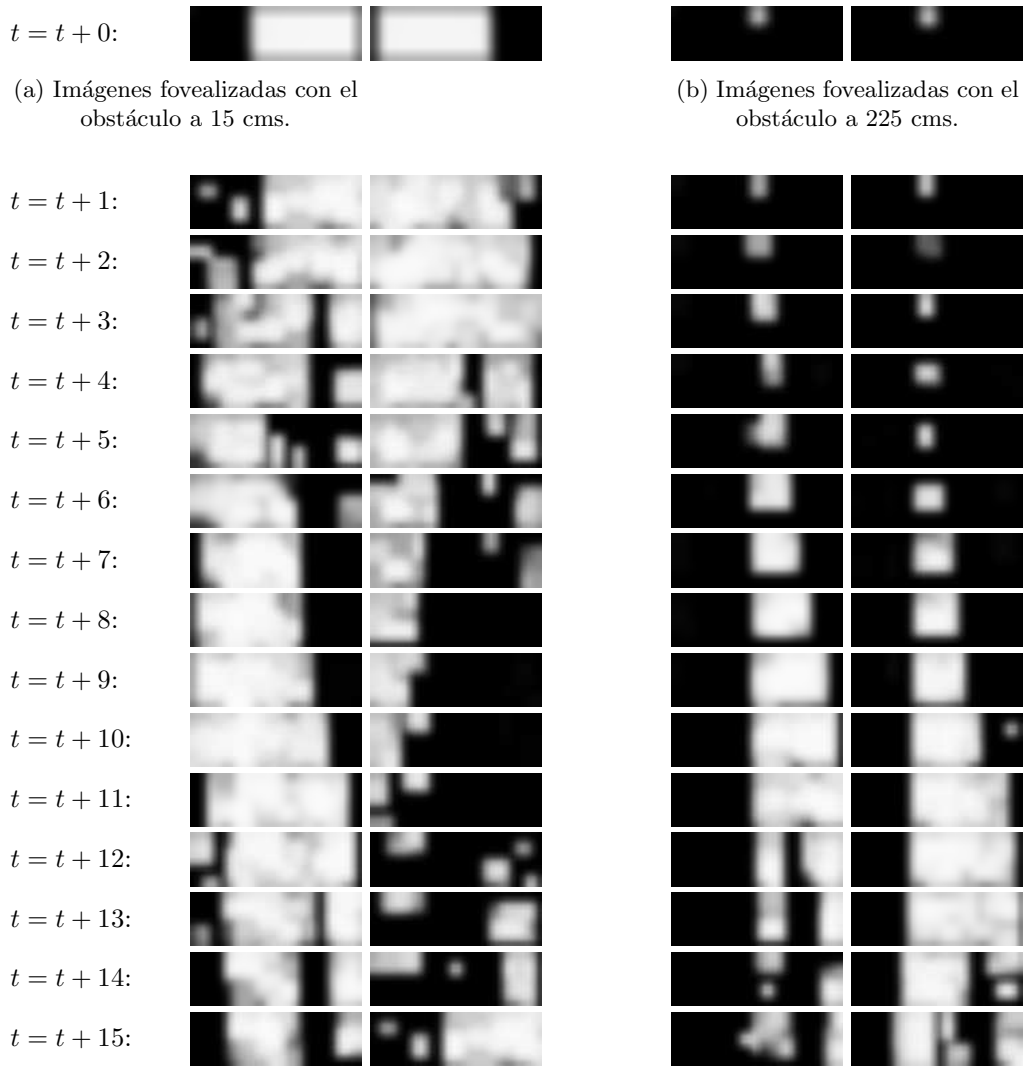
Figura 6.9: Distancia mínima y máxima a la que se ubica el obstáculo. La distancia mínima (a) y la máxima (b), son equivalentes respectivamente, a la ejecución de 1, y 15 movimientos hacia adelante para colisionar con el obstáculo.

6.6.1. Predicciones para los estados visuales

Con el propósito de dar una idea de como se comportan las predicciones visuales para las demás ubicaciones del obstáculo, se muestra en la figura 6.10 el resultado de llevar a cabo una PLP para la mínima y máxima distancia al obstáculo durante el experimento (ver figura 6.9).

En el caso del obstáculo ubicado a 15 cms., en el instante $t = t + 0$ correspondiente a la situación real inicial, se observar que el obstáculo ocupa la mayoría del campo visual en ambas imágenes (ver figura 6.10a). En el instante siguiente $t = t + 1$ de la PLP, equivalente a una colisión si el agente se moviera hacia adelante, se observa como en la predicción derecha la representación del obstáculo cubre casi la totalidad de la imagen. A partir de este punto la información contenida en las predicciones visuales no pueden ser una consecuencia directa del estado anterior al notarse la presencia de ruido en varias zonas de las imágenes (ver figura 6.10c).

Cuando el obstáculo se desplazó hasta una distancia de 225 cms., se puede ver como las imágenes fovealizadas del instante $t = t + 0$ (ver figura 6.10b) muestran un obstáculo ocupando una pequeña zona y ubicado en la mitad superior de estas, describiendo una localización lejana, muy diferente al caso anterior, donde el obstáculo se ubicó a 15 cms. Al realizar la PLP, se puede apreciar como la representación del obstáculo va ocupando una zona de mayor tamaño en la región central de ambas imágenes predichas, figura 6.10d. Sin embargo, al alcanzar el instante $t = t + 15$, esta representación no abarca casi la totalidad de las imágenes y presenta varias zonas de ruido.



(a) Imágenes fovealizadas con el obstáculo a 15 cms.

(b) Imágenes fovealizadas con el obstáculo a 225 cms.

(c) Predicciones visuales con el obstáculo a 15 cms.

(d) Predicciones visuales con el obstáculo a 225 cms.

Figura 6.10: Imágenes de la situación visual inicial y las resultantes en la *predicción de largo plazo* (PLP) de 15 movimientos hacia adelante. En (a) y (c) se muestran, respectivamente, la imagen fovealizada en el instante inicial $t = t + 0$ con el obstáculo ubicado a 15 cms., y las predicciones visuales obtenidas. Las imágenes de la columna de la izquierda corresponden a las predicciones I_i^* mientras que la de la derecha a I_d^*. De forma análoga, (b) y (d) corresponden a la situación cuando el obstáculo se ubicó a 225 cms. Escala 8:1.

6.6.2. Predicciones para los estados táctiles

La información contenida en los mapas de calor tridimensionales con los que se representa a las predicciones de la modalidad táctil, puede condensarse a través de una medida que simplifique su análisis. Para tal efecto es calculó el valor de activación promedio de los mapas de calor en cada instante de la PLP. En la figura 6.11 se muestra este valor para cada una de las 15 situaciones del experimento.

En primer lugar, se puede ver que en cada una de las curvas, la activación se incrementa a medida que el instante simulado se acerca al punto de colisión con el obstáculo, alcanza un máximo y disminuye ligeramente en el instante correspondiente a la colisión para posteriormente decrecer rápidamente y mantenerse oscilando alrededor de cierto valor.

En segundo lugar, se puede apreciar que para las posiciones en las que el obstáculo se encuentra a menos de 10 movimientos para colisionar, el punto en el que ocurriría la colisión (indicado en la figura como un círculo rodeando un punto en cada serie de datos) se correlaciona con los valores de mayor activación de las predicciones táctiles. A partir de la posición del obstáculo que corresponde a una distancia de 11 movimientos, la activación promedio no aumenta de manera significativa ni antes ni después del punto de colisión, concluyendo que el máximo número de instantes simulados para obtener una PLP que pueda anticipar una posible colisión, es de 10.

Finalmente, el valor de la máxima activación, en cada una de las PLP's donde el obstáculo se ubicó a menos de 150 cms. del robot, se encuentra alrededor de 0.8. Estableciendo de esta manera, un valor umbral para indicar un estado de colisión en el instante donde este se presenta o en el inmediatamente posterior. El hecho de que este valor sea aproximadamente el mismo para cada una de estas pruebas, da cuenta de la capacidad en el robot para estimar la distancia a un obstáculo en función del número de movimientos antes de que presente una colisión.

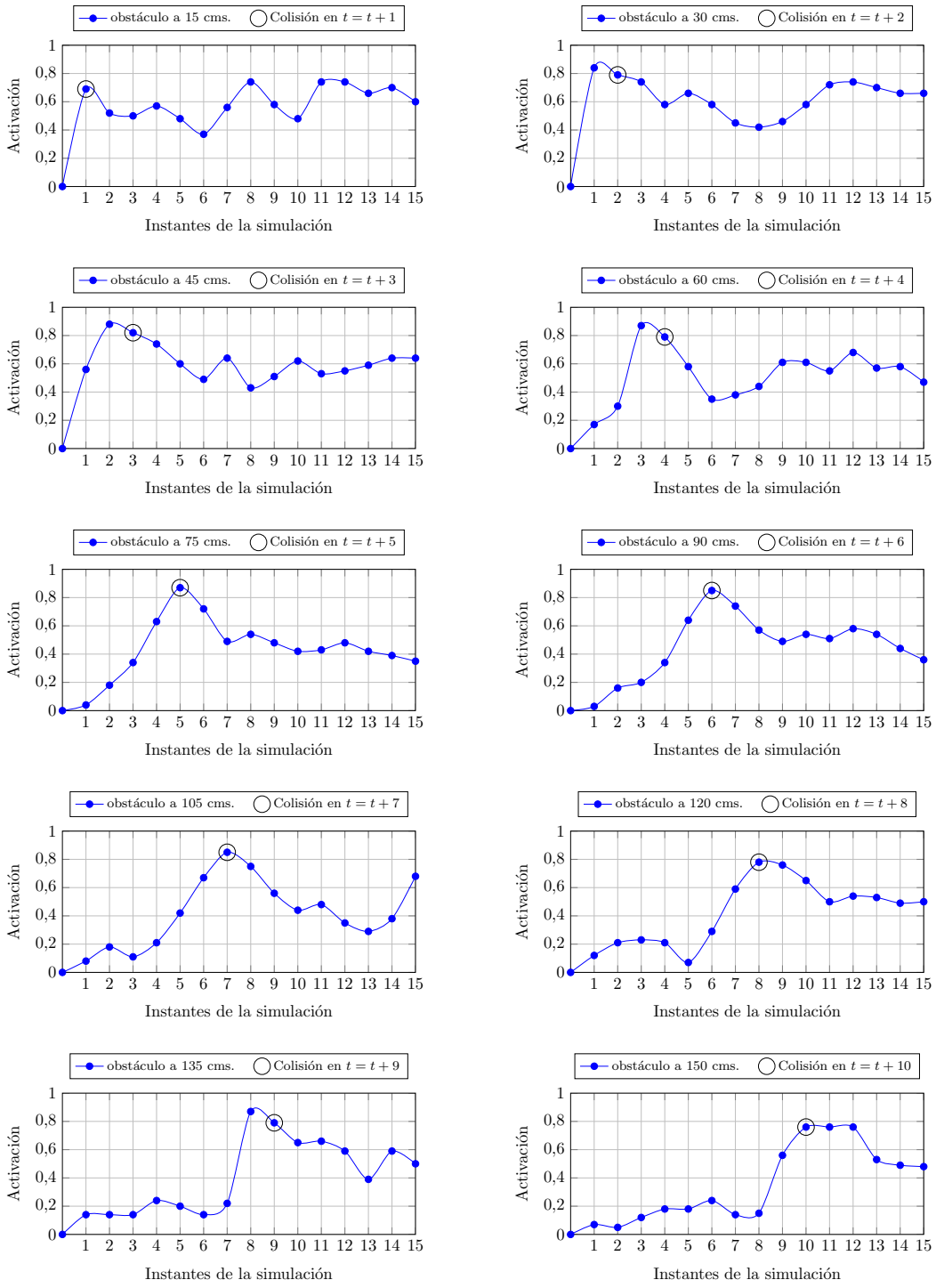


Figura 6.11: Valor promedio de las predicciones táctiles para las 15 diferentes ubicaciones del obstáculo

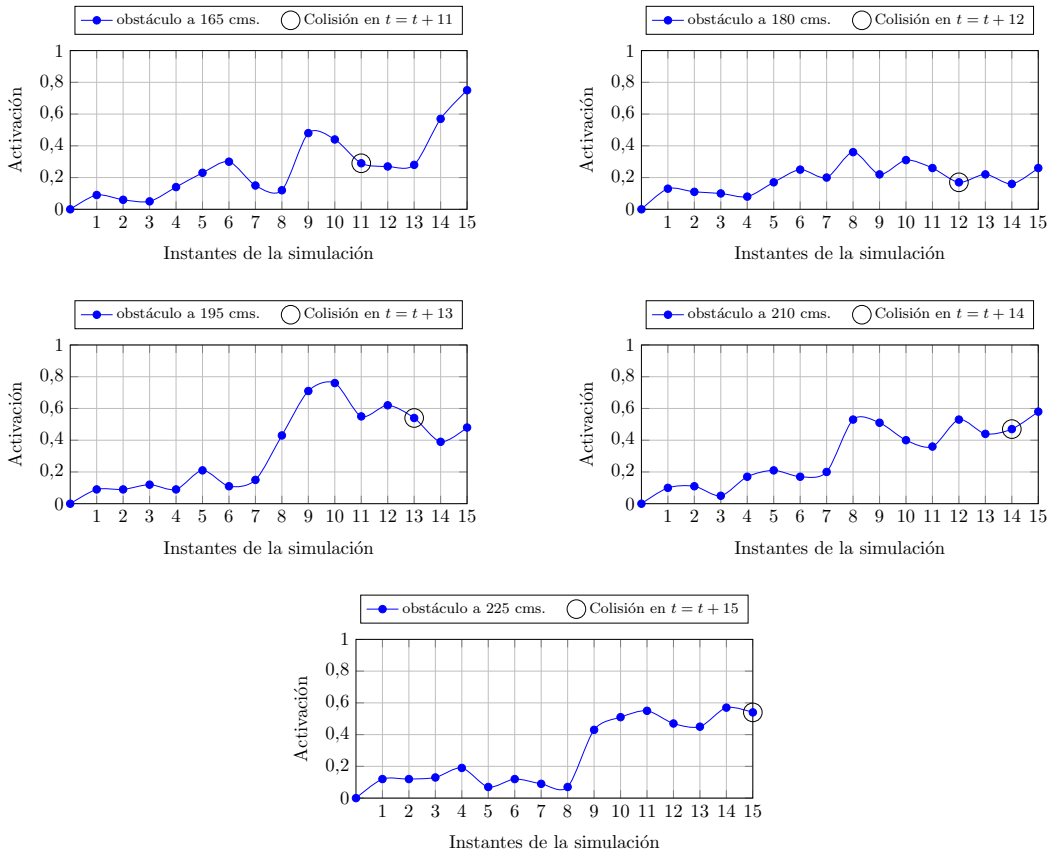


Figura 6.11: Valor promedio de las predicciones táctiles para las 15 diferentes ubicaciones del obstáculo

Capítulo 7

Adquisición del concepto *pasabilidad*

El objetivo de este experimento es mostrar como, desde la robótica cognitiva, es posible proveer a un agente artificial con la capacidad para juzgar la " *pasabilidad*" de una apertura. Este concepto de "*pasabilidad*" es similar al de "*distancia a*" descrito en el experimento anterior.

En humanos se ha encontrado que disponemos de una medida escalada a nuestras dimensiones corporales para juzgar el ancho de una apertura. Específicamente se encontró una relación que da cuenta de la habilidad para juzgar la "pasabilidad" en términos de las dimensiones corporales de cada persona en base al ancho de los hombros y al ancho de la apertura que se pretende atravesar Warren and Whang (1987).

Dado que el espacio motriz está constituido por 3 movimientos diferentes (derecha 5°, adelante 15 cms. e izquierda 5°), la combinación de estos permite la construcción de diferentes PLP's a partir del modelo directo propuesto (ver figura 6.1). Estas combinaciones pueden ser representadas en una gráfica de árbol, figura 7.1, donde el nodo inicial es el estado visual actual del robot y cada rama es derivada a partir del comando motriz a ejecutarse.

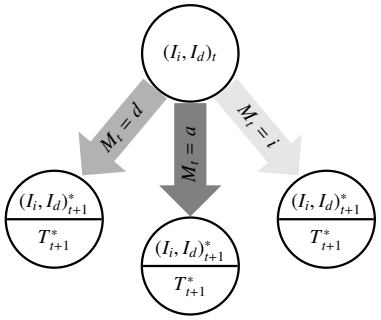


Figura 7.1: Representación del modelo directo como un árbol de predicciones.

Haciendo uso de las asociaciones multi-modales aprendidas, el agente es capaz de generar estructuras de tipo árbol, en donde el análisis de las predicciones táctiles es suficiente para dotar al agente con el concepto de "*pasabilidad*" .

En este contexto, este experimento se dividió en dos etapas. Una etapa inicial de carácter exploratorio y otra de corrección de la trayectoria. El propósito de la primera etapa es reconocer la ubicación de la apertura por la que el agente podría pasar y comenzar a dirigirse hacia ella. Durante la segunda etapa, en caso de ser necesario, se corrige el curso de la trayectoria para que el agente logre pasar a través de la apertura sin colisionar con los bordes de la misma.

Para la primera etapa se construye un árbol de predicciones de profundidad 10 y de 11 ramas (ver figura 7.2a), esto se logra al realizar 11 diferentes PLPs con 10 instantes de simulación cada una, esto es, para cada PLP se simulan 10 comandos motrices. La dirección de las flechas indica el tipo de comando motriz elegido en cada paso de la simulación. Por ejemplo, la rama 1 indica la

simulación de 5 giros hacia la derecha y en seguida 5 movimientos hacia adelante (d-d-d-d-d-a-a-a-a-a), la rama 6 indica 10 movimientos hacia adelante (a-a-a-a-a-a-a-a-a-a) y la rama 11, 5 giros hacia la izquierda y 5 movimientos hacia adelante (i-i-i-i-i-a-a-a-a-a).

La elección de estas ramas para el árbol de predicciones, obedeció al hecho del aumento exponencial del número de estas a medida que aumenta la profundidad del árbol (para una profundidad de 10 existen: $3^{10} = 59049$ ramas). Por esta razón se eligieron únicamente las ramas que indican los comandos motrices mas relevantes para el agente, es decir, aquellas en donde predominan los movimientos hacia adelante, ya que estos son los únicos en los que se podrían presentar colisiones.

Una vez que se realizaron las 11 PLP's se eligió la de menor activación promedio acumulada para la modalidad táctil y se ejecutaron únicamente los movimientos indicados por esta rama hasta el instante $t = t + 6$. Esto permite llevar a cabo correcciones de la trayectoria mediante un segundo árbol de predicciones en caso de ser necesario. La elección del instante $t = t + 6$ para disparar el segundo árbol es por que asegura que cualquiera que haya sido la rama con menor activación como mínimo se ejecute un movimiento hacia adelante (e.g. las ramas en los extremos del árbol).

Para la segunda etapa se ejecuta en primer lugar la rama 2 de un árbol de predicciones de profundidad 3 y 3 ramas: (d-a-a), (a-a-a) y (i-a-a) (ver figura 7.2b). En dado caso de que la activación promedio acumulada para el estado táctil para esta rama llegue a superar el valor umbral de 0.8, las ramas 1 y 3 son llevadas a cabo para corregir el curso de la trayectoria a través de la ejecución de los movimientos indicados en la rama con la mínima activación.

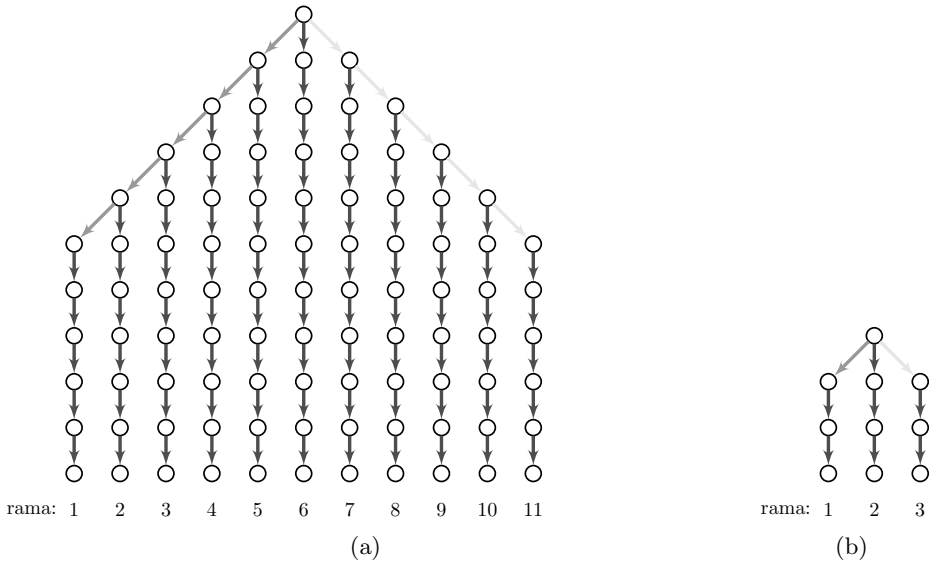
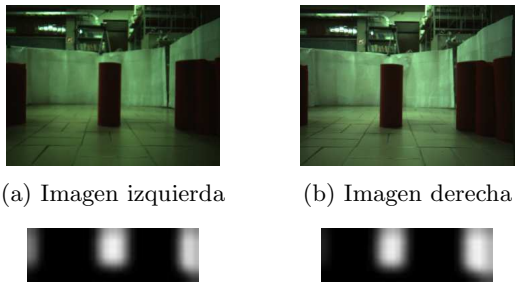


Figura 7.2: Arboles de exploración inicial y para la corrección de trayectoria con profundidad de 10 y 3 niveles, (a) y (b) respectivamente.

Este experimento consistió en situar al agente frente a dos aperturas, donde una de ellas es mas pequeña que el tamaño de su cuerpo y la otra de mayor tamaño como el mostrado en la figura 7.3. Generando un árbol basado en las PLP's, este podría determinar cual de ellas es por la que puede pasar sin colisionar y cual no; esto sin la necesidad de ejecutar de forma explícita una secuencia de acciones. Las imágenes captadas por el pár estéreo y sus correspondientes fovealizaciones para este

ambiente se muestran en la figura 7.4.

Figura 7.3: Ambiente para el experimento de adquisición del concepto de *pasabilidad*. El agente se encuentra situado frente a una escena visual que muestra dos aperturas, según la imagen, por la de la izquierda el agente no puede pasar mientras que por la de la derecha si.



(a) Imagen izquierda (b) Imagen derecha

(c) Imagen fovealizada I_i (d) Imagen fovealizada I_d

Figura 7.4: Imágenes originales y fovealizadas de la situación inicial mostrada en la figura 7.3

7.1. Predicciones para los estados táctiles

Con el objeto de caracterizar el desempeño del árbol de predicciones inicial se muestra la gráfica de las activaciones promedio acumuladas del estado táctil para cada una de las ramas que lo conforman, (ver Fig. 7.5).

En primer lugar se observa que la rama con mayor activación es la rama 6 del árbol, siendo la rama crítica ya que si se llegaran a ejecutar los movimientos hacia adelante indicados por esta, llevaría a que el agente colisione con el grupo de obstáculos en la parte central.

En segundo lugar las ramas de menor valor son las del extremo derecho del árbol (8, 10 y 11) con un valor activación acumulada menor a 2.0. Estas indican una serie de movimientos que dirigirían

al agente hacia la apertura por la que puede pasar, sin embargo la rama con la mínima activación acumulada es la rama 8 (i-i-a-a-a-a-a-a-a-a) con un valor final de 1.14.

Finalmente se muestra en una línea resaltada el instante $t = t+6$ el cual fija el máximo número de movimientos a ejecutarse e inicia la etapa de corrección de la trayectoria.

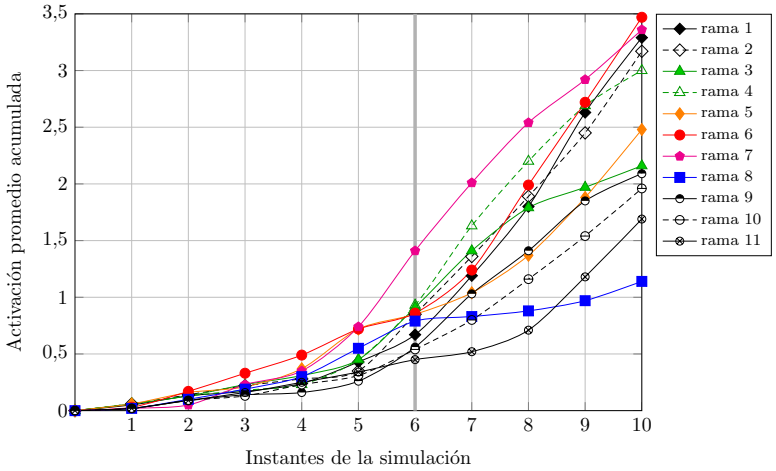


Figura 7.5: Activaciones promedio acumuladas del árbol de predicciones de la primera etapa para el entorno mostrado en la figura 7.3. Se resalta el instante $t = t + 6$.

7.2. Predicciones para los estados visuales

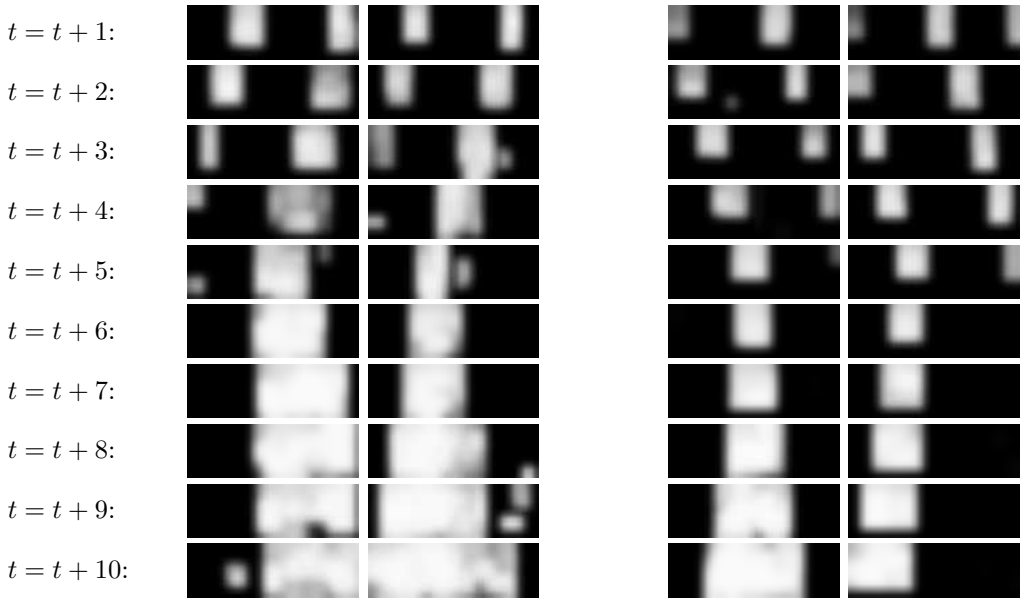
Con el propósito de complementar la descripción del comportamiento del sistema y con respecto al entorno mostrado en la figura 7.3 y la situación visual inicial $t = t+0$ indicada en la figura 7.4, se muestran las predicciones visuales para las ramas 1 y 11 correspondientes a los extremos del árbol de predicciones (ver figura 7.6). Además se muestran las predicciones visuales para las ramas de máxima y mínima activación promedio acumulada para el estado táctil, 6 y 8 respectivamente (ver figura 7.7).

En la figura 7.6 se aprecia como las predicciones visuales reflejan el cambio en la posición de los obstáculos debido a los giros simulados por el agente hasta el punto en que los obstáculos situados en uno de los extremos quedó ubicado en el centro de las imágenes predichas en el instante $t = t+5$. A partir de este instante, en ambas ramas de la simulación este obstáculo ocupó una mayor zona en las imágenes debido a la simulación de los movimientos hacia adelante hasta el instante $t = t + 10$.

En la figura 7.7 muestra dos situaciones contrastantes. En primera instancia, las predicciones visuales de la rama 6 (ver figura 7.7a), la cual tuvo la máxima activación promedio acumulada para el estado táctil, muestra como el grupo de obstáculos de la parte central del entorno crece a medida que avanzan los instantes en la simulación hasta ocupar casi la totalidad para las predicciones izquierda I_i^* y derecha I_d^* en el instante $t = t + 10$.

En segunda instancia, las predicciones visuales de la rama 8 (ver figura 7.7b), la cual presentó la mínima activación, muestra como en los instantes $t = t + 3$ y $t = t + 5$ desaparece del campo visual el grupo de obstáculos ubicado a la izquierda del agente, primero en las predicciones visuales

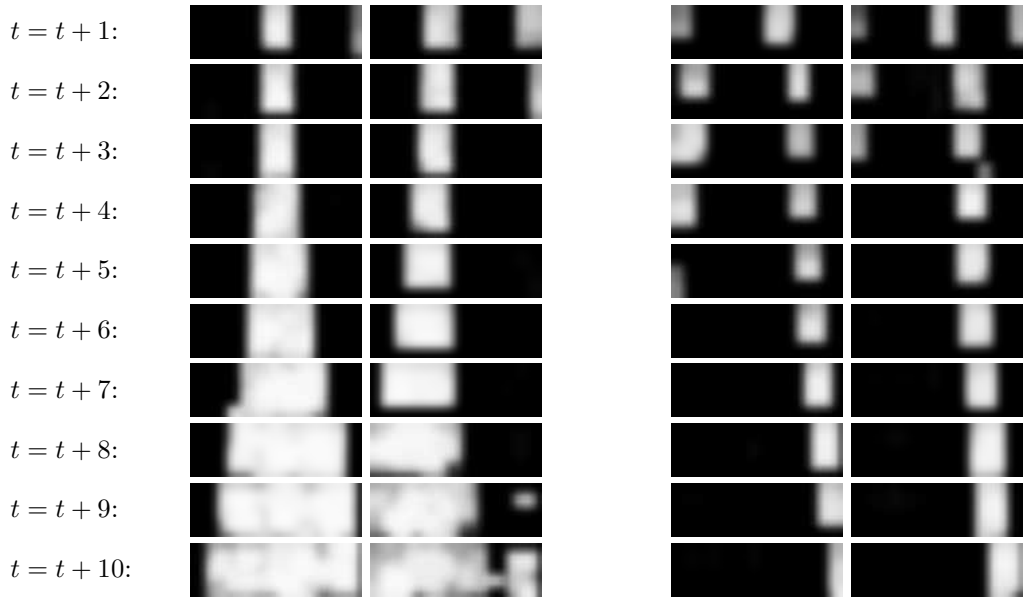
del lado derecho I_d^* y luego en las del lado izquierdo I_i^*. A partir del instante $t = t + 6$, se puede observar como se va formando un claro o una zona libre de obstáculos en la parte central de ambas predicciones, en las que únicamente aparece el grupo de obstáculos del lado derecho.



(a) Predicciones visuales para la rama 1 (d-d-d-d-d-a-a-a-a-a)

(b) Predicciones visuales para la rama 11 (i-i-i-i-i-a-a-a-a-a)

Figura 7.6: Imágenes de las predicciones visuales para las ramas 1 y 11, (a) y (b) respectivamente. Las imágenes de la columna de la izquierda corresponden a las predicciones I_i^* mientras que la de la derecha a I_d^*. Escala 8:1.



(a) Predicciones visuales para la rama 6 (a-a-a-a-a-a-a-a-a-a)

(b) Predicciones visuales para la rama 8 (i-i-a-a-a-a-a-a-a-a)

Figura 7.7: Imágenes de las predicciones visuales para las ramas 6 y 8, (a) y (b) respectivamente. Las imágenes de la columna de la izquierda corresponden a las predicciones I_i^* mientras que la de la derecha a I_d^*. Escala 8:1.

7.3. Elección de la apertura

La trayectoria seguida por el agente se puede observar en la figura 7.8 donde se muestran las activaciones promedio acumuladas en cada instante de la secuencia de movimientos ejecutados por el agente. El color de cada nodo es proporcional a la activación donde el azul y el rojo codifican los valores mínimo y máximo, respectivamente.

Se muestran únicamente los nodos hasta el instante $t = t + 6$, acorde con los discutido anteriormente en la sección 7. Se observa el curso de la trayectoria a través de los primeros 6 movimientos indicados por la rama 8 y en seguida únicamente ejecuta 3 series de 3 movimientos hacia adelante, dado que ninguna de las siguientes PLP's de la rama 2 del árbol de predicciones de la figura 7.2b superó el valor de 0.8, evitando así llevar a cabo una segunda etapa para corregir la trayectoria del agente. El vídeo del experimento se puede encontrar en el siguiente vínculo: `http://youtu.be/ljmaeQQW4Lg`.

De forma similar al experimento anterior, se presenta ahora un caso en el que el agente realizó una corrección de su trayectoria. Para esto se ubicó al agente en el entorno mostrado en la figura 7.9, en el cual el agente se encuentra 30 cms. (el equivalente 2 movimientos hacia adelante) mas cerca al grupo de obstáculos de la parte central que en el caso anterior (ver entorno descrito en la figura 7.3). Las imágenes fovealizadas correspondientes al instante inicial $t = t + 0$ para este ambiente se

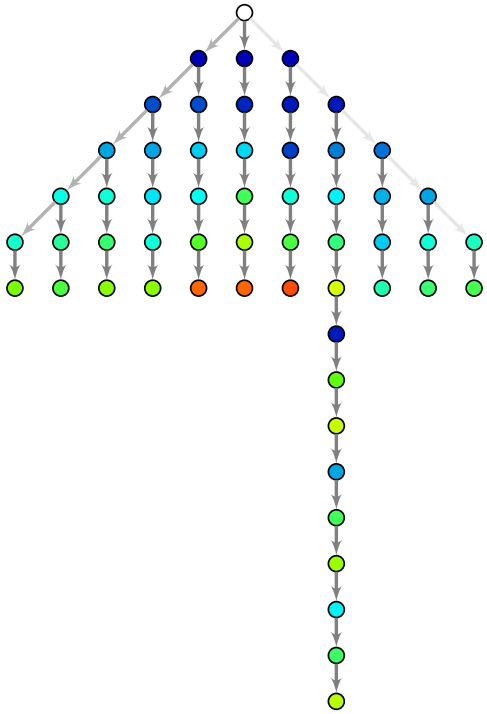


Figura 7.8: Árbol de las activaciones promedio acumuladas para la modalidad táctil durante toda la trayectoria en el entorno 7.3.

muestran en la figura 7.10.

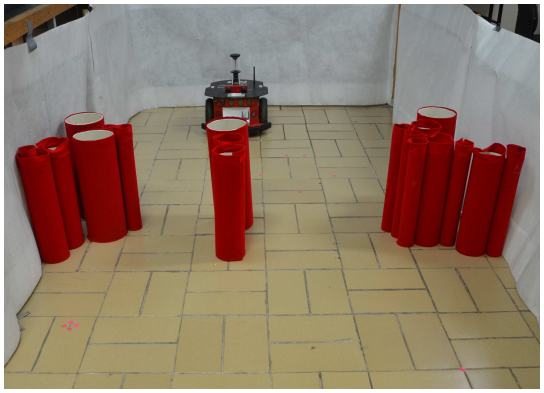


Figura 7.9: Segundo ambiente para el la prueba del concepto de *pasabilidad* realizando una corrección en su trayectoria. El agente puede pasar por la apertura de su lado izquierdo mientras que no por la de su lado derecho.

(a) Imagen fovealizada I_i (b) Imagen fovealizada I_d

Figura 7.10: Imágenes fovealizadas de la situación inicial mostrada en la figura 7.9

De igual manera, para caracterizar el desempeño del árbol de predicciones inicial, se muestra la gráfica de las activaciones promedio acumuladas para cada una de las ramas que lo conforman, (ver Fig. 7.11).

Se observa que la rama con mayor activación es la rama 6 del árbol, la que codifica la sencuencia de 10 movimientos hacia adelante. Además, las ramas de menor valor son las del extremo derecho del árbol (9, 10 y 11), con un valor de activación acumulada menor a 1.5. Estas indican una serie de movimientos que dirigirían al agente hacia la apertura por la que puede pasar, sin embargo la rama con la mínima activación acumulada es la rama 10 (i-i-i-i-a-a-a-a-a-a), con un valor final de 1.24. Nótese que esta rama está ubicada mas hacia el extremo que en el caso anterior (ver 7.5), indicando que esta vez son necesarios la ejecución de mas giros hacia la izquierda para poder pasar a través de la apertura sin colisionar.

Se muestra de igual forma, en una línea resaltada, el instante $t = t + 6$ el cual fija el máximo número de movimientos a ejecutarse e inicia la etapa de corrección de la trayectoria.

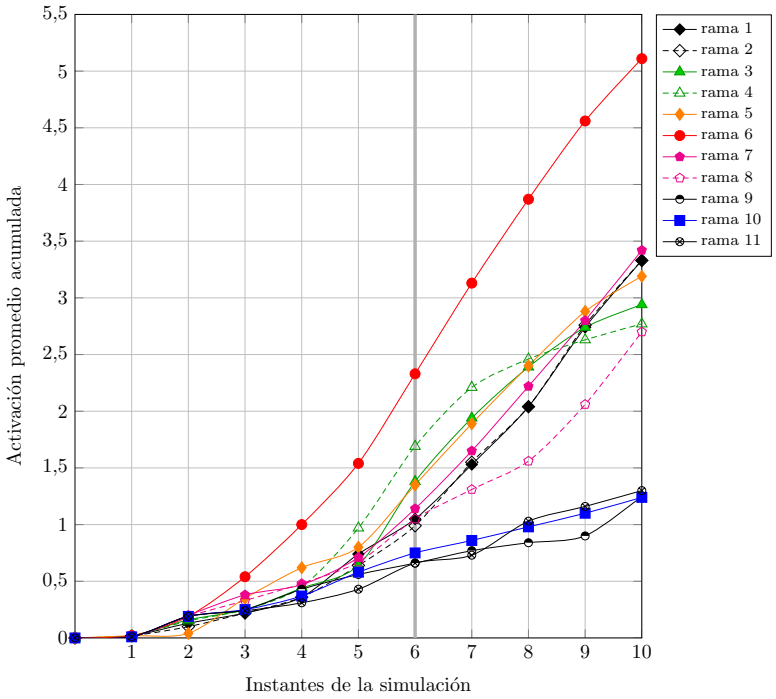


Figura 7.11: Activaciones promedio acumuladas del árbol de predicciones para la primera etapa para el entorno mostrado en la figura 7.9. Se resalta el instante $t = t + 6$.

La tarea restante consiste en corregir el curso de la trayectoria, en caso de ser necesario, a través

de la ejecución del árbol de predicciones de profundidad 3 (ver figura 7.2b).

La primera PLP en llevarse a cabo es la correspondiente a la rama 2 del árbol (a-a-a) con el objeto de determinar si al final de los 3 movimientos el agente llegaría a colisionar, es decir, si la activación promedio acumulada para la predicción táctil superara el valor de 0.8. En caso de que sucediera esto, se llevarían a cabo las PLPs para las otras 2 ramas del árbol con el objeto de determinar cual tendría la menor activación acumulada. En caso contrario, el agente únicamente ejecuta la secuencia de movimientos indicados por esta rama.

Al inicio de esta segunda etapa, la PLP de la rama 2 superó el valor de 0.8, haciendo necesaria la realización del árbol de predicciones con sus 3 ramas. La gráfica que muestra la activación promedio acumulada para este árbol se muestra en la Fig. 7.12, en la que se aprecia como la rama 2 supera el valor umbral, mientras que la rama 1 (d-a-a) es la que tiene el valor mínimo.

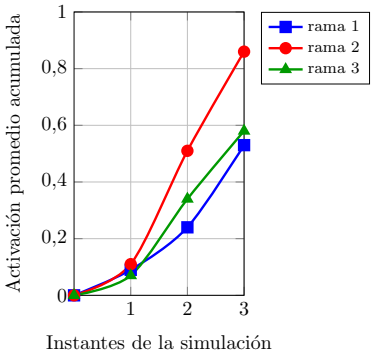


Figura 7.12: Activaciones promedio acumuladas para el árbol de predicciones durante la corrección de la trayectoria del agente.

La trayectoria seguida por el agente se ilustra en el árbol de la figura 7.13, de igual forma, el color de cada nodo de las activaciones promedio acumuladas en es proporcional a la magnitud de la activación, donde el azul codifica el valor mínimo (0) y el rojo y el máximo (1).

Dado que la rama con la mínima activación promedio acumulada fué la rama 10, se ejecutaron la secuencia de movimientos (i-i-i-i-a-a) indicados por esta rama. Se puede observar la tendencia que tuvieron las demás ramas de PLP, de las cuales la rama 6 presenta el valor de mayor activación mientras que las ramas 9, 10 y 11 son las que tienen un menor valor.

En seguida, el agente lanzó un árbol de predicciones de profundidad 3 y corrigió su trayectoria al elegir la rama 1 que corresponde a la secuencia: d-a-a. Posteriormente el agente ejecutó únicamente 5 series de 3 movimientos hacia adelante, ya que ninguna de las PLP para la rama central de estos árboles superó el valor umbral de colisión fijado en 0.8, por lo que el agente no necesitó realizar otra corrección adicional en el curso de su trayectoria, logrando de esta forma pasar a través de la apertura correcta. El vídeo del experimento se puede encontrar en el siguiente vínculo: `http://youtu.be/rEBNrrAymM4`.

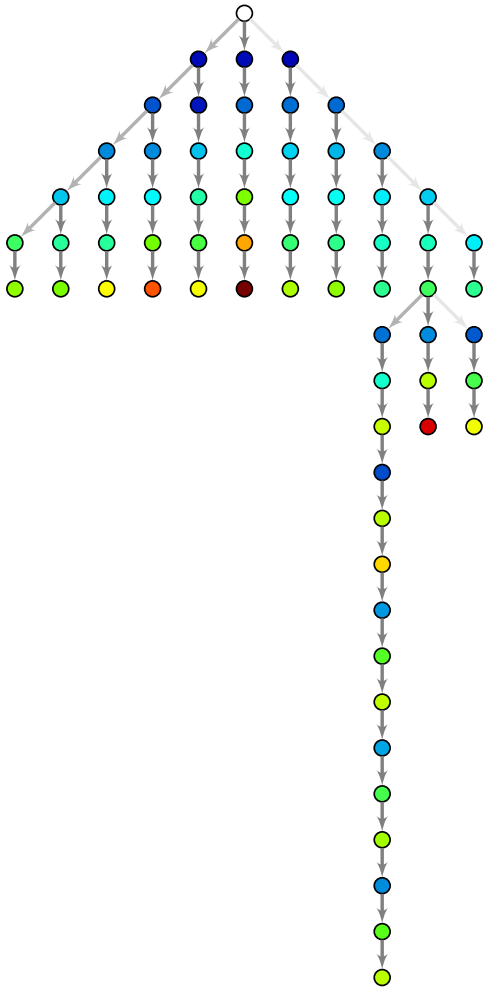


Figura 7.13: Árbol de las activaciones promedio acumuladas durante toda la trayectoria.

Capítulo 8

Navegación a través de un corredor de obstáculos

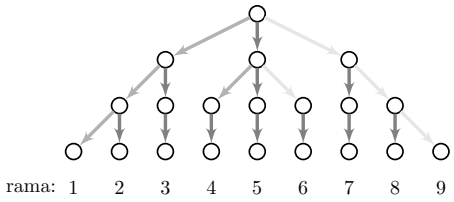


Figura 8.1: Árbol de exploración durante la navegación en un corredor de obstáculos.

Figura 8.2: Ambiente de prueba: Corredor de obstáculos

(a) Imagen fovealizada I_i (b) Imagen fovealizada I_d

Figura 8.3: Imágenes fovealizadas de la situación inicial mostrada en la figura 8.2

(a) Posición inicial

(b) Ejecución de la rama 9

Figura 8.4: Acción correctiva anticipada debida al árbol 2

(a) Posición inicial

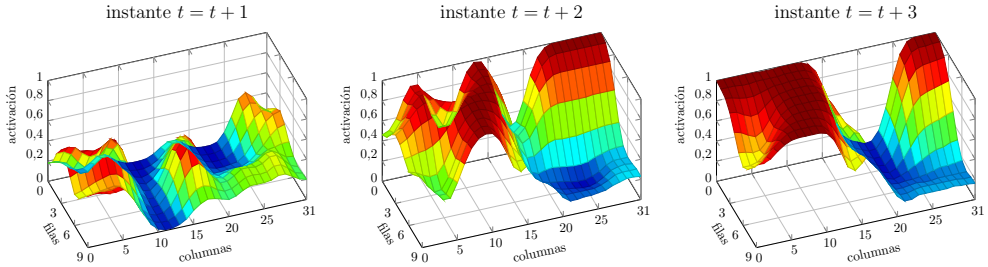
(b) Ejecución de la rama 9

Figura 8.5: Acción correctiva anticipada debida al árbol 3

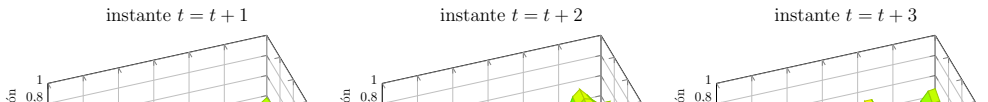
(a) Posición inicial

(b) Ejecución de la rama 1

Figura 8.6: Acción correctiva anticipada debida al árbol 4



(a) Predicciones táctiles para la rama 5 (a-a-a)



(a) Posición inicial

(b) Ejecución de la rama 1

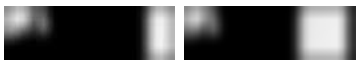
Figura 8.7: Acción correctiva anticipada debida al árbol 5

(a) Orientación hacia la salida

(b) Posición de salida

Figura 8.8: Acción correctiva anticipada debida al árbol 5

$t = t + 0$:

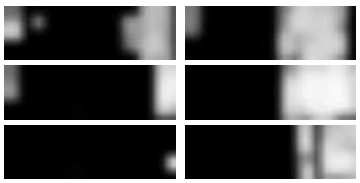


(a) Situación visual inicial

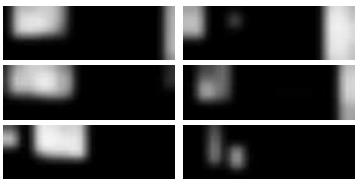
$t = t + 1$:

$t = t + 2$:

$t = t + 3$:



(b) Predicciones visuales para la rama 3 (d-a-a)



(c) Predicciones visuales para la rama 9 (i-i-i)

Figura 8.12: Imágenes de las predicciones visuales para las ramas 3 y 9 del árbol 3, (b) y (c) respectivamente.

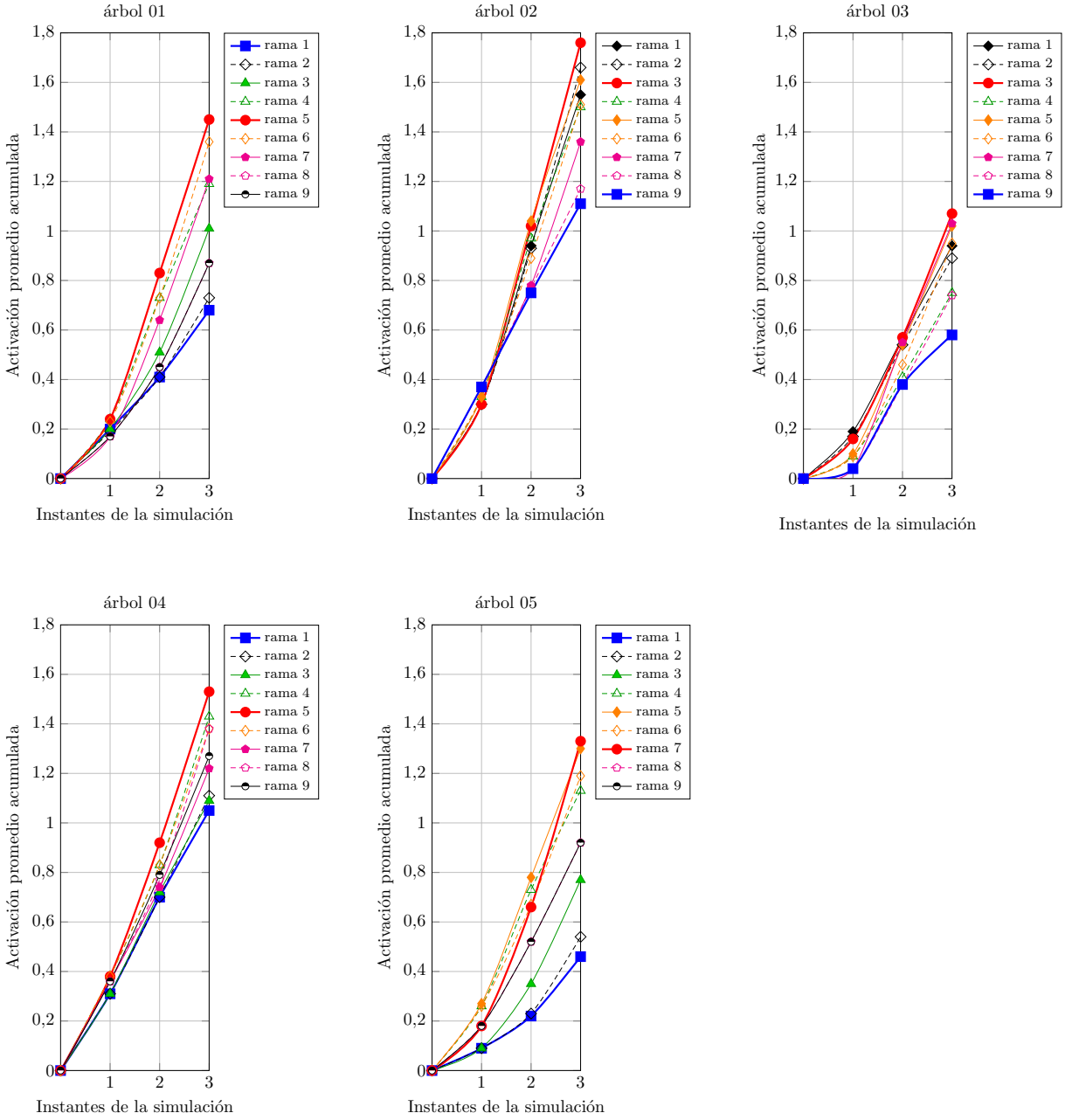
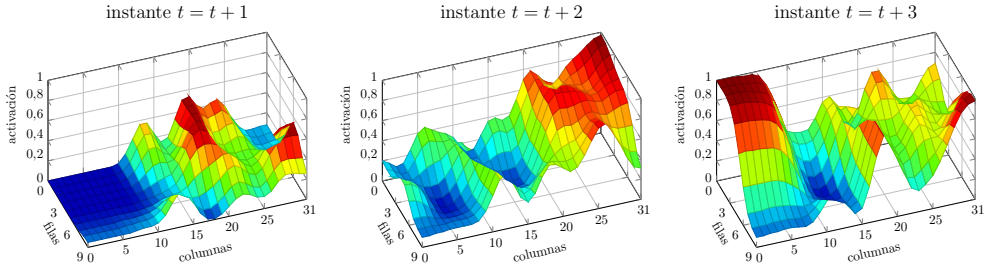


Figura 8.9: Activaciones promedio acumuladas para los árboles de predicciones durante la trayectoria del agente.

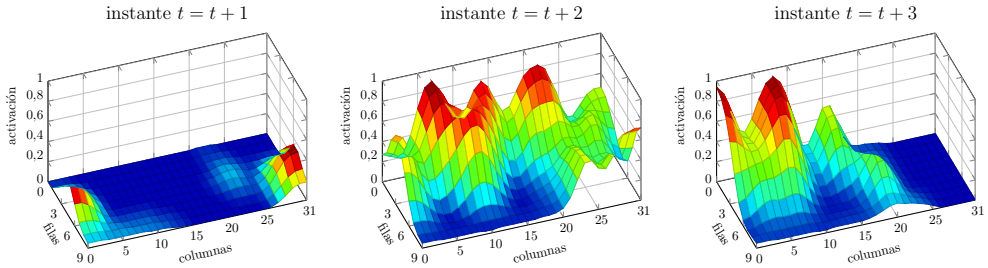
(a) Predicciones visuales para la rama 5 (a-a-a)

(b) Predicciones visuales para la rama 1 (d-d-d)

Figura 8.10: Imágenes de las predicciones visuales para las ramas 5 y 1 del árbol 1, (a) y (b) respectivamente.



(a) Predicciones táctiles para la rama 3 (d-a-a)



(b) Predicciones táctiles para la rama 9 (i-i-i)

Figura 8.13: Gráficas de las predicciones táctiles para las ramas 3 y 9 del árbol 3, (a) y (b) respectivamente.

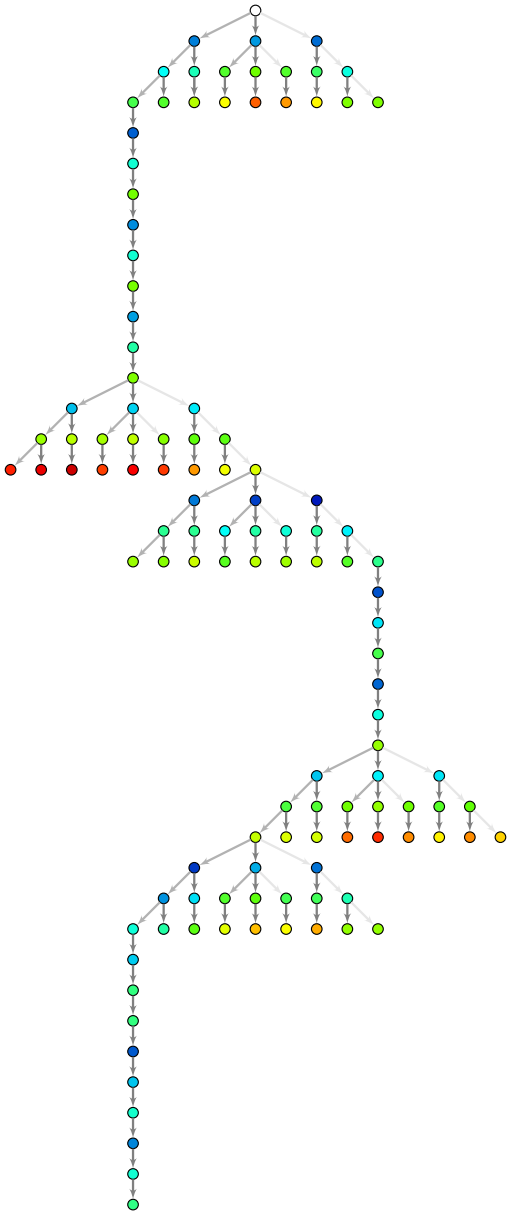


Figura 8.14: Árbol de las activaciones promedio acumuladas durante la navegación en el corredor.

Capítulo 9
Conclusiones

El trabajo presentado constituye un esfuerzo mas por estudiar modelos provenientes de las ciencias cognitivas a través de su implementación en agentes artificiales (Pfeifer (2002)). Todo esto, dentro del marco de trabajo de la cognición cimentada (Barsalou (2008)) la cual enfatiza el papel que desempeña el cuerpo, el entorno y los procesos motrices para la estructuración y el surgimiento de habilidades cognitivas en los agentes.

Este trabajo se originó a partir de diversos estudios acerca de la percepción de distancia en los humanos los cuales proponen que esta no es una capacidad generada a través de un proceso geométrico, sino que por el contrario es el resultado de la asociación multimodal (Braund (2007)), expresado en términos de unidades escaladas a nuestro cuerpo e influenciado por la acción (Proffitt (2006)).

El objetivo principal fué desarrollar un modelo capaz de dotar a un agente artificial con la capacidad de adquirir una noción de distancia a los objetos de su entorno. Este se basó en la implementación de un modelo directo, el cual provee de forma anticipada las consecuencias sensoriales de un comando motriz que se pretende llevar a cabo. El estado sensorial fué representado a través de las modalidades visual y táctil, mientras que el espacio motriz por tres movimientos diferentes, un movimiento hacia adelante y giros hacia la izquierda y derecha.

En los experimentos realizados, se encontró que las predicciones realizadas por el modelo directo para la modalidad táctil estuvieron en un rango continuo de [0-1], a pesar que durante la fase de entrenamiento esta fué codificada con un valor binario representando un estado de colisión o no colisión, este aspecto resultó ser una característica emergente del sistema que le permitió al agente realizar una asociación multimodal de la información visual y táctil percibida.

El modelo implementado le proporcionó al agente la capacidad para realizar un juicio perceptual de la distancia a un objeto, en función del número de movimientos que podría ejecutar antes de que ocurra una colisión.

En el primer experimento se coloco un único obstáculo frente al agente a diferentes distancias en intervalos de 15 cms. Para cada una de las distancias el agente llevo a cabo predicciones de largo plazo (PLP). El propósito de esto fué observar el valor presentado por las predicciones táctiles y caracterizar lo que llamamos un concepto de *distancia a colisión*. Este concepto esta cimentado en las capacidades sensorimotrices del agente y es consistente para PLPs de hasta 10 simulaciones.

Este valor umbral, es un aspecto que va acorde con uno de los principios de diseño de agentes artificiales completos que se propone en (Pfeifer and Scheier (1999)) y concerniente al principio del valor, el cual establece la existencia de alguna medida intrínseca que le indique al agente la conveniencia de experimentar una determinada situación sensorial.

Una vez cimentado este concepto se diseño un experimento buscando escalar la complejidad de los procesos cognitivos. La tarea del agente consistió en determinar de entre dos pasajes o entradas por cual de ellos podía pasar sin colisionar. Para lograr esto el agente utiliza lo que nos atrevemos a llamar un concepto de *pasabilidad*. Haciendo uso de PLPs, el agente encuentra el mejor camino de manera segura antes de ejecutar ningún movimiento explicito. Este experimento se relaciona al reportado en humanos por (Warren and Whang (1987)).

Para tal efecto se implementó un árbol de 11 PLPs con secuencias de 10 combinaciones distintas de los tres movimientos del repertorio motriz del agente. Se mostró como las ramas de PLP que indicaban una secuencia de movimientos hacia la apertura mas amplia, presentaron los valores menores de activación táctil, mientras que las demás ramas tuvieron valores mayores. La máxima activación correspondió a la secuencia de acciones que conducirían a una colisión con los obstáculos frente al agente. En seguida, se realizó una corrección de la trayectoria en plena ejecución de esta, gracias a la implementación de un árbol de PLP de 3 ramas y 3 movimientos cada una. Esto le permitió al agente anticipar una futura colisión con uno de los bordes de la apertura y terminar la tarea de manera exitosa.

Sin lugar a duda, el trabajo a futuro en términos de comportamientos mas complejos presenta retos importantes. Sin embargo, creemos que estos experimentos proveen bases solidas para la cimentación de conceptos y acciones en agentes artificiales, abriendo discusiones importantes sobre la toma de decisiones y la posibilidad de planificación. Todo esto dentro del marco de la robotica cognitiva, la interacción de los agentes con su ambiente y las representaciones y predicciones sensorimotrices.

Bibliografía

Alvarez, L., Deriche, R., Sánchez, J., and Weickert, J. (2002). Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13(1):3–21.

Arceo, D. C., Escobar, E., Hermosillo, J., and Lara, B. (2013). Modelado de un sistema de neuronas espejo en un agente autónomo artificial model of a mirror neuron system in an artificial autonomous agent. *Nova Scientia*, 5(10).

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., and Yoshida, C. (2009). Cognitive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development*, 1(1):12–34.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Physiology*, 59:617–645.

Blakemore, S.-J., Frith, C. D., and Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11(5):551–559.

Bower, T. (1966). The visual world of infants. *scientific American*, 215:80–92.

Braund, M. J. (2007). The indirect perception of distance: Interpretive complexities in berkeley's theory of vision. *Kritike*, 1:49–64.

Breazeal, C. and Scassellati, B. (2002). Robots that imitate humans. *Trends in Cognitive Sciences*, 6(11):481–487.

Bremner, J. G. (2003). Perception, knowledge and action. In Slater, A. and Bremner, J. G., editors, *An introduction to developmental psychology*. Blackwell Publishing.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, 6(1&2):3–15.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

Collins, B. M. and Kornhauser, A. L. (2006). Stereo vision for obstacle detection in autonomous navigation.

Daniel L. Schacter, D. R. A. and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8:657–661.

Dennett, D. C. (1993). Cognitive wheels : The frame problem of ai. In Hookaway, editor, *Minds, Machines and Evolution*, pages 129–151. Cambridge University Press.

DeSouza, G. N. and Kak, A. C. (2002). Vision for mobile robot navigation: A survey. *IEEE, TRANS. PAMI*, 24(2):237–267.

Driskell, J. E., Cooper, C., and Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, 79(4):481–492.

Escobar, E., Hermosillo, J., and Lara, B. (2012). Self body mapping in mobile robots using vision and forward models. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2012 IEEE Ninth*, pages 72–77.

Frith, C. D., Wolpert, D. M., et al. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404):1771–1788.

Gibson, E. J. and Walk, P. D. (1960). The visual cliff. *Scientific American*, 202:64–71.

Gibson, J. J. (1950). *The Perception of the Visual World*. Houghton Mifflin.

Gibson, J. J. (1979). *The Ecological Approach To Visual Perception*. Lawrence Erlbaum Associates, new edition edition.

Goldstein, E. B. (2008). *Cognitive psychology: Connecting mind, research and everyday experience*. Thomsom Wadsworth.

Goldstein, E. B. (2010). *Sensation and Perception*. Cengage Learning.

Goodman, R. and Holland, O. (2003). Robots with internal models: A route to machine consciousness? *Journal of Consciousness Studies*, 10:4–5.

Graffigna, J. P., Romero, L. E., and Romo, R. (2005). Evaluación de métodos para la obtención del mapa de disparidad en sistemas de visión estéreo.

Gregory, R. L. (1966). *Eye and Brain*. Weidenfeld & Nicolson, London.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27:377–342.

Harnad, S. (1990). The symbol grounding problem.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in computer vision*. Cambridge University Press.

Hasan, A. H. A., Hamzah, R. A., and Johar, M. H. (2010). Region of interest in disparity mapping for navigation of stereo vision autonomous guided vehicle. *International Journal of Computer and Electrical Engineering*, 2(2):1793–8163.

Hebb, D. O. (1949). *The Organization of Behavior A Neuropsychological Theory*. Lawrence Erlbaum Associates.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8(6):280–285.

Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of comparative and physiological psychology*, 56(5):872.

Hoffmann, H. (2007). Perception through visuomotor anticipation in a mobile robot. *Neural Networks*, 20:22–33.

Hoffmann, H. and Möller, R. (2004). Action selection and mental transformation based on a chain of forward models. In *The Eighth International Conference on the Simulation of Adaptive Behaviour*, pages 213–222. SAB.

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, 33(11):1419 – 1432. The Neuropsychology of Mental Imagery.

Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervise learning with a distal teacher. *Cognitive Science*, 16:307–354.

Julesz, B. (1971). Foundations of cyclopean perception.

Kiverstein, J. (2007). Could a robot have a subjective point of view? *Journal of Consciousness Studies*, 14:128–140.

Kolb, H. (2003). How the retina works. *American Scientist*, 91(1):28–35.

Kosslyn, S. M. (1994). *Image and Brain*. MIT Press.

Kosslyn, S. M., WL, T., and G., G. (2006). *The Case for Mental Imagery*. Oxford University Press.

Lappin, J. S., Shelton, A. L., and Rieser, J. J. (2006). Environmental context influences visually perceived distance. *Perception & Psychophysics*, 68(4):571–581.

Law, J., Lee, M., Hülse, M., and Tomassetti, A. (2011). The infant development timeline and its application to robot shaping. *Adaptive Behavior*, 19(5):335–358.

Lee, D. N. and Lishman, J. (1975). Visual proprioceptive control of stance. *Journal of Movement Studies*, 1(3):87–95.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, 15:151–190.

Martínez, M. M. (2010). *Técnicas de visión estereoscópica para determinar la estructura tridimensional de la escena*. PhD thesis, Universidad Complutense de Madrid.

Miall, R. C. and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, 9:1265–1279.

Möller, R. and Schenck, W. (2008). Bootstrapping cognition from behavior a computerized thought experiment. *Cognitive Science*, 32(3):504–542.

Moons, T. (1998). Lecture notes in computer science. In Koch, R. and Gool, L. V., editors, *3D Structure from Multiple Images of Large-Scale Environments*, volume 1506, chapter A Guided Tour Through Multiview Relations, pages 304–346. Springer Berlin / Heidelberg.

Murray, D. and Little, J. J. (2000). Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8:161–171.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19:113–126.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27.

Peters, M. W. and Sowmya, A. (1998). A real-time variable sampling technique: Diem. In *International Conference on Pattern Recognition*, Brisbane,Australia.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., Spivey, M. J., et al. (2012). Computational grounded cognition: a new alliance between grounded cognition and computational modeling. *Frontiers in psychology*, 3:612–612.

Pfeifer, R. (1996). Building "fungus eaters": Design principles of autonomous agents. In *From Animals to Animats 4*.

Pfeifer, R. (2002). Robots as cognitive tools. *International Journal of Cognition and Technology*, 1:125–143.

Pfeifer, R. and Scheier, C. (1997). Sensory-motor coordination: The metaphor and beyond. *Robotics and Autonomous Systems*, 20:157–178.

Pfeifer, R. and Scheier, C. (1999). *Undestanding Intelligence*. MIT Press, Cambridge, MA, USA.

Proffitt, D. R. (2006). Distance perception. *Current Directions in Psychological Science*, 15(3):131–135.

Regan, D. and Gray, R. (2000). Visually guided collision avoidance and collision achievement. *Trends in Cognitive Sciences*, 4(3):99–107.

Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *IEEE International conference on neural networks*, pages 586–591.

Rochat, P. (1989). Object manipulation and exploration in 2- to 5-month-old infants. *Developmental Psychology*, 25:871–884.

Salomon, R. (1998). Improving the dac architecture by using proprioceptive sensors. In *In [SAB98*.

Saxena, A., Schulte, J., and Ng, A. Y. (2007). Depth estimation using monocular and stereo cues. In *IJCAI*, volume 7.

Scheier, C. and Lambrinos, D. (1996). Categorization in a real-world agent using haptic exploration and active perception. *Proceedings of the 4th International Conference on Simulation of Adaptive Behavior (SAB'96)*, pages 65–75.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417–424.

Shepard, R. and Metzler, J. (1971). Mental rotation of three-dimensional objects.

Slater, A. (1989). Visual memory and perception in early infancy. *Infant development*, pages 43–71.

Slater, A., Mattock, A., and Brown, E. (1990). Size constancy at birth: Newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology*, 49(2):314–322.

Smith, D., Holmes, P., Whitemore, L., Collins, D., and Devonport, T. (2001). The effect of theoretically-based imagery scripts on field hockey performance. *Journal of sport behavior*, 24(4).

Sporns, O. and Edelman, G. M. (1993). Solving bernstein's problem: A proposal for the development of coordinated movement by selection. *Child Development*, 64:960–981.

Suzuki, M., Floreano, D., and Paolo, E. A. D. (2005). The contribution of active body movement to visual development in evolutionary robots. *Neural Networks*, 18(656-665).

Traver, V. J. and Bernardino, A. (2010). A review of log-polar imaging for visual perception in robotics. *Robotics and Autonomous Systems*, 58:378–398.

Tsai, R. Y. (1987). A versatile camera calibration technique for high-accuray 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Journal of Robotics and Automation*, 3(4).

Turvey, M. T. (2004). Space (and its perception): The first and final frontier. *Ecological Psychology*, 16(1):25–29.

Verschure, P. F. M. J., Kröse, B. J. A., and Pfeifer, R. (1992). Distributed adaptive control: The self-organization of structured behavior. *Robotics and Autonomous Systems*, 9(3):181–196.

Wang, Y., Wu, T., Orchad, G., Dudek, P., Rucci, M., and Shi, B. E. (2009). Hebbian learning of visually directed reaching by a robot arm. In *Biomedical Circuits and Systems Conference, 2009. BioCAS 2009. IEEE*, pages 205–208.

Warren, W. H. and Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, 13:371–383.

Wexler, M. and Boxtel, J. J. A. v. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences*, 9(9):431–438.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9:625–636.

Wolpert, D. M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):593–602.

Wolpert, D. M. and Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18):–729.

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11):487–494.

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882.

Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.*, 11(7-8):1317–1329.

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9):338–347.