UNIVERSIDAD NACIONAL DEL SUR

Tesis de Doctor en Control de Sistemas

Percepción y comprensión autónoma del área transitable

Marcelo Leandro Moreyra

Bahía Blanca Argentina

2013

UNIVERSIDAD NACIONAL DEL SUR

Tesis de Doctor en Control de Sistemas

Percepción y comprensión autónoma del área transitable

Marcelo Leandro Moreyra

Bahía Blanca Argentina

2013

A mis seres queridos.

Prefacio

Esta Tesis se presenta como parte de los requisitos para optar al grado Académico de Doctor en Control de Sistemas, de la Universidad Nacional del Sur y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otra. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el ámbito del Departamento de Ingeniería Eléctrica y Computadoras durante el período comprendido entre el 10 de Agosto de 2008 y el 30 de Mayo de 2013, bajo la dirección del Profesor Dr. Favio Román Masson.

Marcelo L. Moreyra

Bahía Blanca, Argentina
6 de Septiembre de 2013

Agradecimientos

Gracias a Dios tengo muchas personitas a las que quisiera agradecer su apoyo y compañía durante estos años. En primer lugar, a mi querida familia. A mis viejos, a quienes debo lo que soy y lo que tengo. A mi hermana Laura, mis abuelos, mis tíos. Gracias al amor que siempre me han brindado y sin el cual hoy no estaría escribiendo estas líneas.

A mi director, Favio, aunque sea bostero. Por su inagotable predisposición desde el primer momento y por la gran libertad que me dió para trabajar. Por la buena onda, los consejos y el apoyo constante.

A todos mis amigos. A esos locos incondicionales que quiero a montones. Agus, Robin, Mato, Kari, Maite, gracias por estar siempre. A Anggi, por su generosa y sincera entrega. A mi segunda familia, Edgar, Nieves, Franco, Rocío y Lau. Por darle sabor y alegría a tantos momentos. Tan buena gente que me regaló su amistad y compañía.

A todos mis compañeros del Laboratorio de Control y Robótica, que siempre me hicieron sentir parte importante. A los que están, y a los que estuvieron. No podía haberme tocado trabajar en un lugar mejor que éste. En particular a Franquito, el Ruso y Gustavito, por la complicidad en nuestras gloriosas mateadas. A Gerardo, por su desinteresada e importante ayuda con los experimentos.

Todos estas personas me regalaron su granito de arena para construir esta tesis y siempre les estaré agradecido.

Resumen

El diseño y desarrollo de vehículos inteligentes ha sido motivo de investigación durante más de tres décadas mostrando un enorme progreso en los últimos años. La existencia de proyectos a largo plazo impulsados por iniciativas gubernamentales en conjunto con grupos de investigación de la industria automotriz y de la academia, ha permitido que en algunos lugares del mundo los vehículos autónomos ya hayan demostrado con éxito que pueden circular por las calles de una ciudad.

Para que un vehículo de este tipo pueda interactuar en forma segura con otros vehículos conducidos por humanos es necesario que tenga la capacidad de percibir fielmente su entorno, identificando al resto de los participantes del tráfico y los lugares por donde es posible transitar. Actualmente, los proyectos más maduros se basan en modalidades de sensado aún demasiado costosas como para permitir que un producto de este tipo tenga un alcance masivo para la población. Siendo la visión el principal elemento de navegación que utilizan los humanos para conducir un vehículo, resulta algo sorprendente que las cámaras no sean aún protagonistas fundamentales de los actuales sistemas automáticos para la percepción del ambiente, más aún si se tienen en cuenta su bajo costo y su bajo requerimiento de energía para funcionar.

Uno de los problemas donde la visión sí ha permitido un gran avance es la detección del camino por el que puede transitar un vehículo. Para ésto se suele utilizar el conocimiento acerca de la apariencia y la forma geométrica del camino para proponer un modelo que se ajustará en función de las características extraídas de una imagen. Las técnicas modernas del filtrado estadístico son utilizadas para dar seguimiento al modelo a través de tiempo aumentando el rechazo al ruido y las mediciones erróneas, y reduciendo el costo computacional que implica calcular los parámetros. Estos enfoques han permitido alcanzar soluciones con alto grado de robustez ante los cambios climáticos y los cambios drásticos en la iluminación de la imagen. Sin embargo, estos sistemas fallan cuando la forma del camino cambia de una manera tal que el modelo considerado pierde validez. Para poder detectar automáticamente un cambio de este tipo hacen falta nuevas estrategias con un mayor poder de abstracción y que permitan una mayor comprensión de la escena.

Dada la enorme robustez del sistema visual humano y su eficiencia en la utilización de los recursos de procesamiento, resulta de primordial interés aprender acerca de cómo las personas resuelven este problema. Con este objetivo, esta tesis propone estudiar y analizar los patrones de atención visual de las personas cuando reconocen diferentes tipos de topologías como intersecciones, bifurcaciones y uniones de caminos, entre otras. A lo largo de los capítulos se introducen los fundamentos necesarios para comprender el tema abordado y se presentan resultados experimentales que dan soporte a las hipótesis planteadas. Las evidencias encontradas sentarán la base para el desarrollo de nuevos algoritmos para la detección automática de la topología del camino.

Abstract

The design and development of intelligent vehicles has been an active research area for more than 30 years, showing tremendous progress in the last few years. The existence of long-term projects promoted by government initiatives in conjunction with research groups of the automotive industry and academia, has allowed autonomous vehicles to show in some places of the world successful results while driving across urban scenarios.

To be possible for a vehicle of this kind to safely interact with other human-driven cars it is neccessary to have the ability to perceive accurately the environment, identifying all other participants of the traffic and detecting the drivable areas. Currently, the most mature projects are based on sensing modalities still too expensive to allow a product of this type to massively reach the common users. While vision is the primary navigation element that humans use to drive a vehicle, it remains quite surprising that cameras are not yet essential for the current environment perception automatic systems, even more taking into account their low cost and low power requirements for operation.

Road detection is one of the problems where vision has effectively had an important impact. The knowledge about the road appearance and its shape is usually considered to propose a road model that will be fitted according to the features extracted from the image. Modern statistical filtering techniques are used to track the model through time, increasing rejection to noise and erroneous measurements, and reducing the computational cost involved in estimating its parameters. These approaches have achieved solutions with a high degree of robustness to climate changes and drastic illumination variations in the image intensity. However, these systems fail to adapt when road's shape changes in a way that the considered model is no longer valid. New strategies with higher power of abstraction that allow greater understanding of the scene are needed to detect this type of changes.

Given the great robustness of human visual system and its efficient use of processing resources, it results of primary interest to learn about how people solve this kind of problem. To this end, this thesis proposes to study and analize people visual attention patterns when recognizing different types of topologies like road intersections, road splits and road junctions, among others. Throughout the chapters the basics needed to

understand the topic addressed are introduced and experimental results that support the hypotheses are presented. The evidence found will provide the foundations for the development of new algorithms for the automatic detection of the topology of the road.

Índice general

Prefacio I

Agradecimientos III

Resumen V

Abstract VII

1. Introducción 1

1.1. Principales contribuciones 4

1.2. Organización de la tesis 6

2. La percepción visual humana y la visión artificial 9

2.1. El sistema visual humano 9

2.1.1. La percepción de la profundidad y la visión binocular 14

2.1.2. La percepción de los colores 16

2.2. Los principios de la visión artificial 18

2.2.1. La cámara y las imágenes digitales 18

2.2.2. El modelo geométrico de la cámara 19

2.2.3. La representación de los colores 20

2.3. Resumen 23

3. El reconocimiento del terreno transitable 25

3.1. La percepción humana de lo transitable mediante visión 26

3.1.1. Los procesos del reconocimiento visual 26

3.1.2. Las características visuales del terreno transitable 27

3.2. Detección de caminos con imágenes: análisis bibliográfico comparativo . . 33

3.2.1. La detección del carril y la dependencia de la estructura 34

3.2.2. Los sistemas para la detección de caminos en ambientes poco estructurados 36

3.3. Limitaciones, desafíos y oportunidades 48

3.3.1. Problemáticas que se enfrentan 48
3.3.2. Topologías no lineales para representar un camino 55
3.3.3. Hacia una verdadera comprensión de la escena 57
3.4. Resumen 63

4. Patrones de atención visual para la percep. de caminos con topol. no lineales 65

4.1. La atención visual selectiva 65
4.2. Las hipótesis y los antecedentes del experimento 68
4.3. Metodología experimental utilizada 70
4.3.1. Participantes 70
4.3.2. Material visual utilizado 71
4.3.3. Configuración del experimento y el equipamiento 74
4.3.4. Procedimiento 74
4.4. Preproces. de las imágenes para el posterior análisis de las fijaciones . . . 75
4.4.1. Extracción del camino y sus bordes 76
4.4.2. Separación del camino en subregiones 77
4.5. Análisis de los resultados 83
4.5.1. Análisis general de las fijaciones: las estrategias de inspección . . . 83
4.5.2. La importancia de los bordes del camino 89
4.5.3. La influencia de la topología en la distribución de las fijaciones . . 93
4.5.4. Los lugares más visitados 96
4.5.5. El rol de la saliencia 99
4.6. La implicancia de los resultados 102
4.7. Resumen 104

5. Conclusiones y perspectivas a futuro 105

5.1. Trabajos futuros 108
5.2. Comentarios finales 110

A. Material fotográfico y patrones de fijación asociados 113

A.1. Imágenes para el entrenamiento 113
A.2. Imágenes para el experimento 114

Bibliografía **121**

Capítulo 1

Introducción

Las posibles mejoras en un tráfico vehicular cada vez más intenso y desordenado, y la disminución de la cantidad y la severidad de los accidentes de tránsito han motivado durante años el desarrollo de los *vehículos autónomos*. La aparición de este tipo de vehículos promete una reducción en los tiempos de traslado y en el consumo de combustible debido a un mayor y más eficiente flujo vehicular [Luettel et al. (2012)]. Esto tiene un impacto directo en términos de costo pero también en cuestiones como la calidad de vida y el cuidado del medio ambiente. Por otra parte, los tiempos de respuesta que exhibe un sistema automático impulsado por las nuevas tecnologías de sensado y control son considerablemente inferiores a los tiempos normales de reacción de un conductor, aumentando notablemente las probabilidades de evitar un accidente o al menos de reducir su gravedad.

Desde hace décadas algunos grupos de la industria automotriz han trabajado en conjunto con universidades y centros de investigación en lo que conceptualmente se conoce como *sistemas de transporte inteligentes*, impulsados por iniciativas gubernamentales esencialmente de Europa y Estados Unidos. Muchos de los avances tecnológicos obtenidos ya se han convertido hoy en productos comerciales y se pueden encontrar por ejemplo en los *sistemas de asistencia al conductor* que acompañan a determinados vehículos de alta gama. Estos proyectos a largo plazo se han enfocado principalmente en la navegación por autopistas reportando algunos de ellos muchos kilómetros recorridos de forma autónoma [Dickmanns & Zapp (1988), Pomerleau (1995)]. En los últimos años también se han presentado nuevas experiencias en condiciones y con recorridos aún más desafiantes [Wille et al. (2010), Cerri et al. (2011), Bertozzi et al. (2011)].

En los años 2004 y 2005 el Departamento de Defensa de los Estados Unidos (DARPA, por las siglas en inglés de *Defense Advanced Research Projects Agency*) impulsó una competencia de vehículos autónomos denominada DARPA *Grand Challenge* [Iagnemma & Buehler (2006a), Iagnemma & Buehler (2006b)]. Este desafío consistía en que cada

vehículo pudiera realizar en forma totalmente autónoma un recorrido a través de caminos desérticos siguiendo determinados puntos marcados mediante un sistema de posicionamiento global, bien conocido por sus siglas en inglés *GPS*. En el año 2007 se subió la apuesta y se realizó el DARPA *Urban Challenge* en el que los vehículos debían navegar autónomamente en ambientes urbanos a la par de otros vehículos conducidos por humanos, respetando en todo momento las leyes de tránsito del lugar [Urmson et al. (2008), Montemerlo et al. (2008)]. Estas iniciativas impulsaron fuertemente a los investigadores hacia el desarrollo de sistemas autónomos para la navegación en ambientes pocos estructurados y también en ambientes urbanos, que presentan una complejidad mucho mayor respecto al tránsito por rutas y autopistas. El vehículo autónomo desarrollado por la empresa Google [Markoff (2010), Markoff (2011)] a partir de un Toyota Prius es un claro ejemplo de la continuidad de los buenos resultados obtenidos en los desafíos planteados por DARPA.

Al proyecto *Google driverless car* se suman además muchos otros, entre los que se encuentran el *Shelley* de la universidad de Stanford con un Audi TTS [Funke et al. (2012)], el *RobotCar UK* de la universidad de Oxford con un Nissan Leaf [Ackerman (2013)], el *Lexus advanced active safety research vehicle* de Toyota Motor Corporation con un Toyota Lexus [Guizzo (2013)] y el *proyecto SARTRE* en el que Volvo Car Corporation es uno de los principales socios y en el cual se impulsan las formaciones al estilo tren [Robinson et al. (2010)]. Estos proyectos muestran diferentes niveles de autonomía manteniéndose el auto de Google como el líder en lo que respecta al comportamiento autónomo, demostrando múltiples resultados experimentales en situaciones reales de manejo.

Los desafíos que debe enfrentar un vehículo autónomo en lo que respecta a la percepción, la navegación y el control son realmente complejos. La *percepción del ambiente* trata acerca de como un sistema es capaz de entender el mundo que lo rodea a partir de un número finito de sensores, y es quizás el problema que hoy ofrece mayores oportunidades para la innovación. Asimismo, las soluciones que se han propuesto tanto para la navegación como para el control han alcanzado en comparación un nivel de madurez superior. El sistema de percepción de un vehículo autónomo debe ser capaz, entre otras cosas, de: posibilitar la estimación del estado del vehículo y su propio movimiento, detectar y hacer un seguimiento de objetos en movimiento, es decir, de los demás

participantes del tráfico, detectar objetos estáticos como obstáculos y elementos de estructura, estimar la forma o la geometría del camino, y localizarse dentro de un mapa [Luettel et al. (2012)]. Teniendo en cuenta que la completa autonomía de un vehículo es el problema más complejo que se enfrenta, y que ya se han mostrado soluciones para ésto (Google por ejemplo), se podría tener la impresión de que todos los problemas intermedios o de menor complejidad ya estarían resueltos. Sin embargo, varios de estos problemas continúan aún sin soluciones que ofrezcan la robustez necesaria para asegurar la funcionalidad del vehículo ante las diferentes condiciones que comúnmente enfrenta un conductor. El hecho además de que en los desafíos de DARPA no había restricción de costos en la utilización de sensores y que se contaba con la información provista por un mapa digital altamente preciso [Montemerlo et al. (2008)], influyó en que los esfuerzos decantaran en el desarrollo de sistemas con capacidades propias de percepción muy limitadas [Bar Hillel et al. (2012)].

Las modalidades de sensado utilizadas se basaban principalmente en sensores LIDAR[1], un GPS, y en algunos casos sistemas de visión complementarios. Los principales proyectos de hoy en día, como el de Google o el de Toyota, no difieren demasiado de este esquema y presentan un sensor LIDAR 3D de alta gama ubicado en el techo, varios radares a los costados, GPS y unidades de medición inercial (IMU, por las siglás en inglés). Vale la pena aclarar que en todos estos casos hay una gran dependencia de la existencia de mapas del ambiente por donde se transita. Los costos de estas iniciativas son aún muy elevados como para poder generar productos que sean comercialmente viables.

Por otra parte, el recientemente presentado RobotCar UK de Oxford propone una alternativa diferente. A diferencia de los proyectos anteriores, este vehículo no depende de un costoso sensor LIDAR ni de la imprecisión en la posición de un GPS, sino que se basa en el reconocimiento de la escena mediante cámaras de video y sensores de rango láser, bajando considerablemente el costo del equipamiento que se agrega al vehículo. Las imágenes generadas se comparan con una base de datos preexistente para determinar la posición y el nuevo destino del vehículo, además de la existencia de obstáculos. Esta base de datos y mapas se genera a partir del entrenamiento del sistema por los lugares por donde se pretende que el vehículo circule por sí solo. Esto por supuesto limita el rango de acción del vehículo a los lugares conocidos, aunque en un futuro no muy lejano se podría

[1] Acrónimo del inglés *Light Detection and Ranging* o *Laser Imaging Detection and Ranging*

pensar en la utilización de mapas generados por otros vehículos y que se comparten por algún tipo de red de datos.

Dado que una imagen provee gran cantidad de información del ambiente y a muy bajo costo, la cámara pareciera ser un sensor ideal para un vehículo autónomo. Teniendo en cuenta además que los humanos se basan principalmente en la visión para conducir un vehículo, resulta quizás algo paradójico que de los sistemas autónomos más modernos solo uno la utilice como modalidad de sensado primaria, y el resto aún la considere como un complemento. Sin embargo, la alta dimensionalidad de los datos de una imagen exige algoritmos para procesarla que tengan gran capacidad de abstracción y que sean computacionalmente eficientes.

Uno de los problemas de percepción que se han mencionado como parcialmente resueltos es la estimación de la forma o la geometría del camino, y es justamente donde la visión ha sido protagonista permitiendo un gran avance durante los últimos 20 años [Desouza & Kak (2002), Bar Hillel et al. (2012)]. Sin embargo, la capacidad que tienen los sistemas actuales para detectar por si mismo un camino o los posibles caminos presentes en una escena es todavía limitada. En general los enfoques basados en visión suponen un modelo que considera características de apariencia y/o forma del camino, que no siempre logra adaptarse cuando la situación que se presenta cambia y difiere de aquella considerada para definir el modelo.

En particular, el problema que aquí interesa estudiar se relaciona con los cambios en la geometría o la topología del camino que son muy comunes por ejemplo en las intersecciones, las rotondas y las bifurcaciones. La detección de estos cambios a partir de una imagen exige nuevas estrategias y algoritmos que permitan al sistema una mayor comprensión de la escena. Es en este punto donde la visión humana juega nuevamente un papel inspirador para el desarrollo de dichas estrategias. Dado que las personas son perfectamente capaces de reconocer los diferentes tipos de topologías de un camino mientras conducen un vehículo, lo que se propone en esta tesis es estudiar cómo es que lo resuelven y que tipo de estímulos visuales utilizan para ello.

1.1. Principales contribuciones

Los aportes fundamentales de esta tesis se pueden resumir en los siguientes puntos:

- Se presenta un análisis comparativo, exhaustivo y original acerca de los métodos más relevantes que existen para la detección de caminos mediante visión. A partir de ésto es posible reconocer los principales elementos que suelen constituir este tipo de sistemas y cuales son las ventajas y desventajas de cada uno.

- Se identifican claramente las principales problemáticas que deben enfrentar comúnmente los sistemas basados en visión y las limitaciones que derivan en general de los cambios drásticos de escenarios, y las condiciones climáticas y de iluminación. Luego se evalúan y valoran cada uno de los enfoques para la detección de caminos según su comportamiento y robustez frente a estas problemáticas.

- Se determina la necesidad de nuevas estrategias que permitan inferir a partir de la imagen cuándo el modelo geométrico supuesto pierde validez o cuándo las características que se extraen para su ajuste ya no son confiables. Esto es un desafío fundamental para determinar la transitabilidad de un área. Se propone la inclusión de algoritmos con dichas funciones como otra parte constitutiva más de un sistema genérico para la detección de caminos mediante visión.

- Se reconocen además cuáles son las oportunidades de innovación y las estrategias que permitirán lograr la solución completa al problema de detección de caminos.

- Se diseña un experimento sin precedentes en el área que permite establecer las bases de una estrategia para la detección de caminos. Se utiliza para esto un equipo para el seguimiento de los ojos. Así se registró el comportamiento de una persona a través de la actividad de sus ojos cuando trata de identificar la topología de un camino.

- Se presenta un análisis profundo e intensivo de los patrones de atención visual registrados durante el experimento. A través de éste se demuestra la existencia de ciertos lugares del camino que son más relevantes que otros al momento de identificar la topología correspondiente. Se muestra que dichos patrones son influenciados por el tipo de topología presente en la imagen, y que la tarea visual propuesta depende principalmente de procesos cognitivos relacionados con el conocimiento del problema y no de los detalles de alto contraste o que sobresalen de la imagen.

1.2. Organización de la tesis

Esta tesis está organizada en cinco capítulos incluyendo ésta introducción. El Capítulo 2 introduce conceptualmente los sistemas de percepción visual de las personas y de las máquinas, mostrando las analogías que existen entre ellos. En primer lugar, se describe como funciona el sistema visual humano y cuales son las partes que lo componen. Se explican los mecanismos básicos involucrados en la generación de las imágenes y en la percepción de los colores a través de los ojos. Luego, se presenta a la cámara como elemento principal de los sistemas de visión artificial y se describe básicamente como se generan las imágenes digitales. Por último, se analizan las diferentes maneras de representar matemáticamente los colores y como influye la tecnología en su elección.

En el Capítulo 3 se analiza como las personas reconocen visualmente el terreno por el que pueden transitar y como han evolucionado los sistemas para la detección automática de caminos a partir de una imagen. Se describen primero los principales procesos que se dan en el cerebro cuando se identifica visualmente un objeto, para luego presentar cuales son las características visuales que el inconsciente colectivo utiliza para clasificar un terreno como transitable o no. A continuación se realiza un profundo análisis de los sistemas existentes para la detección de caminos en base a imágenes monoculares. Se reconocen los principales elementos que componen un sistema de este tipo y se evalúan las ventajas y desventajas que tiene cada enfoque. Finalmente, se estudian los problemas que tienen los sistemas actuales para enfrentar condiciones reales de uso, y se identifican nuevas oportunidades para la innovación que motivan los esfuerzos de esta tesis.

Dadas las limitaciones que presentan los sistemas modernos cuando se enfrentan a cambios en la topología del camino, este estudio procura aprender acerca de como las personas identifican diferentes tipos de topologías a partir de una imagen. Con este propósito entonces, el Capítulo 4 presenta un estudio experimental de los patrones de atención visual de las personas cuando resuelven esta clase de problema. Inicialmente el Capítulo define el concepto de atención visual y como se refleja en el movimiento de los ojos. Después se analizan los antecedentes de experimentaciones relacionadas que existen en la literatura y se exponen las hipótesis planteadas para el experimento realizado. Se describen en detalle el procedimiento, los materiales, los participantes y el equipo utilizado para las actividades experimentales. Luego se reporta un análisis completo de todos los resultados obtenidos.

Por último, en el Capítulo 5 se comparten las conclusiones finales de la tesis y se delinean los trabajos a futuro. Se analizan y ponen en contexto los resultados obtenidos y se especifican brevemente los pasos a seguir respecto de nuevas experimentaciones y del diseño de un algoritmo para la detección automática de la topología mediante visión monocular. Para finalizar se comenta acerca del rol actual y futuro de la visión en el desarrollo de los vehículos autónomos.

Capítulo 2

La percepción visual humana y la visión artificial

La visión es el principal elemento de navegación y comprensión del ambiente que tienen las personas. Prácticamente la totalidad de sus actividades cotidianas involucra algún tipo de reconocimiento visual de los elementos que las rodean. Con la aparición de las cámaras fotográficas y las cámaras de video ha sido posible el desarrollo de sistemas artificiales basados en la visión para diferentes tipos de aplicaciones como la inteligencia artificial, la robótica y los vehículos inteligentes. A lo largo de los años, muchos de estos sistemas han sido inspirados en la visión humana y de ciertos animales por su enorme capacidad de procesamiento de la información y su eficiencia en la utilización de los recursos. Por eso, resulta imprescindible tener un conocimiento básico acerca de los mecanismos biológicos y artificiales involucrados en la generación de imágenes.

El capítulo se organiza de la siguiente manera: en la sección 2.1 se detalla como está compuesto el sistema visual humano y cual es el mecanismo que permite a las personas percibir los colores y la profundidad a partir de los estímulos de luz que ingresan al ojo; mientras que en la sección 2.2 se describe el principio de funcionamiento de la cámara como principal elemento de los sistemas de visión artificial, se repasa brevemente el modelo geométrico considerado, y se definen los principales espacios de color utilizados para representar el color en las imágenes.

2.1. El sistema visual humano

El ojo es el elemento principal del sistema visual humano (Fig. 2.1). Entre las partes más importantes del ojo se encuentran la córnea, la pupila, el cristalino y la retina. La *córnea* es la membrana externa transparente que protege la superficie anterior del ojo. La *pupila* es la abertura central del iris a partir de la cual se controla la cantidad de luz que ingresa al sistema. El *cristalino* cumple la función de lente óptico y está compuesto

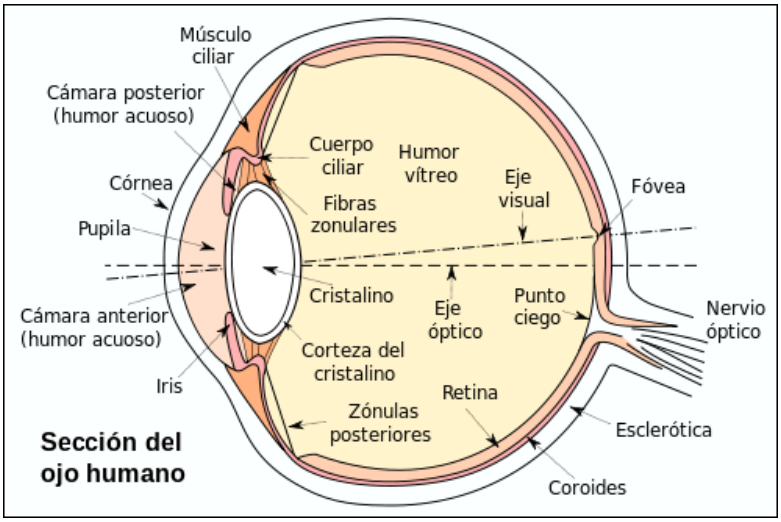


Figura 2.1: Ilustración del ojo humano y sus distintas partes.

por capas concéntricas de células fibrosas y un alto porcentaje de agua. A través de los músculos y las fibras ciliares se controla la apertura focal del lente permitiendo un enfoque a mayor o menor distancia. Los haces de luz provenientes del ambiente se proyectan a través del cristalino hacia la capa interna posterior del ojo llamada *retina*, donde a través de procesos electroquímicos la información de la imagen que allí se forma es transmitida por el nervio óptico hacia el cerebro mediante impulsos eléctricos.

La retina se organiza en diez capas interconectadas por distintos tipos de células, incluyendo células pigmentadas, neuronas y células de sostén. La luz que entra por la pupila debe atravesar todas las capas celulares hasta llegar a las capas más profundas de la retina donde se encuentran las células fotoreceptoras. Estos millones de diminutas células son sensibles a la luz y se dividen en dos grupos: los *conos* y los *bastones*. Los bastones conforman aproximadamente más del 95 % de los fotoreceptores que tiene la retina y son muy sensibles a la luz, a tal punto que podrían detectar un solo fotón [Roorda (2002)]. Son los responsables de que sea posible ver con luz tenue o casi en la oscuridad (*visión escotópica*). Por otro lado, los conos son menos sensibles y se acumulan densamente en una zona llamada mácula que contiene a la *fóvea*, hendidura circular de unos 1.5mm de diámetro. Éstos se clasifican en tres tipos distintos, siendo cada uno sensible a diferentes porciones de la luz visible, como se verá más adelante. La combinación de las señales de estos tres tipos de receptores permiten al ojo percibir los

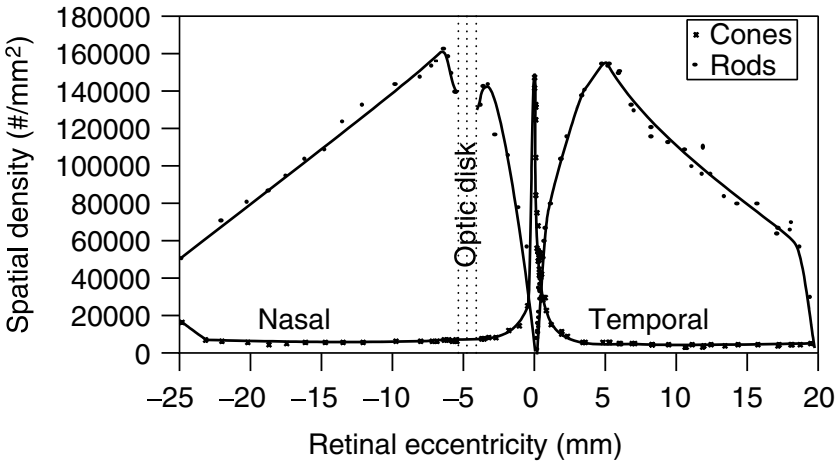


Figura 2.2: Densidad espacial de fotoreceptores (conos y bastones) en la retina del ojo. La imagen proyectada en la retina es muestreada en colores y con mayor resolución en la fóvea, donde existe una gran densidad de células cono. La región nasal de la retina se corresponde con la región más cercana a la nariz mientras que la región temporal de la retina es aquella más alejada. La imagen original pertenece a [Roorda (2002)] y aquí se reproduce en su idioma de origen.

colores cuando la intensidad de la luz es suficientemente alta (*visión fotópica*).

En la Fig. 2.2 se muestra la densidad espacial de conos y bastones en la retina en función de la excentricidad, que se define como el ángulo respecto al eje visual que pasa por la fóvea. Los ángulos negativos se corresponden con la región nasal de la retina, es decir la región más interna y que se encuentra más próxima a la nariz. La región más externa o más alejada de la nariz se denomina temporal. La figura ilustra aquí una gran no uniformidad en la distribución de fotoreceptores, donde los conos solo existen en la fóvea y los bastones dominan rápidamente fuera de esa área. El campo visual puede dividirse entonces en tres regiones: *foveal* (los 2° centrales), *parafoveal* (2° a 5° del centro) y *periférico* (mas de 5°), según [Rayner (1998)]. Los receptores alrededor de la fóvea son responsables de la visión espacial (imágenes estáticas) mientras que los receptores de la periferia permiten detectar el movimiento.

La densidad de conos en la fóvea (aproximadamente 199.000 conos por mm^2) permitiría resolver frecuencias espaciales de hasta 70 ciclos por grado (tomando en cuenta el criterio de Nyquist). Además, para cada fotoreceptor de la fóvea existe una conexión nerviosa dando lugar a la densidad de muestreo más alta en la retina. En las regiones

más alejadas de la fóvea la densidad de muestreo cae abruptamente debido a un aumento en el tamaño de los fotoreceptores y una disminución en la cantidad de conexiones nerviosas a los mismos. Según los datos experimentales de [Westheimer (1987)], a una distancia de solo $\frac{1}{12}^\circ$ respecto del eje visual la pérdida de resolución o *agudeza visual* es medible, mientras que para $\frac{1}{6}^\circ$ la pérdida ya es del 25 %, y para 1° supera el 40 %. Esta falta de resolución se compensa con el movimiento de los músculos del ojo (y la cabeza) que permiten al sistema alinear aquellos objetos de interés directamente hacia la fóvea. El mecanismo a través del cual el cerebro determina que parte del campo visual es la de mayor interés se denomina *atención selectiva*, y es válido no solo para la visión sino también para los otros tres sentidos.

La frecuencia espacial de los estímulos también tiene influencia sobre la sensibilidad al contraste. El *contraste* se define como la diferencia relativa en intensidad entre un punto u objeto y sus alrededores. La respuesta del ojo es logarítmica por naturaleza y se revela con características pasabanda ante estímulos sinusoidales [Gonzalez & Woods (2001)]. Los conos y bastones, que responden de una forma no lineal a las variaciones de intensidad, se combinan en forma ponderada con muchos otros receptores vecinos para generar las señales neuronales. Algunos de estos receptores en realidad ejercen una influencia inhibidora (o negativa) sobre la respuesta nerviosa. Este proceso se conoce como *inhibición lateral* y es el mayor responsable de dicha respuesta en frecuencia.

Retomando la estructura de capas de la retina, las células fotoreceptoras se conectan al nervio óptico a través de otros tipos de neuronas como las células *bipolares*, las células *amacrinas* y las células *horizontales*, hasta llegar a las células *ganglionares*. Estas últimas varían en su tamaño, sus conexiones y su respuesta sensorial ante el color, las formas y la profundidad, y conforman el nervio que finalmente conecta a la retina con el cerebro. Los nervios provenientes de ambos ojos pasan a través del *quiasma óptico* para llegar a cada hemisferio del cerebro mediante dos caminos: la vía colicular hacia el *colículo superior* (CS), que se encarga principalmente de asistir en control del movimiento de los ojos; y la vía retino-genicular que conduce al *núcleo geniculado lateral* (NGL), que es por donde se transmite aproximadamente un 90 % de la información visual [Frintrop et al. (2010)]. Desde el NGL la información se transfiere a la *corteza visual primaria* (V1). En esta vía primaria hacia V1 se realizan ya algunos procesamientos visuales simples, sin dejar de considerar a la retina que, tal como se mencionó antes, posee determinadas células que

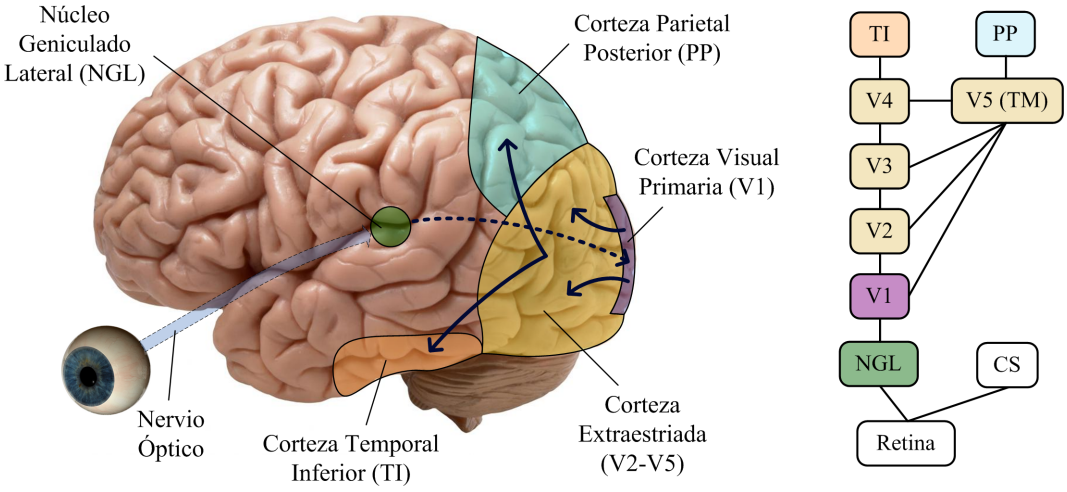


Figura 2.3: Las regiones que componen el sistema visual humano y las conexiones entre ellas. La información visual de la retina viaja por el nervio óptico hacia regiones superiores del cerebro como la corteza temporal inferior (TI) y la corteza parietal posterior (PP), pasando por la corteza visual primaria (V1) y las intermedias (V2-V5). El colículo superior (CS) no se ilustra en el diagrama de la izquierda ya que no es visible para esa vista del cerebro.

responden al contraste de colores y orientaciones. A lo largo de esta vía las células se vuelven más complejas y aumentan su capacidad de procesamiento, combinando además los resultados de células previas. Desde V1 la información es transmitida a regiones superiores del cerebro como las áreas de la *corteza extraestriada* (V2-V5), donde V5 (o TM) se conoce como el *area temporal media*, la *corteza temporal inferior* (TI), y la *corteza parietal posterior* (PP). A grandes razgos, el procesamiento del color y las formas conduce al TI, que es la región donde se realiza el reconocimiento de objetos. En cambio, el procesamiento del movimiento y la profundidad se realiza en el PP. Todas las conexiones y regiones mencionadas del sistema visual humano se encuentran ilustradas en los diagramas de la Fig. 2.3.

Además de las conexiones nerviosas desde la retina hacia cortezas superiores también se conoce que existen muchos circuitos de realimentación desde el cerebro hacia la periferia [Bar (2003), Li et al. (2004), Sigman et al. (2005)]. De esta manera, los sensores periféricos pueden adaptar sus funciones para acompañar procesos cognitivos de más alto nivel. Por ejemplo, si alguien estuviera manejando un vehículo y al llegar a una intersección quisiera saber si existe o no un semáforo, el proceso involuntario de

búsqueda visual del semáforo haría énfasis en detectar estímulos de color amarillo, verde o rojo, dando prioridad a las conexiones nerviosas que tengan estas funciones. En este caso, aquellos estímulos visuales que sean muy distintos a los mencionados podrían hasta pasar desapercibidos.

Resumiendo, el sistema visual humano tiene una enorme capacidad de procesamiento paralelo en sus primeras etapas, principalmente en la retina, el NGL, el CS y la corteza V1. Cerca de la periferia sensorial los campos receptivos se caracterizan por un alto grado de especifidad espacial permitiendo un procesamiento masivamente paralelo de las diferentes partes del campo visual [Niebur & Koch (1998)]. Estas neuronas son las encargadas de extraer características elementales de los estímulos como por ejemplo bordes con determinadas orientaciones. Las células que se encuentran más arriba en la jerarquía cortical tienen generalmente un mayor tamaño por lo que van perdiendo su especificad en el espacio. En la TI las células son más selectivas respecto a las características de los estímulos, es decir que se activan con estímulos visuales más complejos como rostros, manos, objetos, que puedan estar en cualquier lugar del campo visual. Este tipo de procesamiento o reconocimiento más avanzado ya no es paralelo sino más bien secuencial, ya que sólo puede procesarse un estímulo de este tipo a la vez.

2.1.1. La percepción de la profundidad y la visión binocular

El sistema visual tiene la capacidad también de estimar la distancia a los objetos o su profundidad. Para esto suele utilizar diferentes tipos de indicadores visuales que le permiten percibir de algún modo las tres dimensiones del espacio. Estos indicadores se podrían clasificar en aquellos *monoculares*, es decir, los que se pueden representar con dos dimensiones y se pueden observar con un solo ojo, y los *binoculares o estéreo*, que resultan de integrar la información proveniente de ambos ojos.

Según [Howard (2012)], las fuentes de información visual del tipo monocular que generalmente se utilizan para percibir la profundidad son múltiples, y se pueden dividir en las estáticas y las dinámicas. Las estáticas incluyen la perspectiva, la interposición, los tamaños relativos, la iluminación, los efectos del aire y el enfoque, entre otras. La perspectiva tiene que ver con que la textura o las líneas paralelas que convergen a un punto en la distancia, por lo que pueden utilizarse por ejemplo para reconstruir la distancia relativa entre dos partes de un objeto. También es posible estimar cuan

lejano está un objeto en función de su cercanía a la línea de horizonte. La interposición se refiere principalmente a la oclusión de objetos por otros más cercanos que permite establecer algún tipo de ordenamiento de los objetos respecto a las distancias relativas al observador. Por otro lado, cuando se conoce que dos objetos tienen el mismo tamaño, el tamaño relativo con que se observan ayuda a comprender cual está más cerca de los dos. De igual manera, si un objeto tiene un tamaño que resulta familiar y conocido, como por ejemplo un vehículo, se puede estimar una distancia aproximada a partir del ángulo visual que ocupa su imagen en la retina. Respecto de la iluminación, la forma en que la luz se refleja en los objetos y las sombras correspondientes son indicadores efectivos a la hora de determinar la forma y la posición de los objetos en el espacio. Los efectos de dispersión de la luz en la atmósfera pueden alterar la apariencia y el contraste de objetos que se encuentran lejanos, posibilitando diferenciarlos de objetos más cercanos y cuyo contraste es mayor. Por último, el grado de desenfoque de una imagen y la falta de resolución o detalle de la textura son comúnmente signos de lejanía. Los indicadores dinámicos en cambio, se basan justamente en el movimiento del observador y de los objetos. Cuando el observador se mueve, el movimiento relativo aparente de diversos objetos estáticos respecto del fondo da indicios de sus distancias relativas. Además es posible determinar las distancias a partir de la dinámica de los cambios en los tamaños de los objetos. Esto es, a medida que el tamaño de un objeto se vuelve más pequeño se infiere que éste se está alejando, y viceversa.

La percepción de la profundidad a partir de información binocular se basa principalmente en la convergencia y en la estereopsis. En el primer caso, cuando los ojos se enfocan en un objeto se dice que convergen. Al contraerse los músculos del ojo, el cerebro utiliza las sensaciones del tipo kinestésicas para percibir la distancia al punto de enfoque. La estereopsis en cambio, es la percepción de las tres dimensiones a partir de la disparidad entre las dos imágenes retinales que proveen cada uno de los ojos. La disparidad binocular es básicamente la diferencia entre la ubicación del objeto en la imagen izquierda y su ubicación en la derecha, que resulta de la separación horizontal entre ambos ojos (paralaje). Debido a que esta separación es muy pequeña, la estereopsis no sirve para distancias mayores a unos 15 metros, en cuyo caso la visión pasa a ser efectivamente monocular [Gregory (1965)]. Además, la estereopsis usualmente no está presente cuando se observa una escena con un solo ojo, o cuando se mira una fotografía con ambos ojos,

o para alguien que padece estrabismo. De todos modos, en los tres casos las personas pueden percibir ciertas relaciones de profundidad.

Además de las capacidades de percepción mencionadas anteriormente, la visión estéreo permite por un lado mejorar la habilidad para detectar objetos tenues mediante mecanismos de suma o refuerzo binocular, y por otro lado ampliar el campo de visión. Los humanos tienen un campo visual máximo de aproximadamente unos 200 grados con ambos ojos, incluyendo un campo binocular central de 120 grados flanqueado por dos campos de aproximadamente 40 grados en los que solo un ojo puede ver [Henson (1993)]. La visión binocular posibilita además que las personas puedan explorar su futuro inmediato controlando estratégicamente su comportamiento, anticipándose al peligro, y no actuando a base a meros reflejos [Gregory (1965)].

2.1.2. La percepción de los colores

Los células cono de la capa fotoreceptora de la retina se clasifican en tres tipos según las longitudes de onda de la luz a las que son sensibles. Esta selectividad en el espectro de frecuencias se explica por la existencia de unas sustancias (proteínas) llamadas *opsinas*. Cada cono contiene uno de tres tipos de opsinas: la *eritropsina*, que tiene mayor sensibilidad para las longitudes de onda largas (luz roja); la *cloropsina*, que es más sensible a las longitudes de onda medias (luz verde); o la *cianopsina*, que es mayormente sensible a las ondas visibles cortas (luz azul). Los bastones, en cambio, contienen *rodopsina*, que es sensible a frecuencias cercanas a la luz verde azulada, y por eso son los responsables de la visión escotópica, mencionada en la sección anterior.

Si bien generalmente se habla de los "canales" rojo, verde y azul (o R, G, B, según sus iniciales en inglés), en realidad los conos son sensibles a un amplio rango de frecuencias y sus pigmentos tienen curvas de absorción espectral que se solapan entre sí y cuyos máximos se ubican alrededor de 564nm, 534nm y 420nm respectivamente [Gonzalez & Woods (2001)]. Además, el canal B es relativamente mucho menos sensible que el R o el G, y su ancho de banda también es menor que los otros dos.

La interacción de las respuestas espectrales de cada fotoreceptor con la distribución espectral de la luz que ingresa al ojo deriva en lo que el común de las personas conoce como *color*. El color es una propiedad visual que nos permite percibir la diferencia en la composición espectral de la luz proveniente de diferentes objetos. También es posible

asociar físicamente a los colores con determinados materiales, objetos o fuentes de luz basándose en sus propiedades físicas de emisión, absorción y reflexión de la luz (Forsyth & Ponce (2002)). En el caso del sistema visual humano, la combinación de los tres canales *RGB* es la que permite percibir el color, dando lugar a lo que se conoce como *visión tricromática*. Por lo tanto, no importa cuan compleja sea la composición de longitudes de onda de la luz, el ojo la reduce a estas tres componentes.

La teoría tricromática de colores asume que prácticamente cualquier color puede representarse combinando en forma ponderada los colores rojo, verde y azul, también llamados *colores primarios*. Esta representación es justamente muy útil para explicar la captación o la visualización de los colores pero es algo pobre para explicar como realmente los humanos perciben el color [Ballard & Brown (1982)]. El brillo, el tono y la saturación son atributos comúnmente utilizados para describir mejor la sensación ante la luz. El *brillo* permite percibir los distintos niveles de grises independientemente de la cromaticidad de la luz. La intensidad del espectro de la luz no es una medida directa del brillo, pues existen ejemplos de superficies con intensidad espectral uniforme que sin embargo no se perciben con una brillosidad uniforme [Gonzalez & Woods (2001)]. De igual manera, el ojo puede percibir dos tonalidades como idénticas a pesar de que las distribuciones espectrales sean diferentes. El *tono* entonces es el atributo a partir del cual es posible juzgar cuan parecido es un estímulo respecto a aquellos que se describen como rojo, verde, azul y amarillo. Por último, la *saturación* hace posible diferenciar los tonos pasteles de aquellos más vivos, y puede considerarse como la cantidad de luz blanca que posee una fuente.

Con respecto de la transmisión de la información del color hacia el cerebro, hay teorías que proponen que el procesamiento del color se organiza de manera diferente a los mecanismos de la retina. La teoría de los *canales oponentes* propone que a partir de las respuestas de los conos se construyen tres canales: rojo-verde (R-G), azul-amarillo (B-Y), y blanco-negro (o *luminancia*). Permite explicar por ejemplo porque no es posible percibir algunas combinaciones de colores como un “verde rojizo” o un “amarillo azulado”. El mecanismo trabaja a través de un proceso de respuestas excitatorias e inhibitorias, donde las dos componentes son opuestas entre si. Esto es, si un estímulo tiene componentes rojas y verdes de igual intensidad los efectos se cancelan y no se puede ver color alguno.

2.2. Los principios de la visión artificial

2.2.1. La cámara y las imágenes digitales

Una *imagen* se podría definir como una representación visual de un objecto o una escena. Ya sea una imagen mental en el cerebro o una fotografía en una computadora, esta representación es una abstracción de la realidad que rodea a las personas. En este sentido, siempre existirá algún tipo de sistema intermediario a partir del cuál se pueda obtener una abstracción del mundo y que dé origen a una imagen.

De forma análoga a como lo hacen los ojos en una persona, una cámara fotográfica o de video estándar provee información de aquello visible a través de una fotografía o una secuencia de fotogramas, respectivamente. Mediante un proceso meramente óptico, los rayos de luz provenientes del ambiente se enfocan a través de un sistema de lentes y se proyectan hacia el foco de la cámara formando la imagen. Vale la pena mencionar en este punto que existen distintos tipos de cámaras que pueden capturar la luz en distintas zonas del espectro de frecuencias, como por ejemplo las cámaras infrarrojas y las multiespectrales. En esta tesis solo se considerará la utilización de cámaras estándares o convencionales.

En las cámaras más antiguas la imagen se almacenaba mediante la exposición de una película fotosensible a los rayos de luz, reflejando un proceso puramente analógico. Sin embargo, en las cámaras actuales la luz es capturada por un arreglo de sensores electrónicos, usualmente sensores *CCD* (del inglés *Charge Coupled Device*) o *CMOS* (del inglés *Complementary Metal Oxide Semiconductor*), y la imagen es almacenada en formato digital en una tarjeta de memoria u otro dispositivo de almacenamiento que contenga la cámara.

Las cámaras digitales hoy en día son accesibles, compactas y de bajo costo. Permiten extraer gran cantidad de información del entorno y se utilizan permanentemente en áreas pujantes de la investigación científica como son la medicina, las ciencias de la computación y la robótica, entre otras. Así como los ojos juegan un rol fundamental en la navegación de gran parte de los seres vivos, la cámara es protagonista principal en gran cantidad de aplicaciones de la robótica móvil y los vehículos no tripulados o autónomos, áreas de principal interés para esta tesis.

La imagen digital es el resultado de una discretización de la imagen "continua" original. A diferencia de la retina, donde el muestreo es altamente no uniforme, los sensores fotosensibles de la cámara se organizan en arreglos espacialmente uniformes dando como resultado una misma resolución para toda la imagen. Numéricamente, la imagen se representa con una matriz bidimensional de elementos llamados *pixeles* (acrónimo del inglés *picture elements*). Cada pixel adoptará valores que dependerán directamente de la cantidad de luz capturada por el sensor correspondiente, la longitud de onda de la misma, y las características y el tipo de sensor. En el caso de la generación de imágenes a color, las cámaras se construyen con tres fotosensores por pixel que responden diferente ante el espectro de la luz, tal como sucede en el ojo humano. La descripción y representación de los colores de una forma precisa a partir de estas tres mediciones tiene significancias tecnológicas y comerciales, por lo que han requerido de la definición de determinados estándares para su uso (ver 2.2.3).

2.2.2. El modelo geométrico de la cámara

En la sección 2.2.1 se introdujo brevemente el principio básico de funcionamiento de una cámara a partir del cuál se genera una imagen. Si bien el modelo real de una cámara puede ser bastante complejo, aquí solo se considera un modelo óptico sencillo como el de la Fig. 2.4. La relación entre las distancias al objeto z_m, su proyección z_c y la *distancia focal* f_o está dada por la Ec.(2.1). El ángulo fov (del inglés *field of view*) indica el campo visual de la cámara, que depende del ancho W del sensor y la distancia focal f_o. Si $z_m \to \infty$, es decir que se acerca el plano de la imagen al foco, se estará considerando el conocido modelo del tipo "pinhole", que será suficiente para algunos análisis que se harán más adelante.

$$\frac{1}{z_m} + \frac{1}{z_c} = \frac{1}{f_o} \tag{2.1}$$

Son muchos los factores que afectan la información y la calidad de una imagen digital: el enfoque, la exposición, la apertura, la velocidad de obturación, el balances de blancos/gamma, la compresión, el ruido y la resolución de los conversores A/D, entre los más importantes. Dado que el objetivo de esta tesis no está centrado en la captura

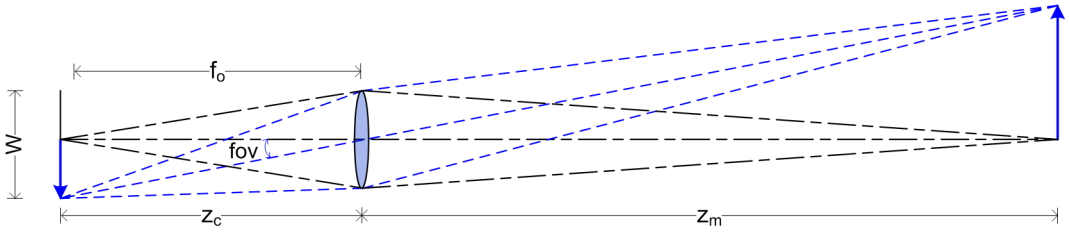


Figura 2.4: El modelo geométrico considerado para la cámara.

de una imagen sino más bien su procesamiento, estas cuestiones no serán analizadas. Tampoco se considerarán las posibles aberraciones y distorsiones de los lentes, así como no se consideraron para el caso del ojo humano. Finalmente, debido a que la imagen es el resultado de un muestreo o una discretización, la teoría de señales justifica la aparición de efectos de solapamiento o *aliasing*, que tampoco se contemplarán durante esta tesis.

2.2.3. La representación de los colores

Las diferentes formas de representar un color han surgido generalmente por cuestiones tecnológicas y/o comerciales. Inclusive a lo largo de la historia se han relacionado algunos colores específicos con determinados productos de la industria [Forsyth & Ponce (2002)]. Sin embargo, las personas no son capaces de reconocer ni identificar todos los colores, por lo que se hizo necesaria la implementación de sistemas estándares que definieran el uso de los mismos. Durante años las organizaciones y las tecnologías relacionadas con la industria del color han impulsado muchos estándares para la representación, transmisión y almacenamiento de imágenes y videos. Aquí solo se hará mención de las más relevantes por lo que para mayores detalles se recomienda acudir a la bibliografía referenciada en el texto. Esto lleva a definir primero lo que hoy se conoce como modelos de color y espacios de color.

El *modelo de color* es un modelo matemático abstracto que describe como representar un color a partir de un conjunto de números (suelen ser 3 o a lo sumo 4). Cada uno de estos valores o "componentes de color" define un eje dentro un sistema de coordenadas para identificar todos los colores que se pueden percibir. Cuando existe además una descripción precisa de como mapear un punto a un color, es decir que cualquier conjunto de coordenadas x, y, z especifica de forma inequívoca un color en sentido absoluto, se

habla de un *espacio de color*. Aunque muchas veces se utilizan ambos conceptos como equivalentes es importante saber diferenciarlos cuando corresponda. De aquí en adelante se hablará de espacios de colores cuando se haga referencia al tema.

En 1930 el organismo internacional CIE (del francés *Commission Internationale d'Eclairage*) estandarizó la representación **RGB** a través de experimentos de correspondencia de colores (*color matching* en inglés) [Szeliski (2010)]. Estos experimentos consistían en combinar el rojo (700nm), el verde (546.1nm) y el azul (435.8nm) para reproducir toda la gama de colores puros o monocromáticos (una sola longitud de onda). Las funciones de correspondencia estándar se obtuvieron promediando los resultados de percepción obtenidos con gran cantidad de personas. Debido a que para algunas regiones del espectro se hizo necesario utilizar cantidades negativas de luz roja para lograr la correspondencia, el CIE desarrolló también un espacio de colores llamado **XYZ**. Este nuevo espacio contiene todos los colores espectralmente puros dentro de su octante positivo y se obtiene a partir de una transformación lineal del espacio RGB. Dicha transformación tiene la particularidad de mapear la *luminancia* del color (brillo relativo percibido) directamente hacia el eje Y, dando a esta componente una significancia especial. En estos dos casos los primarios se combinan en forma aditiva para generar luz de color dando lugar a los sistemas de reproducción de *colores aditivos* [Gonzalez & Woods (2001)]. Se puede utilizar como ejemplo de este mecanismo combinatorio a los fósforos de un monitor de PC, un televisor a color o un proyector de imágenes.

Respecto la transmisión comercial de video en los sistemas de televisión a color, los dos estándares más conocidos y que primero aparecieron son el **NTSC** y el **PAL** (según el país que corresponda). En el caso del NTSC se utiliza el espacio **YIQ**: la componente Y se denomina *luma*, y es proporcional a la luminancia con la corrección gamma[1]; y las componentes IQ se denominan *crominancia*, y representan conjuntamente el tono y la saturación del color. La ventaja de este sistema frente al RGB es que la componente Y podía utilizarse también para los televisores blanco y negro, y las componentes IQ permitían limitar su ancho de banda sin producir una degradación de la imagen perceptible a los ojos. Respecto del sistema PAL, el espacio de colores elegido es el **YUV**, donde Y también representa a la luminancia y las componentes UV

[1]Dado que estos sistemas fueron diseñados para los televisores a color con tubos de rayos catódicos (CRT), a las componentes del color se les aplicaba previamente la *corrección gamma* para compensar la respuesta no lineal del fósforo a la luminancia del color [Gonzalez & Woods (2001)].

se relacionan con las IQ por una simple rotación de coordenadas en el espacio de los colores.

Por otro lado, cuando se trata de pinturas u otros procesos de impresión de colores se utilizan sistemas de *colores sustractivos*. Se llaman sustractivos ya que los pigmentos utilizados absorben ciertas longitudes de onda del espectro. De esta manera, también es posible generar luz de color a través de la utilización secuencial de filtros de colores antepuestos a una luz blanca (W). Los colores primarios sustractivos son el cyan (W-R), el magenta (W-G) y el amarillo (W-B), dando lugar a la representación **CMY** (por las iniciales de cada color en inglés). Por cuestiones de calidad y costos, los sistemas de impresión generalmente utilizan 4 colores, incluyendo también el negro (**CMYK**).

Los espacios de color que se obtienen a partir de combinaciones de fuentes primarias mediante funciones de correspondencia se denominan espacios *lineales*. Mientras que el RGB y el CMY son espacios lineales por definición, el XYZ también lo es porque se obtiene a partir de una transformación lineal de un espacio lineal. El inconveniente que tienen estos espacios lineales es que en realidad no reflejan fielmente como los humanos perciben las diferencias de color y luminancia. En la sección 2.1.2 se introdujeron los conceptos de brillo, tono y saturación como atributos comúnmente utilizados para describir mejor la sensación ante la luz. No es posible para un espacio lineal modelar por ejemplo la intuición humana acerca de la naturaleza "circular" del tono, en el sentido de que éste cambia del rojo a través del naranja hacia al amarillo y el verde, y desde allí al cyan, el azul y el violeta, para llegar nuevamente al rojo.

Los espacios *no lineales* de color se han diseñado justamente para capturar la intuición humana acerca de la topología de los colores. Entre los más utilizados se encuentran los espacios **HSV** y **CIE LAB**. Las siglas del primero surgen precisamente de las traducciones al inglés de tono (*Hue*), saturación (*Saturation*) y valor (*Value*) o brillo. Se obtiene al aplicar una transformación no lineal al cubo tridimensional RGB para obtener lo que se representa esquemáticamente como un cono. En más detalle, la componente V se define como el valor máximo de las tres componentes RGB y se representa como la distancia vertical al generador del cono; la componente H se define como la dirección o ángulo en la rueda de colores; y la componente S se define como la distancia radial a la diagonal [Forsyth & Ponce (2002)]. Por último, el espacio CIE LAB está inspirado en la respuesta prácticamente logarítmica del ojo humano, según la cual se puede percibir

diferencias relativas de luminancia de aproximadamente 1 % [Szeliski (2010)]. Se define como un mapeo no lineal del espacio XYZ a través una función cúbica y una fuente blanca de referencia, donde las diferencias resultantes en luminancia y en crominancia son mucho más uniformes perceptualmente. La componente L está correlacionada con el brillo o luminosidad del color, mientras que las componentes a y b se relacionan con los oponentes $R - G$ y $Y - B$, respectivamente.

2.3. Resumen

En este capítulo se han estudiado los procesos biológicos elementales que se dan en el sistema visual humano y que permiten a las personas percibir el mundo a través de sus ojos. Por otro lado, se ha visto como es el proceso básico de generación de imágenes a color para las cámaras digitales, elemento principal de los sistemas de visión artificial.

En el próximo capítulo se estudiarán los procesos del cerebro involucrados en el reconocimiento visual del terreno y como han evolucionado los sistemas artificiales para su detección a partir de una imagen.

Capítulo 3

El reconocimiento del terreno transitable

Las personas tienen una gran capacidad para identificar visualmente el terreno por el que pueden transitar de forma segura ya sea caminando, andando en bicicleta o cuando conducen un vehículo. El sistema visual humano demuestra una gran robustez para resolver esta tarea para diferentes condiciones de iluminación y del clima, y en diferentes situaciones y ambientes. Esto ha inspirado durante muchos años el diseño de sistemas basados en visión artificial para la detección del terreno transitable en aplicaciones como la robótica móvil, los vehículos autónomos y los sistemas de asistencia al conductor. La utilización de cámaras tiene aquí un rol fundamental ya que todas señales de tránsito y otras señalizaciones relacionadas están pensadas y diseñadas exclusivamente para ser vistas por las personas. Si bien estos sistemas han evolucionado enormemente en su capacidad para procesar estímulos visuales cada vez más complejos y en su robustez frente a diferentes situaciones y ambientes, aún sufren de muchas limitaciones cuando se exponen a las condiciones que normalmente enfrenta un conductor en su vida diaria, dejando abiertas las puertas para nuevas investigaciones y desarrollos en la temática.

El capítulo se organiza de la siguiente manera: en la Sección 3.1 se describen de forma breve los procesos que coexisten en el cerebro cuando las personas reconocen determinados patrones visuales, y se analiza cómo las personas suelen reconocer visualmente un terreno transitable y que tipo de evidencias utilizan para determinarlo; en la Sección 3.2 se estudia en profundidad como han evolucionado los sistemas automáticos para la detección de caminos en imágenes monoculares, incluyendo un análisis conceptual de la bibliografía y una descripción de las partes principales de un sistema; y por último, en la Sección 3.3 se identifican las principales problemáticas que enfrentan los sistemas basados en visión y cuáles son los desafíos que aún restan por resolverse, incluyendo aquello que motiva este trabajo de investigación.

3.1. La percepción humana de lo transitable mediante visión

3.1.1. Los procesos del reconocimiento visual

El cerebro humano tiene una capacidad enorme para reconocer sonidos, olores, objetos o cualquier otro estímulo que le sea familiar, incluyendo aquellos más complejos como el comportamiento de las personas. Constantemente, a través de los sentidos el hombre es capaz de percibir en el mundo que lo rodea similitudes y diferencias con lo que ha aprendido y ya sabe de él. Sin duda, este conocimiento permite extraer de entre tanto estímulo sensorial solo aquella información justa y necesaria para lo que se esté haciendo en el momento.

El sistema visual humano refleja inherentemente las mismas capacidades para reconocer objetos, animales, rostros, y todo tipo de patrones visuales. Tal como se ha descrito en el capítulo anterior, las cortezas jerárquicamente superiores del cerebro como la corteza TI son las encargadas del reconocimiento de estos estímulos más complejos. Sin embargo, se mencionó que las zonas periféricas del sistema como la retina y la corteza V1 son solo capaces de detectar estímulos más elementales como contrastes de colores y bordes a diferentes escalas, y hasta quizás líneas y curvas en las cortezas intermedias. Por lo tanto, no es posible para el sistema detectar o reconocer objetos si no es a través de una composición o integración espacial y/o temporal de características visuales más elementales.

Se podría pensar entonces que el reconocimiento de objetos se realiza por partes, agrupando de alguna manera las distintas evidencias de su existencia. En [Bar et al. (2001)] se dice que las características visuales se extraen primero en las áreas de nivel más bajo y luego se proyectan a regiones de más alto nivel, donde se forma una representación visual de la imagen de entrada. Presumiblemente, esta representación se compara luego con las representaciones de objetos almacenados en memoria. Este tipo de integración de información de "abajo hacia arriba" habitualmente se define como del tipo *bottom-up*. Por otro lado, existe evidencia de que el reconocimiento también está influenciado por procesos del tipo *top-down* o de "arriba hacia abajo", donde información de alto nivel se activa antes que aquella de más bajo nivel facilitando el reconocimiento. [Bar (2003)] propone que una versión aproximada y de baja resolución de la imagen de

entrada se transmite rápidamente a regiones del cerebro como la corteza prefrontal, desencadenando procesos de predicción que reducen la cantidad inicial de hipótesis acerca del objeto y que se refuerzan con los procesos *bottom-up*. En [Gerlach et al. (2002)] se afirma que en el giro temporal inferior (en el lóbulo occipital) la integración se dá según un proceso primariamente *bottom-up*, pero la integración en las partes posteriores de los giros fusiforme y temporal inferior está modulada de alguna forma por el conocimiento estructural almacenado acerca de los objetos. Estos conceptos acerca del flujo bidireccional de la información a través de procesos *bottom-up* y *top-down* se retomarán más adelante cuando se trate la atención selectiva, mencionada antes en la Sección 2.1.

La forma en que el cerebro organiza el conocimiento acerca de los objetos ha sido motivo de estudio por más de 100 años. Sin embargo, todavía se está muy lejos de tener un entendimiento acabado del tema y por lo tanto, aún hoy la investigación en esta área se mantiene muy activa. La idea central que se plantea en [Martin (2007)] es que el conocimiento acerca de los objetos se organiza por características sensoriales (por ejemplo la forma, el movimiento, el color) y por propiedades motrices asociadas con el uso del objeto. Otros modelos también consideran las propiedades funcionales/verbales como por ejemplo donde encontrar típicamente al objeto, su implicancia social, etc. La información sobre los diferentes tipos de propiedades de los objetos se almacenan en diferentes regiones del cerebro. Existe también evidencia de que las propiedades sensoriales y motrices del objeto se almacenan justamente dentro de los sistemas sensoriales y motrices, respectivamente. Por otro lado, las regiones asociadas con diferentes propiedades se involucran de diferentes maneras en función de la categoría a la que pertenezca el objeto. Si bien aún resta identificar aquellos sistemas neuronales que albergan la vasta reserva de conocimientos verbales, no sensoriales/motrices, y formales o del tipo "enciclopédicos", se puede afirmar que el conocimiento acerca de lo que se ve es por naturaleza distribuido y por ende así serán los procesos involucrados en el reconocimiento visual del entorno.

3.1.2. Las características visuales del terreno transitable

La sección anterior ha descrito conceptualmente los procesos básicos involucrados en el reconocimiento visual de objetos. De forma breve se puede decir que una persona será capaz de reconocer un patrón si sus propiedades o características visuales coinciden de alguna manera con aquellas que ya conoce acerca él. Vale la pena repetir que este

"conocimiento", que se encuentra almacenado por partes y en distintos niveles en función de dichas propiedades, interactúa constantemente con la periferia sensorial modulando su comportamiento mientras se inspecciona el campo visual.

El reconocimiento de estímulos visuales abstractos como situaciones, comportamientos o atributos comprendería un patrón similar. Aunque su complejidad será mayor se puede suponer que el proceso tendrá muchos puntos en común con el caso de los objetos concretos. El problema que se plantea en esta tesis está justamente relacionado con este tema. Uno de los objetivos que aquí se persigue es aprender acerca de como las personas determinan si un terreno es transitable a partir de lo que ven con sus ojos. Escrito de otra manera, se pretende estudiar cuales son las características visuales del entorno que permiten decidir si es posible o no transitar por un lugar. La transitabilidad del terreno dependerá por supuesto del medio que se pretenda utilizar para transitarlo. La evaluación que se hará no será la misma si se transita a pie, en bicicleta o en un colectivo.

Dado que el contexto de esta tesis se encuentra vinculado a los vehículos autónomos y a las aplicaciones de vehículos inteligentes, el concepto de transitabilidad estará implícitamente asociado a los automóviles o vehículos similares, aunque parte del análisis será extensible a otro tipo de vehículos terrestres. Entonces, desde un punto de vista algo elemental, se podría imaginar aquello transitable como una zona o un lugar por el que es físicamente posible para un vehículo poder transitar. Si bien esto es un requisito fundamental que se tiene en cuenta cuando se maneja un automóvil, el concepto de transitable que se considerará aquí es un poco más abarcativo. Será transitable todo aquel terreno por el que sea civilmente correcto transitar con un vehículo. Es necesario aclarar sin embargo que no se considerarán cuestiones relacionadas con las leyes y reglas de tránsito, que implican no solo un reconocimiento visual del ambiente sino también un análisis del tipo sociológico y cultural.

En general, cuando se habla del lugar por donde transitan los automóviles inmediatamente se lo asocia con caminos, rutas o autopistas. Inclusive, si se busca en un

diccionario la definición de estas palabras[1], éstas se encuentran íntimamente relacionadas con la transitabilidad. Teniendo en cuenta esto, el problema se traduce a reconocer los estímulos visuales que utilizan las personas para determinar que parte del terreno visible pertenece a un camino.

En relación con el reconocimiento de patrones, para identificar un camino hace falta determinado conocimiento acerca del mismo. Si bien no todos los caminos son iguales, las personas utilizan con éxito distintos tipos de evidencia para resolverlo. Para motivar el análisis acerca del tipo de información que caracteriza a un camino, se realizó una encuesta abierta informal[2] con un grupo de 24 personas mayores de 21 años. Esta actividad consistió en que los participantes observaran un conjunto de 9 fotografías con diferentes tipos de situaciones (ver Fig. 3.1) y para cada una de ellas se respondiera la siguiente pregunta motivadora:

> *¿Que información visual utiliza nuestro cerebro para determinar por donde se puede transitar o no con un vehículo? Es decir, ¿cuáles son las "cosas" que miramos para poder identificar el camino/los posibles caminos?*

Los resultados obtenidos se procesaron luego para eliminar las respuestas repetidas y se agruparon en función del tipo de respuesta. En la Fig. 3.2 se incluye un diagrama con las diferentes categorías[3] de elementos visuales que llaman la atención. Se identificaron dos grandes grupos de categorías: aquellas que brindan información acerca de donde puede estar ubicado el camino y que se llamarán *indirectas* (resaltadas en el diagrama

[1]Definiciones según el diccionario [Espasa-Calpe (1992)]:

- *Camino*: 1. Tierra hollada por donde se transita habitualmente; 2. Vía que se construye para transitar; 3. Dirección que ha de seguirse para llegar a un lugar.
- *Ruta*: 1. Vía/camino; 2. Camino o dirección que se toma para un propósito.
- *Autopista*: 1. Carretera para alta velocidad, con dos direcciones separadas con un seto y desviaciones a distinto nivel.

Se omitieron aquí las definiciones de ruta que se relacionan con la palabra viaje.

[2]La encuesta se considera con carácter informal debido a que no se diseñó con rigurosidad estadística respecto al método, a la muestra de participantes encuestados, ni tampoco respecto al material utilizado. Los objetivos eran simplemente analizar el tema desde un punto de vista mas bien cualitativo y obtener ejemplos reales acerca de los estímulos visuales que utilizan las personas para reconocer un camino.

[3]Las categorías aquí propuestas no son necesariamente mutuamente excluyentes. Esto quiere decir que existen elementos que proveen información de distinto tipo y podrían incluirse en más de una categoría. Se puede citar por ejemplo el caso de un guardarail de una autopista, que define tanto el límite entre lo transitable y lo no transitable, pero también es parte de la infraestructura y se considera un obstáculo

Figura 3.1: Fotografías utilizadas en la encuesta informal acerca de los estímulos visuales que utilizan las personas para determinar cual es el terreno transitable por un vehículo.

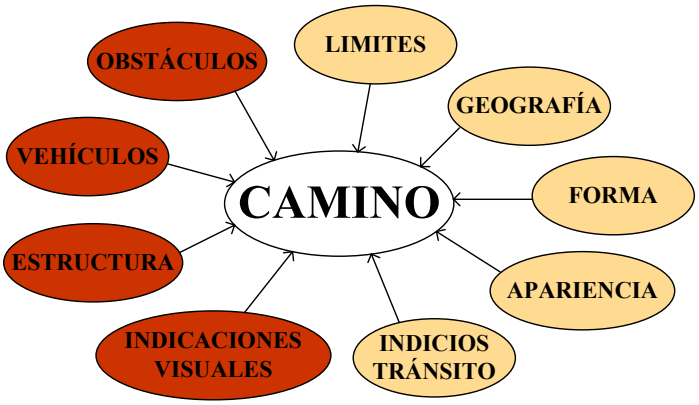


Figura 3.2: Diagrama conceptual de las categorías de elementos visuales que utilizan las personas para identificar la existencia de un camino. En color rojo se identifican los elementos que ayudan indirectamente a encontrar el camino mientras que en color amarillo se identifican aquellos que describen directamente sus propiedades.

con color rojo), y aquellas que describen directamente alguna propiedad del camino y que se llamarán *directas* (color amarillo).

Los elementos indirectos no describen al camino en sí y suelen ser sinónimo de lugares intransitables. Estos elementos incluyen a todo tipo de *obstáculos* y a los demás *vehículos*. Sin embargo, las personas asocian algunos de estos elementos directamente con un camino, ya que solo existen en dichos contextos. Este es el caso de la *estructura* (o infraestructura) y las *indicaciones visuales* mediante señales y carteles. En todos los casos la extracción de información del camino requiere algún tipo de razonamiento o asociación.

Los obstáculos se definen como cualquier objeto que sea físicamente impenetrable o bien implique un riesgo para las personas o los bienes materiales. Las personas asocian el terreno transitable con la ausencia y/o el "espacio libre" de obstáculos. Los participantes consideraron en esta categoría a los peatones, otros vehículos, los árboles, ramas y vegetación en general, los pozos y zanjas, y piedras de gran tamaño. En particular, los vehículos tienen un significado más abarcativo que el de un simple obstáculo. Frecuentemente, cuando se conduce un vehículo se analiza también el comportamiento de otros vehículos para determinar por donde se puede transitar. La existencia por ejemplo de vehículos en movimiento afirma que es posible transitar por un lugar, mientras que los automóviles estacionados indican lo contrario. Los participantes también se refirieron a indicadores como el encendido de las luces de stop, la cantidad de vehículos por carril y los sentidos de circulación, que tienen que ver más con cuestiones de alto nivel como la navegación.

La estructura del lugar se refiere a los elementos artificiales del ambiente cuyo propósito es dar una mayor seguridad a los conductores o bien brindar algún tipo de servicio. Se pueden citar como ejemplos a los servicios de alumbrado (postes de luz), los semáforos, y los separadores del tránsito como guardarails y paredes de contención. Las indicaciones visuales son en realidad parte de la estructura de un camino pero tienen un objetivo particular que es prevenir al conductor acerca de la existencia de obstáculos, cambios en la geometría del camino o bien del tránsito en general. Se utilizan distintos tipos de señalizaciones visuales como carteles o símbolos de colores pintados en el asfalto u otro elemento estructural.

Los elementos directos en cambio, describen las propiedades del camino y se han

dividido en las siguientes categorías: límites, geografía, forma, apariencia e indicios de tránsito. Los elementos dentro de la categoría *límites* son aquellos que justamente delimitan la zona transitable de aquella que no lo es, y que se caracterizan generalmente por algún tipo de contraste de apariencia, de los materiales que conforman la superficie, o bien de las alturas del terreno. Algunos ejemplos que utilizaron los participantes son las líneas de división de carriles, el borde de un pavimento, el contraste de colores entre el camino y la banquina o la vegetación, y las elevaciones del terreno como cordones y guardarails, entre otros. La *geografía* del camino se refiere a una descripción mas bien topográfica del terreno. Esto se entiende como una evaluación de aspectos como desniveles y pendientes, la acumulación de materiales, la solidez del terreno y la existencia de pozos, zanjas o rocas. En general, se espera que el camino sea lo más "plano" posible.

Respecto a la *forma*, si bien no todos los caminos son iguales, en general se busca un patrón que se angosta hacia el horizonte debido a la perspectiva en lo que se mira. La existencia de líneas cuasi-paralelas o líneas de fuga hacia el horizonte ayuda a percibir el sentido y la dirección del camino. La *apariencia* es uno de los atributos que más utilizan las personas para distinguir las diferentes superficies de un lugar. En la mayoría de las situaciones los colores y su textura permiten separar el ambiente en regiones relativamente homogéneas. El terreno transitable se podría pensar como aquellas superficies que tienen una distribución homogénea de color y textura similar al suelo sobre el que se encuentra el observador. Por último, los *indicios de tránsito* son elementos que caracterizan aquellos lugares por donde hubo un tránsito previo de vehículos. Entre ellos se destacan las marcas y huellas de neumáticos, el desgaste del pavimento y las líneas pintadas, las manchas de aceite y/o combustible, y los signos de aplastamiento de la vegetación.

El análisis conceptual presentado es una representación aproximada del conjunto de elementos que conforman el "conocimiento" colectivo acerca de los caminos. Este tipo de conocimiento se acumula de alguna manera con el paso del tiempo y en función de la experiencia de cada persona. Desde pequeño se aprende a reconocer por donde es seguro caminar o andar en bicicleta, y luego de más grandes se aprende a reconocer por donde es viable conducir un vehículo. Este reconocimiento, además de tener un carácter distribuido, es altamente robusto, pues no todos los elementos presentados en el esquema son imprescindibles. Dependiendo de cada caso, solo algunos de ellos son

suficientes para poder identificar el camino. Además, según el contexto, la situación o el tipo de ambiente, determinados elementos son más esperados que otros, o bien algunos son más importantes que otros.

Para terminar, si se tuviera que redefinir el concepto de camino en función de lo planteado hasta ahora se podría resumir de la siguiente manera: es un área suficientemente plana y libre de obstáculos, con cierta homogeneidad en su apariencia y composición tal que parece una extensión del terreno por donde se está circulando, y que además tiene una forma que en perspectiva suele angostarse.

3.2. La detección de caminos con imágenes: análisis bibliográfico comparativo

Hace prácticamente tres décadas que la visión artificial tiene protagonismo en todo tipo de desarrollos con robots o vehículos autónomos [Desouza & Kak (2002)]. Cualquiera sea el tipo de escenario o la aplicación, la visión, junto con otros tipos de modalidades de sensado, ha sido de vital importancia para que el agente autónomo pueda identificar el terreno sobre el que puede transitar en forma segura. La visión ha sido también muy importante a la hora del reconocimiento de posibles obstáculos y potenciales riesgos, especialmente cuando se interactúa con personas [Gerónimo et al. (2010)]. La industria automotriz por ejemplo, hace tiempo que invierte sus esfuerzos en dotar de cierta inteligencia a sus vehículos con el objetivo de asistir al conductor en situaciones de peligro y disminuir de esta forma los riesgos de accidentes [Dickmanns (2002), Luettel et al. (2012)]. A través de múltiples esfuerzos se ha logrado progresar muchísimo en estos campos, incluyendo el gran impacto que ha tenido el desarrollo tecnológico, que ha permitido abordar problemáticas de cada vez mayor complejidad y que exigen sistemas con gran capacidad de procesamiento de la información. Sin embargo, aún hoy existen muchos desafíos y se mantiene el interés por el desarrollo de métodos basados en visión para la navegación autónoma de robots.

La identificación visual de un camino es básicamente un proceso de carácter *monocular* [Caraffi et al. (2007)]. Si bien los humanos suelen apoyarse parcialmente en la visión estéreo cuando por ejemplo esquivan obstáculos con su vehículo, se ha visto anteriormente que éstos son capaces de percibir la profundidad también mediante un solo ojo.

Esto permite manejar fácilmente con un ojo cerrado analizando los diferentes indicadores de profundidad y utilizando la experiencia acumulada. En este sentido, la detección automática del camino se plantea principalmente como un problema de reconocimiento de patrones en imágenes monoculares, en el que las diferentes partes o porciones de una imagen se clasifican como "camino" o "no camino". La visión estéreo suele utilizarse más bien como un complemento válido tanto para la detección de obstáculos como para la identificación de terrenos irregulares.

Respecto al reconocimiento de los caminos, se ha planteado en las secciones anteriores que las personas utilizan su "conocimiento" acerca de éstos para identificarlos visualmente. Las diferentes soluciones que se han publicado en la literatura han ido siguiendo casi de forma natural un esquema similar [Bertozzi et al. (2000)]. Según la aplicación y el tipo de ambiente, se hacen suposiciones acerca del camino que derivan en un conjunto de características visuales que finalmente se buscarán en la imagen. Esta descripción explícita derivada del conocimiento se definirá como *modelo de camino*. Por lo tanto, el modelo considerado determinará que tipos de características se extraen de la imagen y como interactúan entre sí, para luego detectar donde está el camino mediante algoritmos de más alto nivel. En la Fig. 3.3 se incluye un diagrama que muestra el enfoque mencionado del problema. El procedimiento resulta similar a cómo el sistema visual humano extrae características elementales mediante su periferia y las integra en regiones superiores del cerebro, combinando procesos del tipo *bottom-up* y *top-down*.

Los ambientes o entornos donde se utiliza el sistema suelen clasificarse en *estructurados*, *semiestructurados* y *no estructurados*, en función del nivel de infraestructura que acompaña al camino [Bar Hillel et al. (2012)]. Entre los ambientes estructurados se encuentran las autopistas y las rutas cuyas demarcaciones y señalizaciones se encuentran en excelente estado. Los ambientes semiestructurados suelen encontrarse en rutas secundarias y caminos urbanos o suburbanos, mientras que los caminos mas bien rurales generalmente carecen de estructura alguna.

3.2.1. La detección del carril y la dependencia de la estructura

En el caso de las rutas y autopistas, las líneas demarcadas sobre el asfalto son el principal elemento visual utilizado por las personas para identificar por donde conducir su vehículo. Los sistemas automáticos que se basan en estas líneas para identificar el

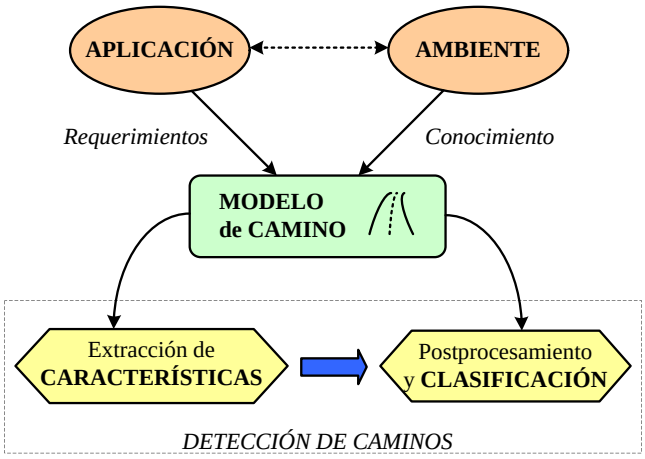


Figura 3.3: Diagrama esquemático del enfoque generalmente utilizado en la detección de caminos. Las propiedades del ambiente y los requerimientos impuestos por la aplicación definen un modelo apropiado para el camino. Este modelo permite elegir las características que se extraen de la imagen y como interactúan. Finalmente, algoritmos de más alto nivel se encargan de procesar y clasificar esta información.

área transitable se denominan sistemas de *detección de carril* (más conocido como *lane detection* en inglés). Aunque esta tesis no tiene particular interés en estos sistemas, se hará una breve introducción dado que es una aplicación desarrollada y presente en vehículos comerciales de alta gama. Los métodos para la detección de carril tienen tres aplicaciones típicas dentro de los sistemas actuales de asistencia al conductor: la advertencia de abandono de carril (*Lane-Departure Warning System* en inglés), el monitoreo de la atención del conductor (*Driver-Attention Monitoring System* en inglés), y el control automático del vehículo (*Automated Vehicle-Control System* en inglés).

En [McCall & Trivedi (2006)] se realizó un análisis exhaustivo del estado del arte en las investigaciones sobre *lane detection* que puede ser muy útil para aquellos que quieran profundizar en el tema. Los autores presentan una interesante comparación entre gran variedad de métodos, indicando las similitudes y diferencias entre ellos así como también en que situaciones cada uno de éstos tienen mayor utilidad. Se tuvieron en cuenta para este análisis los modelos de camino supuestos, las características que se extraen de la imagen, los métodos para postprocesarlas y los métodos para hacer el seguimiento a través del tiempo del camino detectado. Se identificaron además las suposiciones más comunes que se hacen acerca de los caminos: 1) que la textura del camino/carril es

consistente, 2) que el ancho del camino/carril es localmente constante, 3) que las líneas demarcadas siguen reglas estrictas respecto a su apariencia y ubicación, y 4) que el camino es plano o bien sigue un patrón determinado en el cambio de elevación.

Tal como se mencionó, los sistemas de detección de carril tienen una alta dependencia de la estructura del lugar y en general son poco robustos ante situaciones donde por ejemplo las líneas están desgastadas o son discontinuas, o cuando el asfalto cambia de apariencia, o bien cuando el carril cambia de dimensiones. De los métodos analizados aquellos que han tenido un mejor desempeño son los que proponen un modelo geométrico para los bordes del camino. En la literatura se han modelado los carriles mediante líneas rectas [Apostoloff & Zelinsky (2003)], curvas del tipo clotoide [Dickmanns & Zapp (1986), Southall & Taylor (2001)], *splines* [Wang et al. (2004)], hipérbolas [Chen & Wang (2006), Wang et al. (2008)], y parábolas [McCall & Trivedi (2006)]. La ventaja de utilizar un modelo de este tipo es que permite un mayor rechazo a los datos erróneos o aislados (*outliers* en inglés), y una mayor robustez frente al ruido y las zonas sombreadas.

3.2.2. Los sistemas para la detección de caminos en ambientes poco estructurados

En el caso de ambientes semiestructurados o con poca estructura no se conocen a priori las características visuales que puede tener el camino, e inclusive éstas pueden ser variables. Esto hace que el diseño de sistemas inteligentes para la detección de cualquier tipo de camino sea aún hoy un desafío. Cuando no es posible suponer la existencia de líneas o límites bien marcados, en general se supone al camino como una región homogénea (en algún sentido), que contrasta de alguna forma con su entorno y que puede adoptar alguna forma geométrica simplificada. Además, generalmente se consideran condiciones ideales de iluminación. Todas estas suposiciones aplicadas en un algoritmo pertenecen a lo que se ha definido antes como el modelo de camino.

El aprendizaje del modelo y su entrenamiento

Una vez definido un modelo es necesario determinar todos los parámetros del mismo. Dado que éstos son desconocidos o bien solo se conoce su distribución estadística aproximada, el enfoque que se utiliza comúnmente es que el sistema pueda determinarlos por

si solo a través de algún mecanismo de aprendizaje. Es aquí donde entran en juego las diferentes técnicas de la inteligencia artificial y el aprendizaje de máquina, más conocido como *machine learning* en inglés ([Bishop (2006)]). La elección de la técnica utilizada depende en cada caso del modelo planteado, de los parámetros a determinar, y de los requisitos que imponga la aplicación.

El procedimiento a partir del cual se definen los parámetros del modelo se denomina *entrenamiento*. Los métodos de entrenamiento suelen clasificarse en *supervisados* y *no supervisados*, aunque en la literatura que aquí interesa también suelen encontrarse métodos denominados *autosupervisados* y que son una combinación de los dos primeros. Los supervisados se basan en la determinación de los parámetros a partir de un conjunto de casos de ejemplo obtenidos a partir de un "supervisor". En este caso se habla por ejemplo de una persona experta (o un grupo) que, a partir de lo que ve, pueda clasificar las diferentes partes de una imagen como partes de un camino o no. Los casos de ejemplo utilizados para el entrenamiento deben ser suficientemente representativos del tipo de caminos que se pretende detectar para asegurar un buen desempeño con imágenes que no pertenecen al conjunto de entrenamiento. Una de las mayores desventajas de estos métodos es que se necesitan grandes cantidades de imágenes de entrenamiento para obtener un sistema suficientemente robusto. Ésto a su vez implica la inversión de muchas horas para el etiquetado manual, o mediante algún sistema semiautomático, de todas las imágenes. El sistema propuesto por [Rasmussen (2002)] ejemplifica un caso de aprendizaje supervisado. Por otro lado, los métodos no supervisados no suponen un conocimiento a priori acerca de los parámetros del problema. La idea básica es ajustar los modelos en función de las observaciones, identificando patrones y relaciones entre ellas. También se los suele utilizar para generar modelos a priori para el entrenamiento de sistemas supervisados, como es el caso por ejemplo de [Crisman & Thorpe (1991)]. Otros ejemplos de aprendizaje sin conocimiento previo alguno incluyen la detección de lineas [Kang et al. (1996)] y puntos de fuga [Rasmussen (2004a)] del camino a partir de la textura de la imagen. Por último, el término autosupervisado se suele utilizar para aquellos sistemas que generan de forma automática su propio conjunto de datos de entrenamiento. Esto es, a partir de ciertas suposiciones y/o de la utilización de otro tipo de sensores se define una región de la imagen que se acepta como parte del camino. A

partir de esta región es posible generar el conocimiento acerca de los parámetros que luego servirán para clasificar el resto de la imagen. En muchos casos, resulta lógico suponer por ejemplo que la zona justo frente al vehículo es transitable. Esta región se proyecta a la parte inferior de la imagen y suele utilizarse para el entrenamiento de algoritmos posteriores [Guo & Mita (2009)]. En [Moreyra & Masson (2011)] se ha presentado un método para definir esta región en función del punto de fuga del camino, lo que permite luego aprender como se distribuye el color dentro del mismo para clasificar el resto de la imagen. Algunos trabajos suelen utilizar pequeñas regiones predefinidas de la imagen llamadas semillas [Alvarez et al. (2008)]. Otros enfoques utilizan por ejemplo sensores de rango (visión estéreo y sensores láser, entre otros) para definir estas regiones "seguras" [Dahlkamp et al. (2006), Hummel et al. (2006)].

Los métodos de aprendizaje no suelen utilizarse solo para entrenar un modelo de camino que luego se utilizará de forma fija para toda la operación del vehículo. Una de las potencialidades que tiene este enfoque es que permite también que los parámetros se puedan ir actualizando constantemente a medida que se analizan nuevas imágenes del camino, dando al sistema cierta capacidad para adaptarse a los cambios en las características visuales que describen el terreno y en las condiciones de iluminación, entre otros posibles. La robustez final del sistema ante los cambios dependerá fundamentalmente del modelo asumido. Este tipo de integración temporal o "arrastre" de la información a lo largo de una secuencia de imágenes mejora la respuesta frente al ruido, favorece el rechazo de falsas detecciones, y reduce enormemente la carga computacional de los algoritmos.

Los principales elementos de un sistema

La detección de caminos mediante sistemas de visión, incluyendo los sistemas de detección de carril, ha sido objeto de estudio por más de 20 años. Por lo tanto, se pueden encontrar muchísimos trabajos de investigación que han propuesto una gran variedad de soluciones al problema, aunque todavía muchos aspectos siguen sin resolverse. Lo más sencillo aquí sería presentar un resumen atomizado y desconectado de las principales propuestas disponibles en la literatura de la temática. Sin embargo, en esta tesis se ha invertido mucho esfuerzo para poder hacer algo distinto y ayudar al lector a tener un mejor entendimiento del problema. Para ésto se identifican y describen las diferentes

piezas que suelen componer un sistema de este tipo, sus variantes, y las ventajas y desventajas de cada una.

Los primeros pasos en el área se dieron a partir de iniciativas gubernamentales principalmente en Europa y Estados Unidos, cuyo interés era incentivar el desarrollo de *sistemas de transporte inteligentes* que permitieran reducir costos y accidentes a la hora de transportar personas y cargas. Se instrumentó a través de proyectos de colaboración entre muchas unidades de investigación del mundo como universidades, instituciones gubernamentales y fabricantes de automóviles [Bertozzi et al. (2000)]. Algunos de los sistemas pioneros que se pueden mencionar son VITS(ALVINN) [Turk et al. (1988)], NAVLAB [Thorpe et al. (1988)], SCARF [Crisman & Thorpe (1993)], y RALPH [Pomerleau (1995)], entre otros.

Luego, los desafíos que planteó DARPA en los años 2004 y 2005 con los *Grand Challenges* atrajeron un nivel de atención sin precedentes hacia el desarrollo de vehículos autónomos capaces de transitar en ambientes no estructurados. Ambos desafíos consistían en recorrer autónomamente zonas desérticas con la asistencia de mapas digitales precisos y múltiples sensores [Bar Hillel et al. (2012)]. En general, la mayoría de los sistemas que tuvieron éxito se basaron en soluciones con GPS y sensores láser, utilizando la visión más bien como un complemento y no como un sensor primario. Sin embargo, se debe destacar que dichos desafíos despertaron el interés en los sistemas de visión para la detección de caminos con poca o sin estructura [Dahlkamp et al. (2006), Hummel et al. (2006), Alon et al. (2006), Caraffi et al. (2007), Nefian & Bradski (2006), Rasmussen (2006), Broggi & Cattani (2006)].

Haciendo un análisis comparativo y exhaustivo de la bibliografía, incluyendo desde aquellos primeros trabajos hasta los actuales, pasando por los sistemas propuestos para los desafíos planteados por DARPA, se identificaron seis factores que suelen variar en el proceso de clasificación de una imagen:

- *Selección* del sujeto o candidato de análisis
- *Extracción de características* del sujeto
- *Evaluación* o valoración del sujeto
- *Clasificación* del sujeto
- *Integración temporal* de la información
- Método de *aprendizaje*

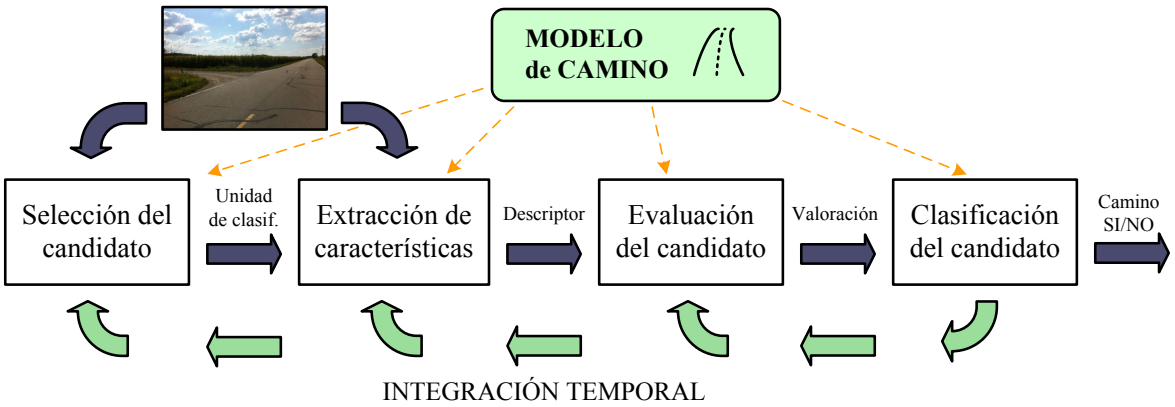


Figura 3.4: Factores involucrados comúnmente en el proceso de clasificación de una imagen. El modelo asumido define el tipo de sujeto o candidato de análisis, que se evalúa y clasifica en función de las características extraídas. El método de aprendizaje no se incluye en el diagrama pero se considera implícito. En el caso de secuencias de imágenes, los resultados se pueden integrar temporalmente de una imagen a otra para mejorar el rendimiento del sistema.

En la Fig. 3.4 se ilustra un proceso de clasificación completo, en el que el método de aprendizaje se encuentra implícito. El diagrama es de carácter genérico y representa a la mayoría de los enfoques, aunque por supuesto siempre existirán excepciones.

A partir de la vasta bibliografía analizada se eligieron los trabajos que se consideraron más relevantes y se los evaluó según cada uno de los factores del proceso de clasificación. Existen algunos casos en los que conviven dos procesos clasificatorios simultáneos o bien secuenciales, donde se utilizan diferentes tipos de candidatos de análisis. Los resultados de la evaluación se resumen en la Tabla 3.1. A continuación se describe cada uno de los factores evaluados:

Selección del sujeto o candidato de análisis

El sujeto o candidato de análisis es la unidad elegida por el modelo para analizar y clasificar. Los diferentes sujetos utilizados en la literatura suelen ser el *pixel*, el *patch*, el *cluster* y grupos geométricos definidos por *líneas*, *curvas* o formas geométricas de distintos tipos. El primero de los candidatos implica que la clasificación se realiza pixel por pixel. En el segundo caso, la clasificación se realiza por regiones o celdas regulares de NxN pixeles. Los *clusters*, en cambio, son conjuntos de pixeles agrupados por algún tipo de similitud mediante algoritmos de segmentación o de *clustering*. Generalmente se agrupan por su color, aunque en ocasiones se utiliza también la posición del pixel en

Tabla 3.1: Análisis de los principales trabajos de la literatura respecto de los factores involucrados en el proceso de clasificación. Para cada trabajo se marcó con un tilde los tipos de factores utilizados. En aquellos casos donde existen dos procesos clasificatorios, éstos se diferenciaron con superíndices 1 y 2.

	Candidato						**Caract.**					**Eval.**		**Clasificación**							**Integ.**			**Aprendiz.**			
	Pixel	Patch	Cluster	Línea	Curva	Otro	Color	Intensidad	Textura	Forma/Tamaño	Otro	Determinística	Probabilística	Umbral/distancia	Matching/voting/scoring	Modelo lineal	Modelo no lineal	Combinación de clasificadores	Estimador MAP	Toma de decisiones	Determinística	Filtrado estadístico	Ninguna	Supervisado	No supervisado	Autosupervisado	Ninguno
[Thorpe et al. (1988)]	√						√		√	√			√	√							√					√	
[Turk et al. (1988)]	√						√					√		√		√					√						√
[Crisman & Thorpe (1991)]			√[a]						√			√			√								√		√		
[Zhang & Nagel (1994)]		√							√		√[b]		√						√		√				√	√	
[Crisman & Thorpe (1993)]						√[c]	√						√						√		√					√	
[Rasmussen (2002)]		√					√		√		√[d]	√					√					√		√			
[He et al. (2004)]	√[2]				√[1]		√[2]	√[1]				√[1]	√[2]	√[2]						√[1]			√[1,2]			√[2]	√[1]
[Rasmussen (2004b)]				√			√		√			√			√							√			√		
[Dahlkamp et al. (2006)]	√						√						√	√							√					√	
[Hummel et al. (2006)]		√					√		√		√[e]		√						√				√			√	
[Alon et al. (2006)]		√[1]		√[2]					√[1,2]			√[1,2]			√[2]			√[1]			√[2]		√[1]	√[1]	√[2]		
[Caraffi et al. (2007)]		√					√			√	√[f]		√							√			√				√
[Franks et al. (2007)]					√		√	√	√				√						√			√			√		
[Alvarez et al. (2008)]	√										√[g]		√	√									√			√	
[Rasmussen & Scott (2008)]			√				√			√		√			√						√			√		√	
[Rasmussen et al. (2009)]						√[h]	√		√[i]				√						√			√			√		
[Guo & Mita (2009)]	√[1,2]							√[1]	√[2]		√[2]	√[j][1,2]		√[2]			√[1]				√[2]		√[1]	√[1,2]			
[Loose et al. (2009)]					√			√			√[k]		√						√			√			√		
[Manz et al. (2010)]					√		√						√						√			√			√		
[Kong et al. (2010)]				√			√		√			√			√								√				√

[1,2] Identifica un proceso de clasificación en aquellos trabajos donde se utilizan dos diferentes.
[a] Se agrupan pixeles por su color y su posición en la imagen.
[b] Se incluye la posición x,y del patch además de la medida de la textura.
[c] Se propone un modelo de intersección de caminos rectos vistos en perspectiva.
[d] Se utiliza un sensor láser 3D para incluir información de rango en cada patch.
[e] Se utiliza información de disparidad a partir de cámaras estéreo.
[f] Se utiliza información de la cinética del vehículo y del espacio libre de obstáculos (visión estéreo).
[g] Se realiza una transformación logarítmica del color a un espacio invariante ante la iluminación.
[h] Se propone un modelo triangular para un sendero visto en perspectiva.
[i] Como medida de la textura de una región se utiliza el histograma de como se distribuyen sus pixeles en grupos previamente segmentados con k-means.
[j] Se utiliza visión estéreo para computar los puntos de correspondencia entre el par de imágenes.
[k] Se incluye información de rango, que bien puede provenir de un radar o de visión estéreo.

la imagen. También es posible por ejemplo agrupar pixeles que pertenezcan a bordes alineados si la imagen es previamente procesada. En [Moreyra & Masson (2010), Moreyra & Masson (2011)] se ha utilizado como sujetos de análisis a los *clusters* generados a partir del color con el método de *clustering* k-means [Bishop (2006)]. A diferencia de los métodos basados en pixeles, los *clusters* permiten reducir significativamente la dimensionalidad de los datos en los posteriores problemas de evaluación y clasificación, aunque agregan un costo computacional extra. Para profundizar en el estudio de las diferentes técnicas utilizadas para segmentar imágenes a color se recomienda la lectura de [Lucchese & Mitra (2001)]. Por último, los grupos que se han denominado geométricos se utilizan cuando el modelo del camino considerado incluye aspectos geométricos o de forma. Tal cual fue mencionado en el Apartado 3.2.1, se suele suponer por ejemplo que el camino está delimitado por dos líneas paralelas en las coordenadas del mundo (que por la perspectiva de la cámara en la imagen se cruzan en lo que se denomina *punto de fuga*), o bien por diferentes tipos de curvas como clotoides, *splines*, hipérbolas, parábolas, etc. Sumado a lo mencionado en la Sección 3.2.1, otras claras ventajas de utilizar candidatos de este tipo es que reducen la cantidad de falsas detecciones y que permiten definir la geometría del camino en forma paramétrica, transformándose el problema de la clasificación en una identificación de los parámetros que mejor lo describen. La gran desventaja frente a los otros candidatos es que solo son válidos para caminos que tengan la forma definida por lo que no pueden adaptarse a situaciones comunes como bifurcaciones, intersecciones o rotondas.

Extracción de características del sujeto

El conjunto de características que se le extraen a cada candidato se denomina *descriptor*. La extracción de dichas características está asociada generalmente a determinados procesamientos de la imagen, que variarán dependiendo del tipo de candidato seleccionado por el modelo. Los elementos visuales que utilizan las personas para identificar la existencia de un camino se traducen principalmente a medidas de homogeneidad y contraste tanto de color como de textura de la imagen. De esta forma, los descriptores comúnmente se construyen a partir de información de *color* y de *textura*, aunque en algunos casos también se incluyen medidas de *forma/tamaño* (solo para candidatos grupales) o de *rango/altura* provenientes de otros sensores como el radar, el láser y las cámaras estéreo.

Respecto al **color**, una de las cuestiones básicas que suelen variar son los espacios de color, que fueron definidos en la Sección 2.2.3. Los más utilizados son los espacios más conocidos como RGB, HSV y CIE LAB, pero según el enfoque también se proponen otro tipo de transformaciones como $c_1c_2c_3$ y $l_1l_2l_3$ [Gevers et al. (1999), Miksik et al. (2011)]. Por otro lado, no siempre se utilizan todas las componentes del color. En [Manz et al. (2010)] por ejemplo, se utiliza solo la componente de saturación ya que se considera que los caminos en ambientes rurales contienen pixeles poco saturados. También se suelen utilizar combinaciones o normalizaciones de componentes como $R - B$ en [Turk et al. (1988)] y $\frac{R}{R+G+B}$ en [Miksik et al. (2011)]. Algunos trabajos proponen características basadas en la intensidad de la imagen, que si bien puede verse como una componente del color, por su significancia se ha mantenido aparte en la Tabla 3.1. Cuando el candidato utilizado es un grupo de pixeles, las medidas de que se proponen se construyen en general a partir de promedios o histogramas, o bien a partir de alguna función de contraste entre el candidato y la región que lo rodea.

Respecto de la **textura**, son muchas las variantes que se utilizan para describir a un candidato. Lo más común es encontrar descriptores basados en información acerca de los gradientes de intensidad y/o color de la imagen. Se utilizan desde simples filtros detectores de bordes o esquina hasta filtrados más complejos con *wavelets* de Gabor [Lee (1996)] u otros filtros direccionales. Tanto la intensidad de los gradientes como la orientación de los mismos sirven para definir como varía la textura de la imagen para diferentes tipos de terrenos. Para candidatos grupales se suelen usar la cantidad de bordes dentro de la región, sumas de intensidades, histogramas de orientaciones o de respuestas a diferentes filtros, histogramas de pixeles que pertenecen a distintos grupos resultantes de segmentaciones, y matrices de covarianza de la intensidad de los gradientes, entre otros.

Una de las líneas de trabajo que se ha profundizado en [Moreyra & Masson (2011)] está basada en la propuesta que originalmente hizo [Rasmussen (2004a)]. En un área utilizada por vehículos es muy común encontrar huellas o marcas que dan indicios de la dirección a seguir. Esto está íntimamente relacionado con los elementos visuales que se denominaron "indicios de tránsito" en la Sección 3.1.2. Esta información puede hallarse mediante un análisis de la orientación de la textura de la imagen. La orientación dominante $\theta(\mathbf{p})$ de un pixel $\mathbf{p}$ es la dirección que describe la estructura paralela más

intensa [Rasmussen (2004a)]. La estimación de dicha orientación se realiza aquí mediante la convolución de la imagen con un banco de filtros de Gabor generados a partir de:

$$g_{\lambda,\phi}(x,y) = e^{\left(a^2+\gamma^2 b^2\right)/2\sigma^2} e^{i\left(2\pi\frac{a}{\lambda}+\psi\right)}, \tag{3.1}$$

donde $a = x\cos\phi + y\sin\phi$, $b = -x\sin\phi + y\cos\phi$, ϕ es el ángulo de orientación de la ondita, λ y ψ representan la longitud de onda y la fase respectivamente, σ es la desvío estándar de la envolvente gaussiana en la dirección x, y γ es la relación entre el desvío estándar en x y desvío estándar en y. Al variar la escala y la orientaciones de los *kernels* es posible detectar estructuras de diferentes tamaños y direcciones. La orientación dominante para cada pixel será el valor de ϕ que maximice esta respuesta.

Si se supone que el camino a transitar es recto y se encuentra sobre una superficie aproximadamente planar, las líneas imaginarias definidas por la dirección de la textura convergerán en promedio a un punto denominado *punto de fuga* (PF). Este punto provee muy buena información acerca del camino que bien podría utilizarse para el guiado del vehículo, o bien para restringir la búsqueda y selección de candidatos como líneas o curvas.

Existen distintas formas de encontrar el punto de fuga en la imagen tal como han demostrado [Wang et al. (2004), Zhang et al. (2009), Wu et al. (2010)]. En esta tesis para el cálculo del PF se utilizó el algoritmo de votación propuesto en [Kong et al. (2010)] (*Locally Adaptive Soft-Voting (LASV)*). La evolución respecto del método original propuesto por [Rasmussen (2004a)] radica en que aquí solo votan aquellos pixeles cuya orientación dominante esté bien marcada y que se encuentren a una distancia menor que un radio predeterminado respecto del pixel candidato a PF.

En la Fig. 3.5 se muestran algunos de los resultados obtenidos en [Moreyra & Masson (2011)] con el algoritmo utilizado para la estimación del PF. En las columnas impares se puede observar el pixel más votado marcado con una cruz sobre la imagen original, mientras que en las columnas pares se encuentra el resultado de la votación, siendo los pixeles más blancos aquellos que recibieron mayor cantidad de votos. Principalmente se quiere ilustrar la potencialidad y la robustez del enfoque frente a distintos tipos de caminos, condiciones de iluminación y contraste.

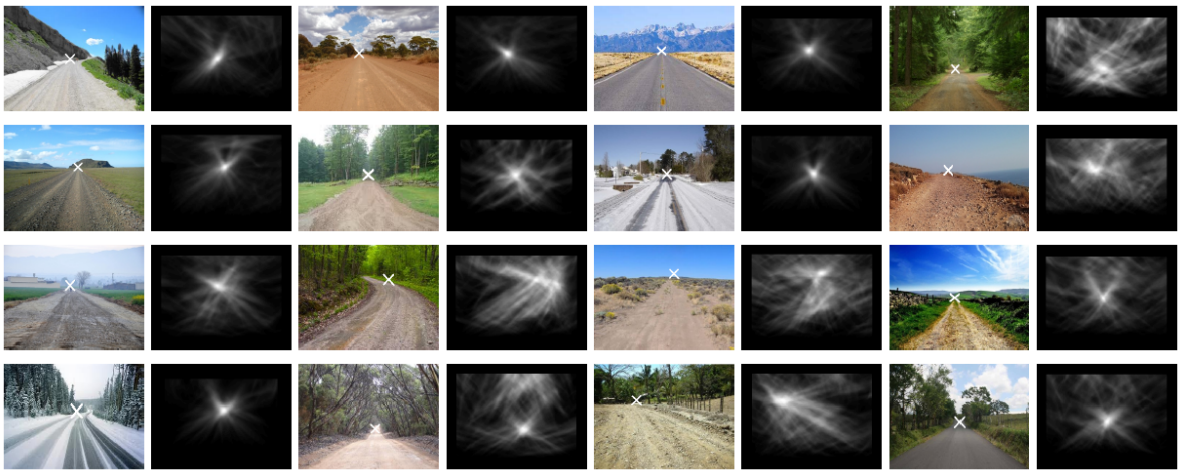


Figura 3.5: Algunos ejemplos de los distintos tipos de imágenes utilizadas en los experimentos presentados en [Moreyra & Masson (2011)]. Se puede observar el punto de fuga estimado marcado con una cruz (columnas impares) y el resultado de la votación (columnas pares).

Evaluación o valoración del sujeto

La evaluación del candidato consiste en valorar o calificar un candidato en función de cuan parecidas sean sus características respecto de aquellas del modelo de camino asumido. Básicamente existen dos maneras de evaluar a un candidato: determinísticamente o probabilísticamente. La forma elegida para la puntuación tiene relación directa con el método de clasificación a utilizar posteriormente.

Los métodos *determinísticos* asignan un valor al candidato a través de mecanismos como: funciones de similitud entre las características extraídas del candidato y las almacenadas para el camino, funciones de contraste entre el candidato y sus alrededores, modelos no lineales de aprendizaje supervisado, o algoritmos de votación. Cuanto más parecido sea el candidato a lo que se pretende de un camino, mayor será el valor asignado. Por otro lado, los métodos *probabilísticos* se basan justamente en asignar una probabilidad (o verosimilitud, *likelihood* en inglés) mediante algún modelo o una distribución previamente definida. El modelo más comúnmente utilizado en la literatura es la suma o mezcla de gaussianas, principalmente para las distribuciones del color. Es posible también tener múltiples modelos probabilísticos para los casos en que se extraen múltiples características. La utilización de funciones de verosimilitud permite además aplicar técnicas de fusión de datos cuando las características provienen de distintos tipos de

sensores.

Clasificación del sujeto

Los métodos de clasificación se utilizan para elegir los candidatos que efectivamente representan al camino en la imagen. Según las diferentes técnicas que conformen las etapas del proceso, se puede definir la clasificación de dos maneras: 1) determinar si existen evidencias suficientes para clasificar el candidato evaluado como un camino o parte de él; o 2) seleccionar a aquel candidato con puntaje más alto o que maximice algún tipo de función objetivo.

Con respecto a la primer definición, uno de los métodos más sencillos consiste en utilizar una función *umbral* a partir de la cuál se determina si el candidato es o no es transitable. Cuando se usan modelos probabilísticos ésto se traduce a una probabilidad mínima que se debe alcanzar. Una técnica similar consiste en utilizar modelos *lineales* o *no lineales* para definir si el candidato pertenece o no a la categoría predefinida como el camino. Se suelen utilizar por ejemplo funciones discriminantes lineales para el primer caso, y redes neuronales o máquinas de soporte vectorial (del inglés *support vector machines*) en el caso no lineal [Bishop (2006)]. Otra técnica relacionada resulta de la *combinación de clasificadores*, dando la posibilidad de entrenar y utilizar diferentes clasificadores según el caso que se presente. *AdaBoost* (proveniente de "Adaptive Boosting") es una de las más utilizadas para esto (ver también [Bishop (2006)]).

En relación a la segunda definición, las variantes se centran en la idea de hallar el candidato que maximice algún tipo de función o cantidad asignada. Las técnicas determinísticas comúnmente utilizadas son los análisis de correspondencia (*matching* en inglés), citando a[Crisman & Thorpe (1991)] como un ejemplo, y los algoritmos de votación (*voting* o *scoring* en inglés), tal como se propone en [Rasmussen (2004a)]. Las primeras consisten en utilizar una librería o lista finita de "plantillas" de distintas formas de caminos que se comparan uno a uno con el candidato evaluado, seleccionándose aquel que minimice el "error" o las diferencias entre ellos. Los métodos de votación en cambio, permiten dar puntaje (o votos) a cada candidato en función de distintos tipos de características que no son geométricas, pues el candidato las considera inherentemente. Aquel candidato que reciba más votos o mayor puntaje es el seleccionado. Por otro lado, cuando se utilizan probabilidades se suele denominar al clasificador como un "estimador", ya que a partir de la evidencia estima cual el mejor candidato en sentido probabilístico.

Tanto los estimadores basados en la *teoría de decisiones* como los estimadores de máximo a posteriori o *estimadores MAP*, pertenecen a esta categoría. En este último grupo se ubican por ejemplo los métodos de filtrado estadístico como son los filtros de Kalman [Chen (2012)] y los filtros de partículas [MacCormick & Isard (2000)]. Las ventajas de los filtros son tanto la capacidad de parametrización de los candidatos del tipo geométrico como la integración temporal de la información, más allá de de permitir la fusión de datos de múltiples sensores y múltiples características.

Integración temporal de la información

En la Sección 3.2.2 se mencionó a la integración temporal como una manera de aprovechar los resultados obtenidos con imágenes anteriores para facilitar y mejorar el rendimiento del proceso de clasificación en imágenes posteriores de una secuencia. En el análisis se identificaron métodos del tipo determinístico y otros del tipo probabilístico. Las técnicas *determinísticas* utilizan en general información proveniente de sensores capaces de medir el movimiento y la orientación del vehículo para proyectar los resultados obtenidos al espacio de la imagen siguiente (por ejemplo [Dahlkamp et al. (2006)]). Existen otros casos donde los modelos de color que se utilizan para clasificar los pixeles, se entrenan con información de la imagen actual y de imágenes anteriores (por ejemplo [Miksik et al. (2011)]). De esta forma, se reduce en parte el impacto de los cambios bruscos en la iluminación que podrían alterar drásticamente los colores aparentes del camino. En el caso de las técnicas del *filtrado estadístico* (filtro de Kalman, filtro de Partículas) se proyecta a la siguiente imagen una distribución probabilística que describe a los parámetros del modelo del camino (por ejemplo [Southall & Taylor (2001), Loose et al. (2009)]). El filtro de partículas por ejemplo, supone cada partícula como un posible conjunto de parámetros que describen al camino, esto es, un candidato. Cada partícula o candidato se proyecta a la siguiente imagen mediante la información acerca del movimiento del vehículo/cámara. Finalmente se actualiza la probabilidad asociada a cada una, en función de las características extraídas en la nueva imagen.

Método de aprendizaje

El último de los factores evaluados es la forma en que el sistema "aprende" los parámetros del camino. Los diferentes métodos de entrenamiento son el *supervisado*, el *no supervisado* y el *autosupervisado*, que ya se describieron en detalle en la Sección 3.2.2.

3.3. Limitaciones, desafíos y oportunidades

3.3.1. Problemáticas que se enfrentan

Independientemente del método o el enfoque utilizado, todos los sistemas de detección de caminos con imágenes se enfrentan a los mismos tipos de problemas e inconvenientes. En general, estos problemas son el resultado de cambios en las condiciones de iluminación, en las condiciones climáticas y del ambiente, y en las condiciones del camino que se pretende detectar. Cabe destacar que en general los sistemas se diseñan para funcionar en condiciones prefijadas según la aplicación y bajo suposiciones casi ideales acerca del ambiente y las situaciones a enfrentar. Sin embargo, muchas veces las condiciones distan de ser ideales y son pocos los trabajos en la literatura que analizan estas cuestiones. A continuación se listan la mayoría de los inconvenientes que suelen enfrentar dichos sistemas:

- *Cambios drásticos de iluminación* - Los ejemplos más comunes se dan cuando el automóvil (y la cámara) entra o sale de una zona oscura como túneles o sombras producidas por árboles o edificios. El reflejo de la luz sobre el asfalto u otros elementos, que luego incide directamente sobre la cámara, también puede producir estos inconvenientes. Si bien las cámaras de video suelen tener mecanismos de control automático de la cantidad de luz que ingresa, éstos no son suficientemente rápidos y actúan globalmente provocando cambios en la intensidad de toda la imagen. Esto tiene un impacto directo sobre la distribución de los colores que definen la apariencia visual de un camino complicando por ejemplo la integración de esta información entre imágenes consecutivas.

- *Zonas de muy alta y muy baja intensidad en una misma imagen* - Es muy común encontrar en una misma imagen zonas con sombras bien marcadas y zonas muy soleadas. Esta situación en general sobrepasa los límites físicos de una cámara ya que el rango dinámico justamente no es ilimitado. Las cámaras controlan la cantidad de luz que ingresa en función de las zonas más intensas de la imagen capturada previamente, oscureciendo el resto de la imagen.

- *Pérdida de información del color* - Tanto los pixeles muy oscuros como aquellos muy iluminados o sobreexpuestos carecen de información válida acerca del color.

Los pixeles muy oscuros son aquellos que prácticamente no recibieron luz, siendo la respuesta de los fotosensores muy afectada por el ruido mismo del detector. En el caso de los pixeles sobreexpuestos lo que se produce es una saturación de los sensores, deformándose la relación entre los canales de color, y provocando que estos pixeles se vean prácticamente blancos. En ambos casos la información de color no es confiable.

- *Sombras, marcas y manchas en el camino* - La existencia de sombras es uno de los mayores problemas para los sistemas de visión ya que genera fuertes contrastes en la imagen y suele provocar confusión por su diferencia en apariencia respecto de una zona similar sin sombras. También es muy común encontrar en el camino marcas y huellas de tránsito, manchas de combustible y/o aceite, y arreglos de pozos, baches o grietas del asfalto. Se podrían considerar también aquí a los charcos de agua debidos a pérdidas o acumulación de agua por las lluvias. En todos los casos la apariencia de dicha zona es muy diferente a la del camino y puede confundirse con la de un obstáculo.

- *Colores en común entre el camino y lo que no lo es* - En muchas ocasiones la apariencia del camino puede confundirse con la de otros elementos que lo rodean como muros y veredas. También se pueden encontrar objetos que yacen tanto en el camino como fuera de él, como son por ejemplo la acumulación de hojas o nieve.

- *Cambios de apariencia del camino* - Esta situación es habitual por ejemplo en caminos urbanos y rurales de nuestro país. En ocasiones el material con que se construye el camino, y por ende sus colores, suele variar en lugares donde se han realizado reparaciones del pavimento o bien donde existen cambios de jurisdicciones provinciales o contratos de concesión del mantenimiento del mismo. También es habitual encontrar caminos secundarios de ripio en barrios alejados o en construcción que se comunican con caminos pavimentados. La transición entre uno y otro puede causar confusión para algunos sistemas.

- *Condiciones climáticas y ambientales cambiantes* - Los sistemas deben enfrentar situaciones como lluvias, nieve, niebla, polvo, humo y también cambios en la posición e inclinación del sol para los distintos horarios del día. Estos escenarios modifican considerablemente la apariencia de los objetos y del camino.

- *Diferentes tipos de caminos y estructuras* - Si bien es posible diseñar sistemas para ambientes específicos y controlados, un sistema preparado para funcionar en ambientes urbanos o rurales se puede encontrar con distintos tipos de caminos que difieren en apariencia, forma, y cuya estructura y alrededores es muy variable. Vale la pena aclarar que aquí no se refiere al cambio del tipo de camino durante la navegación, que fue considerado más arriba.

- *Vehículos u obstáculos que ocluyen el camino* - Todos aquellos elementos que impiden la correcta o completa visibilidad del camino y/o sus límites pueden dificultar la identificación del mismo. Lo más probable es que el resto de los vehículos que forman parte de la escena, ya sea en movimiento o estacionados a un costado, ocluyan parte del camino y hagan que la información que se obtiene de la imagen termine siendo incompleta.

- *Vibraciones y movimientos de la cámara* - Aunque no es uno de los problemas más importantes, las vibraciones de la cámara entre fotogramas consecutivos de video y el rastro de desenfoque dejado por los objetos en movimiento pueden traer algunas complicaciones menores.

Para analizar los algoritmos respecto a estas problemáticas se han separado en dos grupos según el modelo utilizado. Los modelos basados en candidatos como pixeles, *patchs*, o *clusters* se denominaron como *modelos de apariencia*, ya que utilizan principalmente el color y/o la textura para identificar si el candidato se parece o no al camino. Por otro lado, los sistemas que suponen algún tipo de forma o geometría para el camino se denominan *modelos geométricos*. En la Tabla 3.2 se analizan los efectos de las distintas problemáticas sobre cada uno de los modelos y se describen algunos mecanismos para atenuarlas. También se incluye una valoración acerca del rendimiento general del sistema ante la problemática ("$\surd$" para indicar un rendimiento positivo, "$\bigotimes$" para uno negativo o "$\sim$" cuando es dudoso y depende de cada caso).

Tabla 3.2: Análisis de los efectos causados por las distintas problemáticas existentes sobre los modelos de apariencia y los modelos geométricos. Se presentan además algunas propuestas para atenuar o disminuir su efecto. El símbolo "$\surd$" indica una buena performance del sistema ante el problema, mientras que "$\otimes$" indica una performance insuficiente, y "$\sim$" indica que el caso es dudoso y dependerá de cada situación y sistema particular.

Problema	**Modelos de Apariencia**	**Modelos Geométricos**
Cambios drásticos de iluminación.	La distribución de los colores puede cambiar bruscamente cuando hay cambios repentinos en la iluminación de la imagen o cuando hay reflejos de luz directos hacia la cámara, afectando gravemente a un sistema basado en los colores del camino. Para atenuar su efecto algunos sistemas transforman las componentes de color hacia determinados espacios de colores o características que son más inmunes a la variación de la intensidad. Si es posible estimar de alguna manera como han variado los parámetros de la cámara entre una imagen y la siguiente también sería posible aplicar algún tipo de transformación a la distribución de colores del camino que se arrastra de imágenes anteriores. Valoración: $\otimes$	La información que se integra de una imagen a otra es la distribución de los parámetros del modelo geométrico y no una distribución de colores. En este sentido, los cambios de iluminación no afectan gravemente al sistema. Además, si las características que se utilizan para evaluar un candidato son medidas de contraste entre las regiones que están adentro y las que están fuera del camino, el método es aún más robusto, ya que ambas regiones están afectadas de igual manera por el cambio. Valoración: $\surd$
Zonas de muy alta y muy baja intensidad en una misma imagen. Pérdida de información del color.	Los colores se perciben diferentes en zonas más oscuras complicando la comparación con otras zonas de la imagen. La información de color de los pixeles sobreexpuestos y los muy oscuros no debería considerarse por no ser confiable. Para esto debería preprocesarse la imagen para detectar estas zonas y considerarlas mediante algoritmos de más alto nivel. La utilización de espacios de color o características con mayor inmunidad a la variación de intensidad pueden también ser útiles. Por último, los procesamientos localizados o por regiones de interés podrían mejorar la detección en aquellas zonas de poco contraste. Valoración: $\otimes$	Dado que la evaluación de cada curva/candidato involucra muchos pixeles (podría ser la imagen entera), estos problemas afectan poco al sistema. En casos algo extremos, la generación de falsos contrastes podría ser problemática. De igual manera que para los modelos de apariencia, los pixeles con información poco confiable deberían detectarse tempranamente y no utilizarse en la evaluación. Valoración: $\surd$

Continúa en la página siguiente

Tabla 3.2 – continuación

Problema	**Modelos de Apariencia**	**Modelos Geométricos**
Sombras, marcas y manchas en el camino. Colores en común entre el camino y lo que no lo es.	El problema principal es que la apariencia de las zonas sombreadas o manchadas es diferente a la del camino por lo que el sistema puede confundirlas con obstáculos. Las sombras en particular también suelen traer muchísimos inconvenientes cuando se utilizan medidas de gradientes y contrastes. En este caso puede ser posible algún tipo de procesamiento para su detección temprana y su posterior evaluación a más alto nivel, aunque no es sencillo. Por otro lado, aquellas zonas que tengan apariencia similar al camino sin duda se confundirán con él. Estos problemas podrían atenuarse mediante un análisis localizado en diferentes partes de la imagen para hallar continuidades en bordes o líneas del camino cuando se pasa de una zona sombreada a una que no lo es, o viceversa. Valoración: $\otimes$	Idem caso anterior. Valoración: $\surd$
Cambios de apariencia del camino.	Estos sistemas se basan principalmente en la premisa de que la apariencia del camino varía muy poco entre dos imágenes consecutivas. Cuando la superficie o las condiciones del camino cambian abruptamente este tipo de sistemas generalmente deja de funcionar. En algunos casos, es posible readaptarse a la nueva situación mediante métodos de aprendizaje autosupervisado, aunque durante el transitorio el sistema puede generar estimaciones del camino insuficientes. Mediante la utilización de procesamientos locales debería ser posible detectar una continuidad en los límites del camino. Hacen falta algoritmos de más alto nivel que puedan determinar el tipo de características que conviene utilizar ante diferentes situaciones. Valoración: $\otimes$	Si las características utilizadas para evaluar el candidato son medidas de contraste entre la región contenida por el camino y la que no, el sistema no es afectado por estos cambios. Valoración: $\surd$
Condiciones climáticas y ambientales cambiantes.	La lluvia suele afectar considerablemente la apariencia de los objetos, empeorando además la situación respecto de los reflejos de la luz hacia la cámara. La niebla, el polvo y el humo pueden degradar el contraste de las imágenes hasta el punto de inutilizar algunos sistemas. Se pueden utilizar procesamientos específicos para mitigar el efecto de estos problemas pero en general los sistemas se ven muy afectados. Valoración: $\otimes$	Si las condiciones ambientales son severas los sistemas de este tipo también se ven afectados, principalmente en la estimación del camino a distancias lejanas donde por ejemplo la niebla y el polvo destruyen de alguna manera la información de la imagen. En algunos casos más leves, si las condiciones de contraste dentro y fuera del camino se mantienen, estos sistemas pueden adaptarse. Valoración: $\sim$
Continúa en la página siguiente		

Tabla 3.2 – continuación

Problema	**Modelos de Apariencia**	**Modelos Geométricos**
Diferentes tipos de caminos y estructuras.	Si las características elegidas para evaluar al candidato son suficientemente robustas y no dependen de un caso específico, estos sistemas pueden funcionar sin inconvenientes. Se pueden adaptar a cualquier forma ya que la clasificación es por grupos de pixeles. Valoración: $\surd$	De igual manera que los modelos de apariencia, estos sistemas pueden funcionar para caminos con diferentes apariencias. Sin embargo, no pueden adaptarse a formas que no son contempladas por el modelo geométrico asumido. Comunmente los modelos de curvas utilizados fallan en detectar y representar situaciones comunes como por ejemplo bifurcaciones, intersecciones y rotondas. Es necesario algún mecanismo de alto nivel que pueda redefinir el modelo de camino utilizado según el caso que se presente. Valoración: $\otimes$
Vehículos u obstáculos que ocluyen el camino.	Salvo en aquellos casos en los que la apariencia del vehículo/obstáculo sea muy similar a la del camino, en general no hay inconvenientes considerables. Además, como no hay una forma predefinida para el camino, los algoritmos pueden agregar o quitar una zona de la imagen sin afectar al sistema. La utilización de algoritmos para la detección de obstáculos mediante visión u otra modalidad de sensado pueden complementar al sistema dándole mayor robustez. Valoración: $\surd$	Estos sistemas por definición no incluyen a los obstáculos dentro del modelo geométrico. La existencia o no de un obstáculo puede pesar mucho a la hora de evaluar un candidato, debido principalmente a los contrastes que éstos pueden generar en la imagen. La previa detección de obstáculos juega aquí un papel más importante que en los modelos de apariencia. Valoración: $\otimes$
Vibraciones y movimientos de la cámara.	El nivel de difuminado o desenfoque producido por el movimiento en general no es suficiente para que la apariencia del camino se vea afectada (salvo casos extremos). Las vibraciones y corrimientos no tienen efecto alguno. Valoración: $\surd$	De igual manera, el desenfoque no presenta complicaciones en condiciones normales. El corrimiento de una imagen debido a movimientos bruscos de la camara por desniveles del terreno podría traer problemas a la hora de integrar la información de los parámetros del camino proveniente de imágenes anteriores. Valoración: $\sim$

Para la mayoría de las problemáticas planteadas, los sistemas basados en modelos geométricos parecen tener un mejor rendimiento que aquellos basados en modelos de apariencia. Estos últimos sufren principalmente cuando cambian las condiciones de iluminación de la escena y cuando aparecen sombras o zonas sobreexpuestas intensas. Si bien la utilización de cámaras con mayor rango dinámico y con filtros especiales pueden mejorar el rendimiento, aquí se han analizado las propuestas respecto al procesamiento de la imagen y no respecto a las configuraciones del sensor en sí. En la literatura en general se propone utilizar determinados espacios de colores que son más inmunes a las variaciones de intensidad [Sotelo et al. (2004), Tue-Cuong et al. (2008)] o bien aplicar transformaciones especiales a la información de color [Ghurchian & Hashino (2005), Alvarez et al. (2008)].

En cambio, los sistemas que asumen un modelo geométrico han mostrado ser muy robustos a las variaciones en la condiciones de iluminación (ver por ejemplo [Manz et al. (2010)]). También se comentó aquí que son robustos ante *outliers* y datos faltantes. Sin embargo, su principal limitación radica justamente en aquello que le brinda las ventajas mencionadas: el modelo geométrico. La suposición de que el camino tiene una determinada forma geométrica imposibilita que el sistema puede adaptarse a situaciones que este modelo no contemple. Los modelos de curvas comunmente utilizados dejan de funcionar en casos como intersecciones de diferentes tipos, bifurcaciones, y rotondas, entre otros. En la Sección 3.3.2 se profundiza un poco más acerca de la detección de topologías no lineales para los caminos. La existencia de vehículos u obstáculos que ocluyan la visión de los bordes del camino tampoco está contemplada, por lo que en algunos casos pueden provocar errores groseros en la estimación de los parámetros del modelo. [Chapuis et al. (2002)] por ejemplo utiliza procesamientos en zonas locales para aumentar la robustez frente a las oclusiones y las imperfecciones en las marcas del camino.

Uno de los desafíos de mayor complejidad es que todos los sistemas sean funcionales en diferentes condiciones climáticas y en cualquier horario del día. Tanto los modelos de apariencia como los geométricos en general no logran adaptarse aún a estas exigencias. Es posible sin embargo diseñar algunas estrategias para mitigar los efectos de condiciones climáticas adversas como la lluvia, la niebla y el polvo, tal como han demostrado por ejemplo [Peynot et al. (2009), Hautière et al. (2010)].

Concluyendo este análisis se puede decir que ambos tipos de sistemas de alguna forma se complementan, ya que donde uno de los enfoques tiene debilidades el otro tiene fortalezas. Se ha mencionado en varias oportunidades la necesidad de utilizar métodos específicos para la detección temprana tanto de sombras como de zonas donde el color no es confiable para su evaluación posterior mediante algoritmos de más alto nivel. También puede ser necesario en muchos casos el procesamiento localizado de la imagen. Justamente estos dos tipos de procesamiento tienen mucho en común con la manera en que los humanos suelen procesar visualmente una escena. Esto refuerza la idea subyacente por la que en esta tesis se pretende estudiar como los humanos perciben aquello transitable.

Por último, se quiere hacer énfasis en que el diseño de un sistema basado solo en visión que sea robusto ante todas las problemáticas presentadas es una tarea muy compleja

y que los investigadores aún no han resuelto. Por este motivo, siempre que la aplicación y los recursos lo permitan, puede ser muy ventajoso utilizar distintos tipos de sensores para combinar lo mejor de cada uno en sistemas más estables y más robustos. La información obtenida a través de otras modalidades de sensado puede en ocasiones reducir enormemente el costo computacional que implica obtener el mismo tipo de información a través de algoritmos de visión monocular. En [Moreyra & Masson (2010)] por ejemplo se han realizado experimentos que ilustran la potencialidad de una simple combinación de visión e información de rango provista por un sensor láser para diferenciar aquellas superficies que en apariencia parecen indistinguibles pero que en realidad pertenecen a objetos diferentes.

3.3.2. Topologías no lineales para representar un camino

Al manejar, las personas enfrentan constantemente situaciones que difieren unas con otras y que difícilmente pueden aproximarse con modelos simples. Si bien los típicos modelos de dos líneas rectas o curvadas han demostrado ser funcionales durante la mayor parte de los recorridos en autopistas o rutas pavimentadas, inclusive en caminos secundarios sin asfalto, es necesario modificarlos o cambiarlos para poder adaptarse a escenarios como intersecciones y bocacalles, así como también a bifurcaciones, uniones y rotondas que son muy comunes llegando a zonas suburbanas y urbanas. A este tipo de topología de caminos se la denominará *no lineal*, ya que no puede modelarse con un solo par de curvas o rectas.

Teniendo en cuenta que la utilización de la visión para guiar vehículos se ha investigado durante muchos años, sorprende que estos casos hayan sido tan poco estudiados. Sin embargo, hay algunas propuestas que han intentado modelar algunas de las situaciones "no lineales" que se mencionaron en el párrafo anterior. [Crisman & Thorpe (1993)] ha propuesto un método para detectar intersecciones del tipo Y o λ utilizando una clasificación de colores del tipo bayesiana para medir la correspondencia del camino actual con una librería de candidatos. Si bien se propone un mecanismo para reducir el espacio de búsqueda de los candidatos, el sistema es computacionalmente costoso. Se asume casi idealmente que el camino difiere en colores de su entorno y que los caminos son localmente rectos, planares y de ancho constante. En [Jochem et al. (1995), Jochem et al. (1996)] se utiliza lo que los autores llaman *cámaras virtuales* para la búsqueda de

intersecciones en la imagen. Las cámaras virtuales son cámaras imaginarias que pueden ubicarse en cualquier posición respecto del vehículo y proveen una imagen virtual del ambiente a partir de la correspondiente transformación geométrica de la imagen original. La detección del camino en sí está basada en un sistema llamado ALVINN que fue propuesto en [Pomerleau (1992)], y que utiliza una red neuronal entrenada para mapear la información obtenida en la imagen directamente hacia un ángulo de avance del vehículo de manera de guiarlo dentro del camino. Si a priori se define una ubicación y una orientación de la cámara virtual respecto del vehículo, sería posible buscar aquellos caminos que se encuentren en dicha dirección. El sistema entonces trabaja suponiendo una determinada topología del camino y ubica diferentes cámaras virtuales de forma de que el sistema indique cuando ha detectado un camino en cada sensor virtual. Las desventajas que tiene este enfoque es que no es posible a priori saber que forma tendrá el camino ni cuando cambiará, y que las transformaciones geométricas dependen de que el terreno sea totalmente plano y que los parámetros de la cámara se conozcan sin errores. El sistema ALVINN además no permite extraer información acerca del camino ya que la salida de la red neuronal es una señal de control para el vehículo. En el caso de [Ekinci & Thomas (1996)] se propone un método para detectar intersecciones del tipo $\dashv$, $\vdash$ o similares que se activa con la asistencia de un mapa almacenado. Con ayuda del mapa y según la información de las imágenes anteriores, se identifican los lugares de la imagen donde deberían encontrarse los bordes del camino pero que en realidad se detecta una discontinuidad. A partir de estos puntos se utiliza información de intensidad y de bordes para segmentar los caminos que se unen al principal y ajustar modelos de rectas para cada uno de ellos. El trabajo presentado por [Heimes & Nagel (1998)] también utiliza la información de un mapa para ajustar en tiempo real un modelo fijo de intersección urbana. Los autores utilizan el seguimiento de dicho modelo en la imagen para complementar la información de otros sensores y mejorar la estimación de la posición y orientación del vehículo dentro del camino. Por otro lado, el sistema presentado por [Lutzeler & Dickmanns (2000)] puede ajustar modelos de intersecciones del tipo $\top$ o λ, aunque no soporta cruces de caminos. El sistema conoce de antemano la ubicación aproximada de la intersección e interactúa con un sistema de control activo de la dirección de visión para controlar el vehículo mientras transita la intersección.

Mucho más reciente es el trabajo de [Manz et al. (2011)], donde se propone un

método que utiliza tanto un modelo de un solo camino como un modelo de intersección de caminos, que alternan según la situación que se presente. Este último permite modelar por ejemplo cruces, bifurcaciones y algunos tipos de intersecciones a partir de un modelo de ramas basado en un filtro de partículas. El cambio de modelo geométrico se realiza con la asistencia de un GPS, un láser 3D y un sistema de información geográfica (GIS). La localización del vehículo en el mapa permite predecir automáticamente a que distancia cambiará la forma del camino y que tipo de topología tendrá, de tal forma que el modelo geométrico predicho sea siempre el apropiado. Los muy buenos resultados que los autores muestran son obtenidos en ambientes poco estructurados por lo que el enfoque está muy relacionado con el problema que esta tesis pretende analizar.

A criterio del autor de esta tesis, este último trabajo propone un método robusto y que se logra adaptar a varias topologías. Sin embargo, la selección del modelo geométrico se realiza directamente a partir de un mapa preciso del ambiente. En muchas regiones no es posible disponer de un mapa del lugar, por lo que este método sería inaplicable. Esto evidencia la necesidad de que los sistemas puedan reconocer una situación solo a partir de los sensores que lleva el vehículo consigo. En el caso de la visión, ésto se traduce a algoritmos con la inteligencia suficiente para comprender los estímulos visuales de la escena que se presenta en una imagen para proponer un modelo adecuado.

3.3.3. Hacia una verdadera comprensión de la escena

Tal como se ha discutido hasta ahora, en general, los sistemas para la detección de caminos mediante visión se diseñan bajo fuertes supuestos acerca de las condiciones del camino y del ambiente. A pesar de que los investigadores han trabajado en el diseño de estos sistemas por más de 20 años, la mayoría de los enfoques que se encuentran en la literatura carecen de suficiente capacidad de adaptación para mantenerse funcionales cuando la situación que se enfrenta difiere de la ideal. Aún los sistemas más maduros, que ya muestran muy buenos resultados para algunos tipos de situaciones, no son suficientemente abarcativos para cubrir todas las problemáticas que normalmente se le presentan a una persona que maneja su vehículo. Los sistemas de asistencia al conductor modernos por ejemplo tienen una gran dependencia de la estructura de los caminos para los que se los diseña y no están preparados para asistirlo durante todo el trayecto que éste realiza cotidianamente desde que sale de su casa hasta que llega a su trabajo.

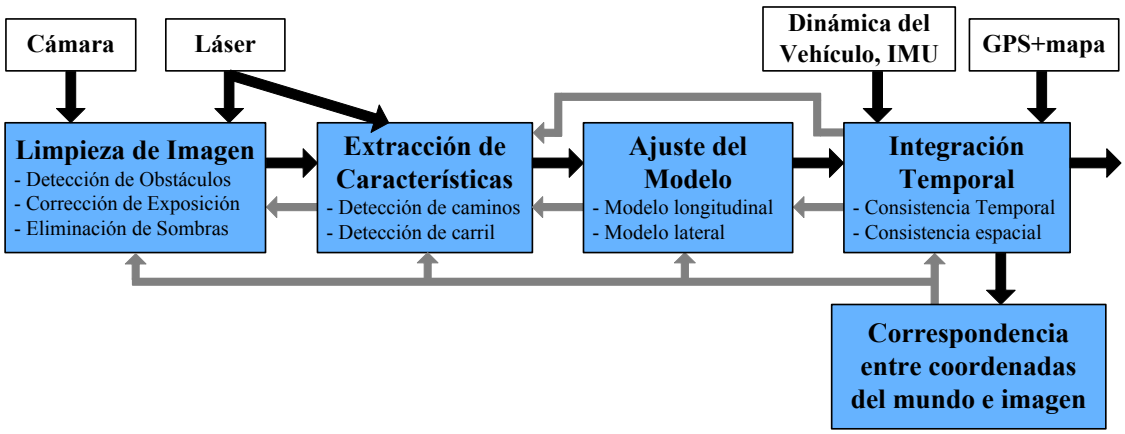


Figura 3.6: Descomposición funcional de un sistema genérico para la detección de caminos presentado por [Bar Hillel et al. (2012)]. Esta figura es una reproducción en español de la figura original.

En [Bar Hillel et al. (2012)] se presenta un muy interesante análisis del estado del arte de los sistemas basados en visión para la detección de caminos. Uno de los enfoques utilizados por los autores se basa en descomponer las partes de un sistema genérico en bloques según las funciones de cada una, tal como se reproduce en la Fig. 3.6. Los bloques identificados reflejan un razonamiento de alguna forma similar al que se ha presentado en el proceso de clasificación de la Fig. 3.4. La diferencia más importante quizás es que el significado de lo que aquí se ha denominado "candidato" es más abarcativo que el "modelo" allí utilizado, ya que considera también la clasificación por pixeles, *patchs* y *clusters*. Los algoritmos de "limpieza de la imagen" no se han incluido como un bloque del proceso pero sí se han considerado muy importantes tanto la detección previa de sombras como la de obstáculos y zonas sobreexpuestas de la imagen. De igual forma con la utilización de otro tipo de sensores para complementar el sistema.

Tanto el enfoque de [Bar Hillel et al. (2012)] como el aquí propuesto han sido útiles para describir de forma general la gran mayoría de las propuestas que se pueden encontrar en la literatura de la temática. Sin embargo, estos modelos genéricos no contemplan de forma explícita mecanismos concretos que permitan identificar de forma autónoma qué características o modelos conviene utilizar según la situación que se presente. Son pocos los sistemas que tienen por ejemplo la capacidad de medir el nivel de confiabilidad de la estimación del camino que están entregando. Es deseable que un sistema pudiera

alertar al usuario o bien cambie de estrategia cuando la confiabilidad de la estimación es insuficiente para la aplicación. Esto es, que sea capaz de reconocer cuando las condiciones han cambiado. Estos cambios se traducen principalmente a cambios en la apariencia del camino, es decir, en las características que lo definen, o a cambios en la forma o geometría del mismo.

En este punto, se entiende que para obtener sistemas con gran capacidad de adaptación es necesario incluir algoritmos de alto nivel que permitan primero analizar y comprender la situación presente en la escena antes de ajustar cualquier modelo. En la Fig. 3.7 se ha modificado el sistema genérico propuesto por [Bar Hillel et al. (2012)] para incluir un bloque denominado *Análisis y Comprensión de la Situación* que utiliza tanto la información proveniente de la imagen como aquella que se obtiene de la integración temporal. Esta última incluye el estado actual del vehículo y la estimación del camino. En función de la situación que se identifique el sistema puede determinar el modelo y las características que mejor se adapten. Para representar esto se incluyeron los bloques *Selección de Características* y *Selección del Modelo*. Los bloques originales de la figura se mantuvieron en color azul mientras que los nuevos se identificaron con un color naranja claro.

En la literatura casi no existen registros de trabajos que propongan algoritmos de procesamiento de imágenes para la selección automática tanto de las características como del modelo geométrico utilizados para la detección del camino. Sin embargo, se pueden encontrar algunas aproximaciones o soluciones parciales.

Respecto del primer caso, existen algunas propuestas en la bibliografía que utilizan subsistemas basados en diferentes características que trabajan en serie o en paralelo para entregar la mejor estimación posible. En [Apostoloff & Zelinsky (2003)] se presenta un método basado en el filtro de partículas para fusionar dinámicamente múltiples características. Constantemente se evalúa la utilidad de cada característica a lo largo del tiempo y se seleccionan aquellas cuyo rendimiento es óptimo. El resto de las características se continúan procesando en segundo plano y a una velocidad menor a la deseada, reservando la mayor cantidad de recursos de cómputo para la estimación principal. Cuando el monitoreo de estas características secundarias detecta que alguna de ellas puede contribuir a mejorar la estimación global del sistema se las agrega al conjunto primario. En [Rasmussen et al. (2009)] se realiza un procedimiento algo similar. Para cada imagen

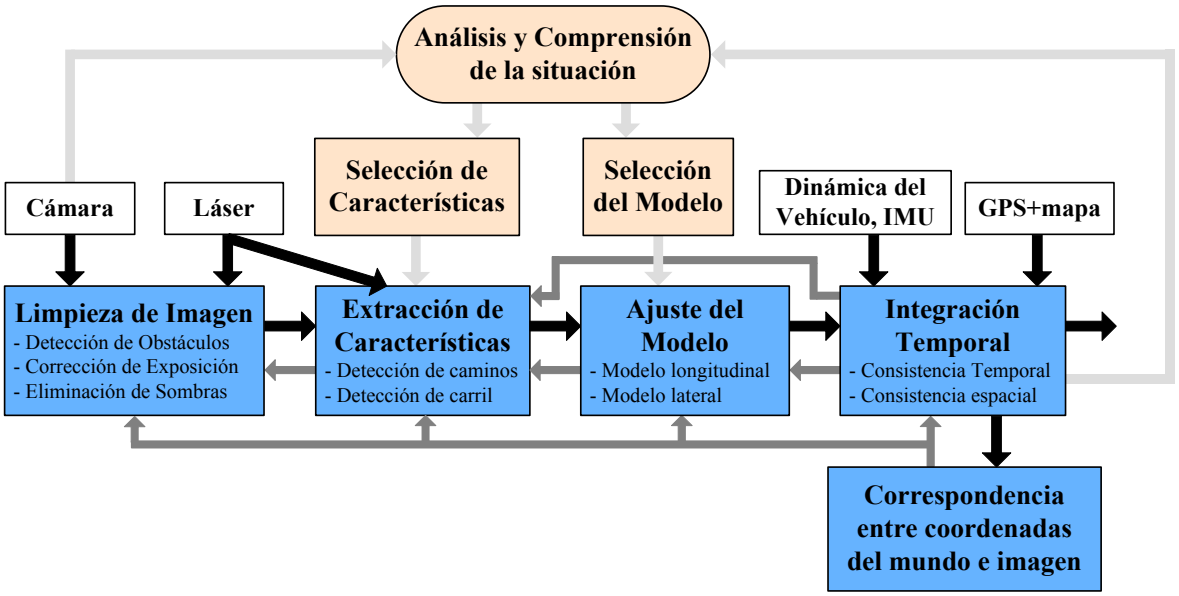


Figura 3.7: Diagrama de bloques de un sistema genérico para la detección de caminos que se ha modificado para incluir funciones de alto nivel como el análisis y comprensión de la situación y la selección de las características y el modelo utilizado para el ajuste. El modelo original pertenece a [Bar Hillel et al. (2012)] .

de una secuencia se realiza un mismo preprocesamiento utilizando una, dos o las tres componentes de color en el espacio CIE LAB. Si bien no se corren tres filtros por separado, cada cierta cantidad de iteraciones se generan tres grupos gemelos de partículas y se las evalúa mediante los correspondientes preprocesamientos para determinar cual de ellos genera la mayor verosimilitud. Las partículas asociadas con este grupo son las únicas que se propagan con el filtro. Por otro lado, [Alon et al. (2006)] incorpora dos modalidades diferentes de detección que corren en paralelo, siendo la salida del sistema gobernada por el módulo que tiene la mayor confianza en la estimación. Por un lado se utiliza un método de clasificación de la textura de la imagen, mientras que por el otro se realiza la búsqueda de los bordes del camino mediante la transformación a las coordenadas del mundo, suponiendo que los límites del camino son paralelos y que el terreno es totalmente plano. Previamente, se estiman los parámetros de la cámara mediante visión estereoscópica.

Se puede observar que los sistemas que utilizan filtros bayesianos, como el de partículas, facilitan de alguna manera la utilización de múltiples características extraídas

de la imagen, al mismo tiempo que favorecen a una mejor estimación del camino. Por naturaleza, los valores de verosimilitud de cada partícula pueden ser utilizados para medir cuan bueno es el ajuste de un modelo a una situación determinada. Además, por la naturaleza del filtro también se facilita la fusión de datos con otro tipo de sensores. Esto se suma a las ventajas comentadas anteriormente sobre la integración temporal de la información y el rechazo a las mediciones erróneas.

Respecto de la selección automática del modelo geométrico, las referencias disponibles son aún más escasas. Cabe señalar en este punto que la estimación de los parámetros de un modelo predefinido no se considera como una selección de un modelo, aunque el conjunto posible de parámetros genere por ejemplo muchas curvas diferentes. En [Rasmussen (2003)] se propone uno de los primeros algoritmos para la detección visual de una intersección sin utilizar ningún conocimiento a priori, que puede ser provisto por ejemplo por un mapa. Allí se ha descrito un sistema para la detección de intersecciones del tipo $+$, $\top$ y $\rceil$, y sus respectivas rotaciones, a través del análisis de la forma resultante de una segmentación por colores de una imagen panorámica obtenida con múltiples cámaras. La clasificación de las formas se realiza en las imágenes segmentadas transformadas a coordenadas del mundo mediante un método de aprendizaje supervisado. Éste es entrenado con secuencias de imágenes capturadas en un barrio residencial, donde los caminos son asfaltados y en general muestran bastante contraste con las partes que no son transitables. La aplicación del sistema parece bastante acotada a las circunstancias para las cuales éste ha sido entrenado. En [Lombardi et al. (2005)] se consideran tres modelos distintos de caminos: curva a la izquierda, curva a la derecha, y camino recto. Estos modelos están prefijados como imágenes binarias y se seleccionan probabilísticamente en función de la textura de la imagen. Si bien el enfoque es aún más limitado, se puede clasificar como un intento de selección automática de un modelo. En [Bai et al. (2010)] se presenta una metodología para estimar caminos que son difíciles de ajustar con modelos geométricos de hipérbolas o clotoides, a partir de la descomposición en una secuencia de formas más simples que se unen en determinados puntos llamados de transición (donde se producen cambios considerables en la curvatura). En lugar de utilizar un solo filtro para el seguimiento del camino completo, se realiza un cambio de modelo en los puntos de transición, obteniéndose un método más robusto y que se ajusta mejor a las condiciones del camino. No es posible con este sistema afrontar topologías no lineales.

Por otro lado, [Hummel et al. (2008)] introduce un nuevo enfoque basado en la lógica descriptiva para la representación formal del conocimiento necesario para la comprensión de escenas de carreteras e intersecciones complejas a partir de sensores. Se utilizan los datos de video provenientes de una cámara sobre un vehículo y un mapa digital como evidencia para una intersección particular. La observabilidad parcial de los datos y su pertenencia a diferentes niveles de abstracción se pueden manejar naturalmente dentro del marco formal presentado. Este trabajo es uno de los pioneros en presentar un marco de razonamiento y representación del conocimiento para tareas de alto nivel como la interpretación de escenas. Si bien no presenta resultados experimentales, el enfoque es muy interesante y refuerza la idea que aquí se plantea acerca de que el diseño de sistemas verdaderamente robustos para la detección de caminos necesita complementarse con algoritmos de más alto nivel de procesamiento.

Muy recientemente los autores de [Chiku & Miura (2012)] presentaron un método para la detección y el seguimiento de caminos con diferentes topologías, que tiene la capacidad de alternar diferentes modelos en el marco de estimación de un filtro de partículas. El sistema aproxima los bordes del camino mediante modelos lineales a trazos (*piece-wise linear (PWL)* en inglés) y utiliza gradientes de color, de intensidad y de altura (visión estéreo) como evidencia para evaluar los diferentes modelos que cada partícula propone. Este tipo de aproximación permite que el sistema pueda adaptarse a caminos de diferentes formas aunque podría no ser suficientemente preciso para determinadas aplicaciones. Dado que las aproximaciones se realizan por tramos es posible detectar aquellos segmentos cuyo grado de ajuste es pobre para inferir que el modelo propuesto para esa región no es el más adecuado. De esta forma mediante el seguimiento y análisis de la verosimilitud de las partículas y sus tendencias tanto en el tiempo como en el espacio, es posible detectar ciertos "signos" de falta de ajuste del modelo actual. En función de estos signos se activa un mecanismo de transición entre modelos que incluye por ejemplo el agregado de partículas del nuevo modelo a las ya existentes.

La propuesta que realiza este último trabajo es la que más se aproxima al problema que se propone estudiar en esta tesis. El sistema descrito se ha mostrado robusto frente a los cambios en la forma de los caminos utilizados sin la utilización de un conocimiento a priori de las topologías que el vehículo debía transitar. Aunque los modelos propuestos no incluyen topologías del tipo $\top$ o λ (donde el camino principal llega a un fin), seguramente

en un futuro cercano éstos se extenderán para incluir dichas situaciones. Una de las desventajas que tiene el enfoque es que depende de la existencia de un mínimo nivel de contraste entre el camino y el resto del terreno u obstáculos. En general, en ambientes poco estructurados los caminos se caracterizan por una existencia difusa de los bordes del mismo y sin suficiente contraste de color entre ambas zonas. El análisis de la orientación de la textura de la imagen podría quizás incluirse de alguna manera para robustecer al sistema. Otro de los potenciales problemas que tiene el sistema es que los cambios de modelo se estabilizan de forma aceptable cuando la distancia respecto del vehículo es bastante reducida, lo que obligaría a reducir la velocidad del mismo si su trayectoria depende de la topología del camino.

3.4. Resumen

Hasta aquí se describieron los procesos básicos que se dan en el cerebro para el reconocimiento de estímulos visuales y cómo influye el conocimiento a priori del objeto de interés en dichos procesos. Se han estudiado y analizado con gran profundidad cómo los sistemas automáticos para la detección de caminos han ido evolucionando hacia sistemas cada vez más robustos. Sin embargo, también se han descrito las limitaciones actuales y los desafíos que aún restan por alcanzarse, subrayando la necesidad de nuevos algoritmos y nuevas estrategias para una verdadera comprensión de la escena que se presenta en la imagen. En lo particular, este trabajo está enfocado hacia el diseño de algoritmos de visión monocular que permitan determinar a mediano y largo alcance la topología de un camino en ambientes poco estructurados.

Dado que las personas son totalmente capaces de resolver esta tarea mientras conducen su vehículo, aquí se propone estudiar y aprender como los humanos inspeccionan visualmente una imagen con escenas estáticas similares para determinar la topología de un camino. En el capítulo siguiente se describe en detalle el experimento diseñado para dicho propósito y se analizan los resultados y sus implicancias.

Capítulo 4

Patrones de atención visual para la percepción de caminos con topologías no lineales

Las limitaciones que tienen los sistemas actuales para la detección de caminos mediante visión empujan permanentemente a los investigadores hacia el diseño y desarrollo de nuevos algoritmos que permitan comprender mejor la situación que se presenta en una escena. Dada su gran capacidad de adaptación y su eficiente utilización de los recursos de procesamiento, el sistema visual humano siempre es motivo de inspiración para cualquier sistema basado en imágenes. En este sentido, el estudio y el análisis de los patrones de atención visual de las personas que aquí se presenta pretende ser un puntapié inicial para el futuro diseño de estrategias para la detección automática de topologías no lineales de caminos.

El capítulo se organiza de la siguiente manera: en la Sección 4.1 se introduce el concepto de atención visual selectiva y como se manifiestan los cambios en el foco de atención a través del movimiento de los ojos; en la Sección 4.2 se enumeran las hipótesis del experimento y cuáles son los antecedentes en la literatura que se relacionan con el mismo; en la Sección 4.3 se describe en detalle todo el método experimental utilizado, incluyendo el material, el equipamiento, y el grupo de personas que participaron; en la Sección 4.4 se presentan los algoritmos utilizados para preprocesar la imagen antes del correspondiente análisis de los resultados; en la Sección 4.5 se reportan todos los resultados obtenidos; y por último, en la Sección 4.6 se comenta acerca de sus implicancias.

4.1. La atención visual selectiva

Como se ha visto en el capítulo anterior, los sistemas diseñados para la percepción del área transitable mediante visión suelen adolecer de una gran dependencia de la

estructura de los caminos y no siempre están preparados para enfrentar las condiciones reales de uso como son los cambios drásticos en la iluminación, las condiciones pobres de visibilidad y una gran variedad de caminos e infraestructuras, entre otras cosas. Todas estas limitaciones advierten acerca de la necesidad de sistemas que sean capaces de reconocer cuando un modelo geométrico ya no es válido o cuando las características de la imagen necesarias para su ajuste pierden vigencia, haciendo necesario su reemplazo por otras más acordes. En otras palabras, se requieren nuevas herramientas o estrategias que permitan alcanzar un mayor entendimiento o una mayor comprensión de la escena.

Por otro lado, el sistema visual humano ha demostrado siempre ser lo suficientemente robusto y flexible como para permitir por ejemplo que un conductor experimentado pueda manejar con éxito en condiciones extremas. Este tipo de capacidades junto al aprovechamiento óptimo de los recursos de cómputo, han atraído la atención de los investigadores hacia el desarrollo de nuevas estrategias y nuevos sistemas inspirados en la visión humana [Frintrop et al. (2010)]. En este caso, se hará énfasis en el estudio de los patrones de atención visual de las personas mientras se intenta reconocer la topología no lineal de un camino con poca estructura en diversas fotografías. Dentro de los alcances del conocimiento del autor, es la primera vez que se realiza un experimento de este tipo por lo que constituye uno de los aportes más importantes de esta tesis. De esta forma, se pretende obtener evidencia experimental acerca de aquellos elementos visuales que utilizan las personas para resolver el problema planteado, y a partir de un análisis profundo y sin sesgo determinar los elementos fundamentales que permitan el diseño de nuevos algoritmos de procesamiento de imágenes para la detección automática de topologías de caminos.

Antes de continuar será necesario repasar el concepto de *atención visual selectiva* que fue introducido en la Sección 2.1. Es sabido que cuando se lee, se observa una escena, o se busca un objeto, los ojos se mueven continuamente de un punto a otro para obtener la información que se necesita. Estos movimientos producidos por los músculos del ojo, junto a los movimientos de la cabeza, permiten compensar la falta de resolución en gran parte del campo visual alineando aquellos objetos de interés directamente hacia la fóvea. Se debe recordar que la agudeza visual del ojo humano cae abruptamente con la excentricidad, es decir, con la distancia angular al punto de la fijación, que se encuentra justamente sobre el eje foveal. El mecanismo a través del cual el cerebro determina que

parte del campo visual es la de mayor interés se denomina *atención visual selectiva*. Es un proceso cognitivo por el cual el sistema visual se concentra selectivamente en un determinado estímulo ignorando al resto, enfocando todos sus recursos de procesamiento a atenderlo. Dado que dichos recursos son limitados este proceso se realiza de forma serial o secuencial.

Los movimientos de gran velocidad y aceleración que realizan los ojos se denominan *sacádicos*. Entre sacádicos, los ojos permanecen relativamente quietos durante períodos cortos de tiempo que se denominan *fijaciones*. Aunque generalmente se asocia el foco de atención con estos movimientos, en lo que se denomina la *atención manifiesta*, el sistema visual también es capaz de enfocarse en determinados lugares de la periferia sin que los ojos muestren movimiento alguno. Este último fenómeno se llama *atención encubierta* y es un mecanismo masivamente paralelo que comúnmente precede al manifiesto. Por este motivo en ocasiones se lo llama preatencional, aunque existe evidencia de que no siempre es condición necesaria para un sacádico [Frintrop et al. (2010)]. Pero normalmente la atención encubierta y los movimientos sacádicos trabajan en conjunto, direccionándose primero el foco atencional hacia una región de interés, generándose luego un sacádico que permita realizar una fijación en dicho lugar para extraer la información de alta resolución disponible.

El estudio de la atención visual a través de equipos para el seguimiento de ojos permite analizar de forma muy directa los patrones de atención manifiesta. Los corrimientos de atención encubierta son sin embargo mucho más difíciles de detectar y se necesita diseñar experimentos que de alguna manera hagan manifiesto estos "movimientos" ocultos. A esto se suma la complejidad de que el movimiento de los ojos refleja momento a momento procesos cognitivos que se dan en el cerebro y que pueden solaparse con los procesos particulares de la tarea que se está resolviendo, enmascarando los movimientos que verdaderamente interesan. Desde hace años se conocen estudios que avalan esta teoría [Yarbus (1967), Hoffman & Subramaniam (1995), Rayner (1998)].

El experimento realizado en esta tesis está principalmente enfocado a los patrones de atención visual manifiesta que son los corrimientos de atención más comunes. Fue diseñado con el objetivo de maximizar la concentración del participante mientras realiza la tarea propuesta, llevando al mínimo posible las distracciones.

4.2. Las hipótesis y los antecedentes del experimento

Existe una importante cantidad de investigaciones acerca del rol del movimiento de los ojos en la ejecución de tareas cotidianas donde se utiliza la visión. En muchos de estos casos se han utilizado mediciones experimentales obtenidas con equipos especiales para el seguimiento de ojos, como se ve por ejemplo en [Land et al. (1999), Hayhoe et al. (2003)]. Los principales resultados de estos trabajos remarcan la importancia tanto de la tarea que se está realizando como de los objetivos y las recompensas internas que tiene la persona, en la selección del lugar donde fijar y el momento en que se hace [Land (1997), Hayhoe & Ballard (2005)].

En relación al experimento en sí que aquí se presenta, ya se ha adelantado que no existen antecedentes directos. Sin embargo, existen algunos grupos que han realizado estudios relacionados en los que se analizan los patrones de fijación mientras las personas manejan un vehículo. Uno de los primeros estudios se reportó en [Mourant & Rockwell (1970)] donde se demostró experimentalmente que los patrones de búsqueda y exploración reflejan diferencias según la familiaridad que se tenga del camino y según las condiciones de manejo. Se encontraron además algunas de las funciones que tiene la visión periférica mientras se maneja, como son el monitoreo constante de otros vehículos y de la línea de marcación del carril. Luego en [Mourant & Rockwell (1972)] se compararon los patrones de búsqueda y exploración de conductores experimentados con los de conductores novatos, encontrando que el proceso de adquisición visual de los novatos es muy deficiente para detectar circunstancias que tienen un alto potencial de accidente. Los autores recomendaron fuertemente que se prohíba manejar a los conductores novatos en rutas públicas hasta que demuestren experiencia suficiente, sugiriendo además el desarrollo de programas de entrenamiento para mejorar el rendimiento de estos conductores. En [Shinar et al. (1977)] se estudiaron los patrones de fijación mientras se conduce un vehículo por una curva, encontrando que dichos patrones persiguen la geometría del camino y que existen muchas más fijaciones en los costados del camino, a diferencia de los caminos rectos con los que se utiliza para esto más frecuentemente la visión periférica. Además se muestra que el proceso de percepción de la curva comienza un par de segundos antes de efectivamente llegar a la misma, y que los patrones de búsqueda de curvas a la izquierda y curvas a la derecha son asimétricos. En [Land & Lee (1994)] también se ha trabajado en la comprensión de como los conductores se comportan durante las

curvas y cuál es la relación entre los movimientos oculares y el control de dirección del volante. Allí se ha encontrado que los conductores se basan particularmente en el "punto tangente" del interior de cada curva, buscando este punto 1 o 2 segundos antes de cada curva. Por otro lado, los resultados de [Underwood et al. (1999)] sugirieron que los conductores experimentados fijan su atención en el punto tangente menos frecuentemente que aquellos principiantes, y que en las curvas cerradas, donde existe una limitada visión que permita anticipar potenciales peligros, la cantidad de fijaciones en el punto tangente se reduce en comparación con las curvas más abiertas. Más recientemente, el trabajo de [Lehtonen et al. (2012)] ha observado que los conductores se anticipan a las curvas abiertas alternando su atención visual entre el camino y un punto al final de la curva llamado punto de oclusión. Se afirma que otras tareas que son demandantes de memoria de trabajo reducen la anticipación visual y que dicho comportamiento debería entenderse en términos de una competencia por los recursos de ejecución de dicha memoria.

Respecto al experimento que se presenta en este capítulo, se espera que los patrones de fijación resultantes del experimento muestren ciertas características que son propias de la tarea visual que se propone al observador. Uno de los objetivos es encontrar evidencias acerca de la existencia o no de lugares específicos de la imagen que sean estratégicos para determinar la topología del camino. Se pretende además investigar el orden y el momento en que el observador enfoca su atención en cada uno de ellos. Los resultados obtenidos deberían dar soporte a una estrategia de inspección del tipo *top-down* [Hoffman (1998), Gilbert & Sigman (2007)]. Estos son procesos que guían constantemente la atención y en los que los objetivos de la persona, como su memoria y la tarea que está realizando influyen involuntariamente sobre la selección de los estímulos visuales que se van a atender [Bar (2003), Li et al. (2004), Sigman et al. (2005)]. Estos comportamientos y la interacción entre ellos es muy difícil de modelar y ha sido mucho menos estudiado en la literatura.

Cabe recordar que los conceptos de los procesos del tipo *top-down* y también de aquellos del tipo *bottom-up* se definieron por primera vez en la Sección 3.1.1. Además en la Sección 3.1.1 se describió como éstos procesos se integran y conviven en los diferentes elementos del cerebro que conforman el sistema visual humano, permitiendo un flujo bidireccional de la información. El análisis del aspecto *bottom-up* de la tarea que aquí se propone será a través de la observación y el estudio del efecto de las saliencias de la

imagen en el comportamiento de los ojos durante la misma. La saliencia de un pixel, o un grupo de pixeles, es un atributo por el cual sobresalen o contrastan con sus vecinos. Los modelos computacionales de la atención visual que son guiados por las saliencias como estímulo se conocen justamente como modelos del tipo *bottom-up* [Itti (2000)].

Por otro lado, es de esperarse que los resultados muestren algunas tendencias en común con otras investigaciones donde también se analiza como los humanos inspeccionan visualmente una fotografía [Rayner (1998)]. En [Tatler et al. (2005)] por ejemplo, los autores encontraron que la consistencia en la posición de las fijaciones seleccionadas por los participantes decrece después de un número pequeño de fijaciones posteriormente a la aparición del estímulo. Esto significa que los observadores se comportan de una manera mucho más uniforme en las primeras fijaciones en comparación con las posteriores, sugiriendo que las tendencias que se marquen en las primeras fijaciones tendrán mayor peso. Este comportamiento se analizará más adelante mediante diferentes análisis dinámicos.

El experimento que se diseñó y que se llevó a cabo con un grupo de voluntarios consiste básicamente en que cada persona observe un conjunto de fotografías y determine para cada una de las escenas la correspondiente topología del camino que observó en la escena. Se registraron los movimientos oculares de los participantes mientras realizaban la tarea a través de un equipo especial para el seguimiento de ojos. A continuación se describen en detalle los procedimientos que se siguieron y se analizan en profundidad todos los resultados obtenidos.

4.3. Metodología experimental utilizada

Esta sección presenta las bases del experimento utilizado. Aquí se describen todos los detalles acerca de los participantes, el material visual y el equipamiento utilizado, además del procedimiento completo que se siguió para la experiencia.

4.3.1. Participantes

Participaron de la experiencia 34 voluntarios (3 mujeres, 31 hombres) del Instituto de Investigaciones en Ingeniería Eléctrica (IIIE), incluyendo estudiantes de posgrado y profesores (Fig. 4.1). Las personas tenían entre 25 y 51 años de edad y todas ellas poseen

Figura 4.1: Se pueden observar aquí algunos de los voluntarios que participaron del experimento. Cada uno de ellos realizó la actividad sentado frente a un monitor de PC y las cámaras del equipo Eyelink 1000, con el mentón y la frente apoyados sobre un soporte fijo.

al menos un título de licenciatura o ingeniería. Una pequeña minoría de ellos todavía no había aprendido a conducir un vehículo, o bien habían obtenido recientemente la licencia para conducir. Todos los participantes mostraron tener una visión normal.

4.3.2. Material visual utilizado

Un total de 33 fotografías fueron seleccionadas para mostrarse a los participantes durante la actividad experimental. Todas estas fotografías son de autoría propia y fueron capturadas en una resolución de 12 megapixeles. Para los fines de la experiencia la resolución de las imágenes luego se redujo a 1024x768 pixeles mediante el software de edición *Gimp*. Cabe mencionar que esta es la resolución óptima elegida para la pantalla donde se reproducen las imágenes. Los escenarios retratados reflejan diferentes tipos de topologías con diferentes niveles de complejidad y en ausencia total de vehículos. Se eligieron fotografiar algunos caminos secundarios o alternativos y caminos sin asfaltar

ubicados en barrios alejados del centro urbano de la ciudad de Bahía Blanca. El objetivo era reducir al mínimo la existencia de elementos distractores, por lo que se seleccionaron lugares que tengan muy poca estructura, o bien no tengan ninguna. Todas las fotos se tomaron desde una altura similar (aproximadamente 1.5m) y con la cámara ligeramente inclinada hacia abajo. Esta configuración sería similar a si se ubicase la cámara sobre el techo de un vehículo para testear futuros algoritmos de percepción autónoma. Por último, la apertura angular del lente se configuró en cada caso lo suficientemente ancha como para permitir al observador apreciar claramente la topología del camino en la imagen.

Como se mencionó anteriormente, para cada una de las fotografías inspeccionadas el participante debe reconocer la correspondiente topología del camino presentado. Para esto, después de cada imagen se presenta al observador una pantalla con 4 topologías diferentes para que pueda seleccionar aquella que considera la más similar a lo que acaba de observar en la fotografía. La topología se define aquí como una representación simplificada de la forma del camino a través de una serie de segmentos rectos o curvos conectados entre sí. Cada una de las pantallas de respuesta también se diseñó con una resolución de 1024x768 pixeles. En la Fig. 4.2 se ilustran algunos ejemplos del material visual mencionado.

Si bien este análisis no está particularmente interesado en el estudio de los patrones de atención mientras se elige y selecciona la correspondiente topología, se cree fuertemente que esta tarea de reconocimiento fuerza a la persona a prestar mucho más atención durante la etapa de inspección de la imagen. Así, el participante se enfocará preferentemente en aquellos lugares que considera relevantes para la tarea que intenta resolver. Además, se les indicó a los voluntarios que para cada una de las imágenes solo existía una respuesta correcta, y que el objetivo del experimento era obtener la mayor cantidad de respuestas correctas posibles. Estas instrucciones apuntaron a incentivar aún más al observador a realizar la tarea de forma consciente y con la mayor concentración posible. Las 4 opciones de respuestas se diseñaron de la siguiente manera: además de la topología correcta, se incluyó una muy diferente, una algo más parecida y otra muy similar a la correcta. El objetivo que se buscó era no facilitar la tarea de reconocimiento. La respuesta correcta se define a partir de la topología real del camino que se observó en el campo en el momento en que se tomó la fotografía.

Figura 4.2: Ejemplos de las imágenes utilizadas durante la experiencia. Las fotografías del camino se muestran en las columnas impares, mientras que las correspondientes pantallas con las opciones disponibles para seleccionar una topología se muentran en las columnas pares.

Durante la preparación del experimento las fotografías fueron ordenadas de forma aleatoria, manteniendo luego dicho orden para todos los participantes. De forma similar, la opción de respuesta correcta para cada caso se ubicó aleatoriamente en una de las cuatro posiciones posibles. Este orden también se mantuvo para todas las experiencias. Del total de 33 fotografías se utilizaron 3 para la etapa de entrenamiento o práctica, mientras que las otras 30 se incluyeron en el experimento en sí.

La totalidad del material utilizado junto con todos los datos adquiridos en el experimento y algunos resultados de ejemplo se pueden encontrar en un sitio web especial preparado para esto. Estos datos están disponibles en [Moreyra & Masson (2012)] y pueden ser de utilidad para otros investigadores interesados en la temática. Para comodidad del lector, en el Apéndice A se exhibe cada una de las fotografías junto a su correspondiente pantalla de respuesta y un histograma de la posición de las fijaciones en dicha imagen.

Tabla 4.1: Parámetros de configuración del Eyelink 1000

Parametro	**Valor**	**Unidad**
Frecuencia de muestreo	1000	Hz
Umbral de desplazamiento	0.1	$^\circ$
Umbral de velocidad	30	$^\circ/s$
Umbral de aceleración	8000	$^\circ/s^2$
Cota del error promedio de posición	0.5	$^\circ$

4.3.3. Configuración del experimento y el equipamiento

El equipo utilizado para registrar los movimientos oculares durante el experimento es un Eyelink 1000 de la empresa SR Research. En la Tabla 4.1 se detallan la frecuencia de muestreo y los umbrales de detección y de error del equipo. Todas las calibraciones del equipo durante las actividades se realizaron con el método de 9 puntos, que consiste básicamente en que la persona fije su mirada secuencialmente en 9 puntos diferentes de la pantalla mientras el equipo ajusta internamente sus parámetros.

Todos los participantes realizaron la actividad sentados con su mentón y frente apoyados sobre un soporte fijo ubicado a 59cm de la pantalla (ver Fig. 4.1). Las cámaras especiales del Eyelink se ubican justo debajo del monitor de la PC. Todas las imágenes se mostraron en la pantalla de dicho monitor con una resolución de 32 pixeles/cm, lo que resultó en imágenes de 32cm de ancho por 24cm de alto.

4.3.4. Procedimiento

Algunos días previos al experimento se envió a cada uno de los voluntarios un documento detallado con todas las instrucciones para la realización del mismo. A pesar de que un 25 % de los voluntarios admitió no haber leído las instrucciones, este entrenamiento previo fue de gran utilidad para acelerar las actividades y reducir considerablemente el tiempo total de duración de dichas experiencias. De todas formas, con cada uno de los participantes se realizó un breve repaso del procedimiento antes de comenzar la actividad, independientemente de las instrucciones que aparecían en pantalla ya durante el experimento.

El procedimiento completo incluía cuatro etapas: una calibración inicial, la práctica o entrenamiento, una segunda calibración y finalmente el correspondiente test. Después

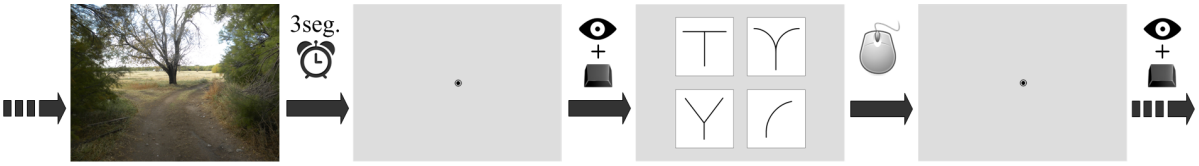


Figura 4.3: Diagrama esquemático de la secuencia de imágenes presentadas al observador. Este ciclo se repite de igual manera para todas las imágenes del experimento.

de la primera calibración, el participante pudo practicar con algunas fotografías con el objeto de familiarizarse con el equipo, los comandos de control necesarios y la tarea asignada. Esta sesión de práctica consistía en una versión muy reducida del verdadero test donde se utilizaron 3 fotografías predeterminadas del conjunto total de 33. Una vez terminado este entrenamiento se permitió al participante relajarse unos segundos y acomodarse si fuese necesario. Finalmente, la segunda calibración se realizó antes de comenzar el experimento. Durante el mismo, se mostró una por una el conjunto restante de 30 fotografías junto a sus correspondientes pantallas de respuestas. Cada una de ellas se mostraba en pantalla solo por 3 segundos de tal manera de presionar a la persona a enfocarse solo en lo que fuera relevante para la tarea asignada. A continuación, la pantalla con las opciones disponibles se mantenía todo el tiempo que fuera necesario hasta que la persona pudiera tomar una decisión. La selección de la respuesta se realizaba mediante el puntero del ratón de la PC haciendo clic con el botón izquierdo. Después de cada imagen (fotografía o respuestas) el equipo exigía al participante una recalibración rápida que consistía en mirar un punto en el centro de la pantalla y pulsar una tecla determinada. Este mecanismo permitía al equipo eliminar posibles errores debidos a los pequeños movimientos de la cabeza de la persona durante la inspección de las imágenes. En la Fig. 4.3 se presenta un diagrama representativo que ilustra claramente la secuencia que se acaba de describir.

4.4. Preprocesamiento de las imágenes para el posterior análisis de las fijaciones

Para analizar las fijaciones en el contexto apropiado resulta necesario identificar aquellas regiones de la escena que tienen algún tipo de significado o implicancia para la

tarea visual que se realiza. En primer lugar, para cada imagen se identifican los pixeles que pertenecen al camino. Luego, en base a éstos es posible definir aquellos pixeles que se encuentran en la región del borde o los límites del camino, que también son de mucho interés para el análisis que sigue.

Más adelante se presenta un método automático que permite separar la región del camino en subregiones específicas. Este algoritmo es esencial para el posterior análisis que se hará con los resultados y representa un enfoque innovador para el estudio de las fijaciones en este experimento. La comparación directa de la posición de las fijaciones no es correcta cuando éstas pertenecen a imágenes con escenas diferentes. Este método que se propone hace posible la comparación sin sesgo de los resultados obtenidos con diferentes participantes y con caminos de diferentes formas y dimensiones y es aplicable para todas las imágenes donde el camino pueda modelarse como una bifurcación de dos vías. Algunas topologías representadas serían: $\top$, $\dashv$, $\vdash$, λ, Υ, entre otras.

4.4.1. Extracción del camino y sus bordes

Cada una de las imágenes se etiquetó manualmente para obtener un mapa binario de la región que representa al camino. Un ejemplo de este mapa se puede observar en la Fig. 4.4(a) donde se superpone en color azul con rayas blancas sobre la correspondiente imagen. Una vez disponible el mapa del camino es posible definir si una fijación se hizo o no sobre él.

Como se adelantó, también interesa determinar si una persona ha obtenido información acerca de los bordes del camino, es decir, de los límites entre el área transitable y aquella que no lo es. Para esto hace falta definir una región alrededor del borde, que se puede hallar mediante el perímetro de la región binaria que define al camino. El ancho de la región estará relacionado con la resolución que tiene el ojo cuando fija un punto en la imagen y se define a continuación.

La agudeza visual y el ancho del borde

Se ha visto en la Sección 2.1 que la agudeza visual es mucho mayor en el centro de la fijación debido a una gran densidad de células cono en la fóvea (repasar Fig. 2.2). Los datos experimentales de [Westheimer (1987)] permiten afirmar que a una distancia

(a) (b)

Figura 4.4: La región del camino junto a la región de los bordes para una de las imágenes del test. El mapa del camino (a) y el mapa de los bordes (b) se solapan en color azul con rayas blancas sobre la correspondiente fotografía. El ancho de la región del borde se define a partir del umbral e_{th}.

de 1° del eje foveal ya existe una pérdida de agudeza visual de aproximadamente un 40 %. Este valor será utilizado para definir el umbral de percepción e_{th}, que indicará la excentricidad a partir de la cual cualquier información visual se considerará como no vista.

Teniendo en cuenta los detalles de configuración de la Sección 4.3.3, este umbral en la excentricidad se puede redefinir como una distancia medida en pixeles a partir de la Ec. 4.1. Para $L = 59cm$, $e_{th} = 1°$ y una resolución de 32 $pixeles/cm$, el umbral de percepción es de unos 33 pixeles. Por lo tanto, cualquier pixel que pertenezca al borde será considerado visto si para una determinada fijación el círculo de radio 33 pixeles centrado en ella contiene a dicho pixel. Un ejemplo de la región del borde del camino se ilustra con color azul a rayas blancas en la Fig. 4.4(b).

$$e_{th_{pixeles}} = \text{Resolución} \cdot L \cdot \tan(e_{th}) \tag{4.1}$$

4.4.2. Separación del camino en subregiones

Además de definir si una fijación se realizó en el camino, se tiene particular interés en determinar que ubicación específica se visitó y si ésta tiene algún tipo de relación con la topología del camino. Por esta razón se diseñó e implementó un método para dividir de forma automática el área del camino a partir de ciertos puntos que se llamarán

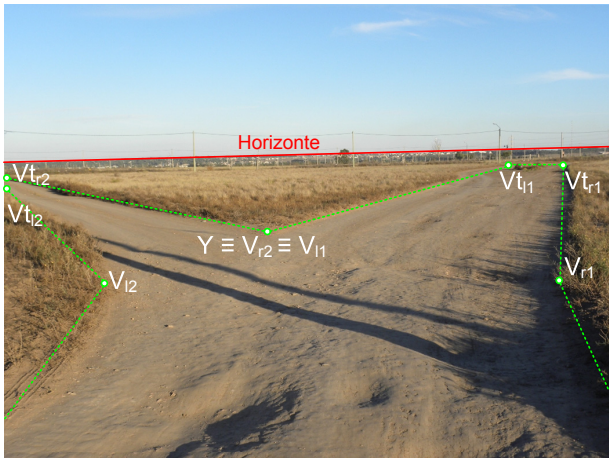


Figura 4.5: Los puntos llamados vértices y la línea del horizonte se utilizan para particionar el camino. Estos se marcan manualmente sobre la imagen y se proyectan hacia las coordenadas del mundo.

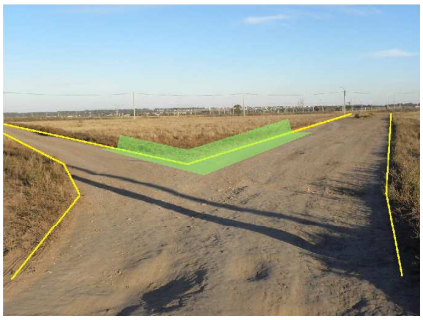
vértices (ver Fig. 4.5). $\mathbf{Vt}_r$ y $\mathbf{Vt}_l$ definen los vértices de arriba a la derecha y arriba a la izquierda, mientras que $\mathbf{V}_r$ y $\mathbf{V}_l$ definen los vértices de abajo a la derecha y abajo a la izquierda, respectivamente. Se supone aquí un modelo de bifurcación de dos vías que contiene la siguientes subregiones: los finales de camino, los comienzos de camino, la intersección de caminos, la región media del camino, y la región inferior del mismo. Esta subdivisión posibilitará el análisis respecto del rol que tiene cada una de ellas sin importar la diferencia de tamaño o posición que presente para los distintos casos. Las Figs. 4.6(a) a 4.6(g)) muestran un ejemplo de las regiones resultantes al aplicar el método propuesto (regiones de color verde superpuestas sobre la imagen). Se puede observar también en la imagen que el correspondiente borde del camino se agregó a cada una de dichas regiones.

El método se inicializa seleccionando manualmente los vértices de cada una de las imágenes. Luego éstos se proyectan hacia coordenadas del mundo bajo la suposición de que pertenecen al plano del suelo ($y_w = 0$). Esto asegura que aunque las fotografías se tomen bajo diferentes parámetros internos de la cámara, éstas puedan ser tratadas de forma equivalente, favoreciendo la comparación entre ellas.

Las Eq. 4.2 y Eq. 4.3 definen la proyección de un punto $\mathbf{P}_c = (x_c, y_c)$ de la imagen hacia coordenadas del mundo. Nótese que primero el punto se pasa del sistema de coordenadas centradas en la parte superior izquierda de la imagen al sistema de coordenadas

(a) Fin de camino 1

(b) Fin de camino 2



(c) Región de Intersección

(d) Comienzo de camino 1

(e) Comienzo de camino 2

(f) Región media del camino

(g) Región inferior del camino

Figura 4.6: Regiones resultantes de aplicar el algoritmo de visión de la región del camino. Cada una de ellas, junto al correspondiente borde, se superpone en color verde sobre la imagen en (a), (b), (c), (d), (e), (f) y (g). Los parámetros del algoritmo que se utilizaron aquí son $K_L = 0.85$, $K_I = 0.6$, y $K_B = 0.45$.

con su origen en el centro de la imagen $(\frac{C}{2}, \frac{F}{2})$, obteniéndose $\mathbf{P}_c^o = (x_c^o, y_c^o)$. Aquí C y F indican la cantidad de columnas y filas de la imagen, respectivamente.

$$x_w = \frac{h_c \cdot x_c^o}{f_0^p \cdot \sin\theta_t - y_c^o \cdot \cos\theta_t}, \tag{4.2}$$

$$z_w = \frac{h_c \cdot f_0^p \cdot \cos\theta_t + h_c \cdot y_c^o \cdot \sin\theta_t}{f_0^p \cdot \sin\theta_t - y_c^o \cdot \cos\theta_t}, \tag{4.3}$$

Una vez proyectado debe trasladarse y rotarse respecto de la posición y orientación de la cámara, resultando en $\mathbf{P}_w = (x_w, 0, z_w)$. Como se mencionó en la Sección 4.3.2, las fotografías se tomaron con la cámara ligeramente inclinada hacia abajo y desde una altura similar al techo de un vehículo (aproximadamente $h_c = 1.5m$). Este ángulo de inclinación, que se llamará ángulo de *tilt*, se puede estimar a partir de la posición del horizonte en la imagen. El horizonte también se define manualmente para cada figura, siendo H_y la fila que corresponde al mismo. El ángulo de *tilt* θ_t se calcula entonces con la Ec. 4.4. Ninguna otra rotación es considerada. La distancia focal en pixeles f_0^p se calcula para cada una de las fotografías haciendo $f_0^p = C * f_0/W_{CCD}$, donde f_0 es la distancia focal en mm y W_{CCD} es el ancho del sensor CCD.

$$\theta_t = \arctan\left(\frac{H_y - \frac{F}{2}}{f_0^p}\right). \tag{4.4}$$

El procedimiento para separar el camino se ejecuta casi por completo en el sistema de coordenadas del mundo. Los detalles del algoritmo utilizado se describen en el *Algoritmo* 1 junto con los diagramas de la Fig. 4.7. Brevemente, la región *Fin del camino* se delimita mediante una línea perpendicular (y_{Lp}) a la línea central del camino (y_L), que se ubica a una distancia $K_L \cdot L$ del punto $\mathbf{Vt}_i$, definido por los vértices superiores (ver Fig. 4.7(a)). L representa el largo estimado del camino y K_L es un parámetro de diseño tal que $0 \leq K_L \leq 1$. La *Región de Intersección* se encuentra a partir del punto de intersección de las líneas centrales ($\mathbf{I}$) y algunos puntos y líneas auxiliares (ver Fig. 4.7(b)). La región se define en función del ancho estimado del camino R_W y el parámetro de diseño K_I ($0 \leq K_I \leq 1$). Las regiones denominadas *Comienzo de camino* se definen a partir del área resultante de la simple unión de 4 puntos ($\mathbf{q}_{r_i}$, $\mathbf{q}_{l_i}$, $\mathbf{V}_{r_i}$ y $\mathbf{V}_{l_i}$) sustrayéndose el área ya ocupada por la región de intersección (ver Fig. 4.7(c)). Luego, la *Región inferior del camino* se define como el área frente a la cámara que se encuentra a una distancia menor

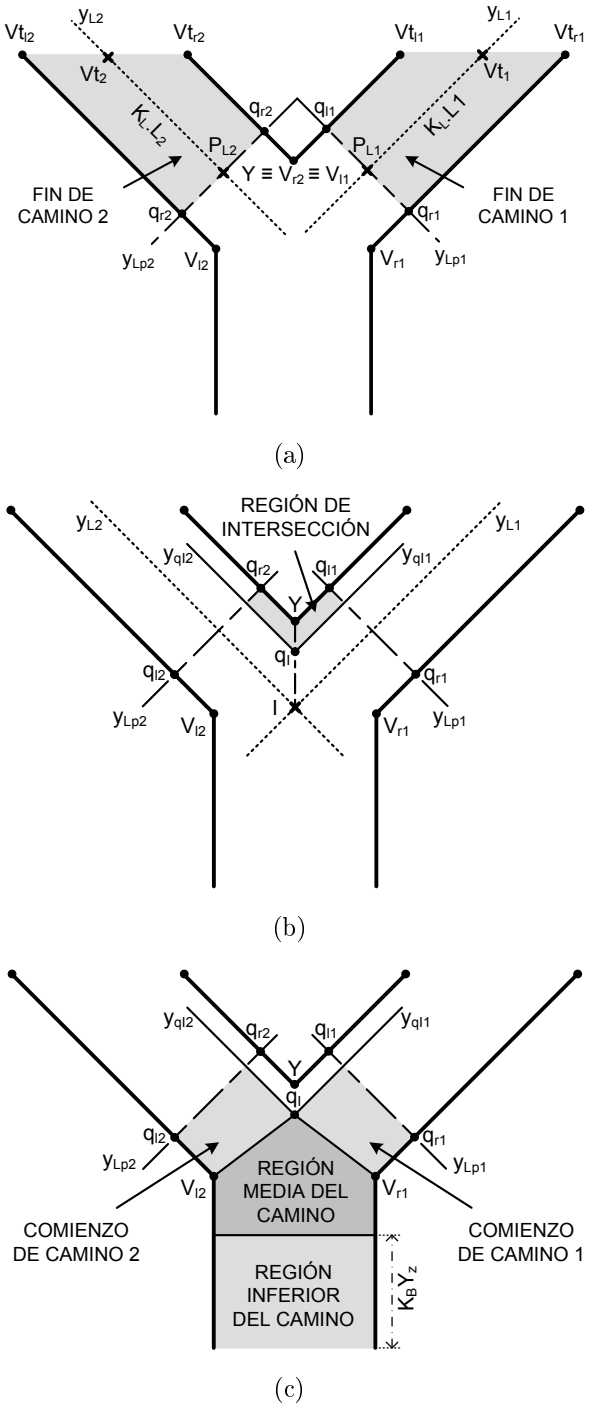


Figura 4.7: Diagramas esquemáticos de los procedimientos realizados para separar (a) los *Fines de camino*, (b) la *Región de intersección*, y (c) los *Comienzos de camino*, la *Región media del camino* y la *Región inferior del camino*. Cada punto y cada línea referenciada en el algoritmo denominado Algorithm 1 se ilustra en estos diagramas para ayudar al lector a una mejor comprensión del método planteado.

Algoritmo 1 Separación del camino en subregiones

Hallar fines de camino:

for $i = 1 \rightarrow 2$ **do**
$\measuredangle y_L \leftarrow (\measuredangle(\mathbf{Vt}_{r_i}, \mathbf{V}_{r_i}) + \measuredangle(\mathbf{Vt}_{l_i}, \mathbf{V}_{l_i}))/2$ ▷ Ángulo de línea calculado con los ángulos segmentos a los vértices.
$\mathbf{Vt}_i \leftarrow (\mathbf{Vt}_{r_i} + \mathbf{Vt}_{l_i})/2$
Hallar línea y_L con $\measuredangle y_L$ que pase a través de $\mathbf{Vt}_i$
$L = \|\mathbf{Vt}_i - (\mathbf{V}_{r_i} + \mathbf{V}_{l_i})/2\|$ ▷ Largo estimado del camino.
Hallar el punto $\mathbf{P}_L$ en y_L tal que $\|\mathbf{Vt}_i - \mathbf{P}_L\| = K_L \cdot L$
Hallar línea y_{Lp} perpendicular a y_L que pase por $\mathbf{P}_L$
Hallas los puntos $\mathbf{q}_r, \mathbf{q}_l$ como la intersección de y_{Lp} con los segmentos $\overline{\mathbf{Vt}_{r_i}\mathbf{V}_{r_i}}, \overline{\mathbf{Vt}_{l_i}\mathbf{V}_{l_i}}$
if $(\|\mathbf{Vt}_{r_i} - \mathbf{q}_r\| > K_L \cdot \|\mathbf{Vt}_{r_i} - \mathbf{V}_{r_i}\|)$ ó $(\|\mathbf{Vt}_{l_i} - \mathbf{q}_l\| > K_L \cdot \|\mathbf{Vt}_{l_i} - \mathbf{V}_{l_i}\|)$ **then**
Actualizar $\mathbf{q}_r$ ó $\mathbf{q}_l$ a las cotas correspondientes $K_L \cdot \|\mathbf{Vt}_{r_i} - \mathbf{V}_{r_i}\|$ ó $K_L \cdot \|\mathbf{Vt}_{l_i} - \mathbf{V}_{l_i}\|$
Actualizar y_{Lp} para que pase por el punto $\mathbf{q}$ más cercano a $\mathbf{Vt}_i$ (igual pendiente)
end if
Hallar la región *Fin de camino* delimitada por y_{Lp} ▷ ver Fig. 4.7(a)
Hallar la intersección del perímetro de la región del camino con $y_L\perp$ para definir el ancho de camino R_W.
end for

Hallar región de intersección:

Encontrar el punto $\mathbf{I}$ como la intersección de las líneas y_{L_1} y y_{L_2}
if $\mathbf{I} \notin$ cuadrilátero $\mathbf{V}_{r_1}\mathbf{V}_{l_1}\mathbf{V}_{r_2}\mathbf{V}_{l_2}$ **then**
$I \leftarrow (\mathbf{V}_{r_1} + \mathbf{V}_{l_1} + \mathbf{V}_{r_2} + \mathbf{V}_{l_2})/4$
end if
Encontrar el punto $\mathbf{q_I}$ en la línea $\overline{\mathbf{YI}}$ tal que $\|\mathbf{Y} - \mathbf{q_I}\| = K_I \cdot (R_{W_1} + R_{W_2})/2$ ▷ El punto Y se define en la Fig. 4.7(b)
Hallar líneas y_{qI_1} y y_{qI_2} que pasan por $\mathbf{q_I}$ con pendientes $(\angle y_{L_1} + \angle(\mathbf{q}_{l_1}, \mathbf{Y}))/2$ y $(\angle y_{L_2} + \angle(\mathbf{q}_{r_2}, \mathbf{Y}))/2$
Hallar *la región de Intersección* como la región del camino que está delimitada por y_{qI_1}, $y_L\perp_1$, y_{qI_2} y $y_L\perp_2$ ▷ Ver Fig. 4.7(b)

Hallar los comienzos de camino:

for $i = 1 \rightarrow 2$ **do**
Hallar la región *Comienzo de camino* como el área definida por $\mathbf{q}_{r_i}$, $\mathbf{q}_{l_i}$, $\mathbf{V}_{r_i}$ y $\mathbf{V}_{l_i}$, sustrayendo el área ocupada por la *Intersección* ▷ Ver Fig. 4.7(c)
end for

Hallar las regiones media e inferior del camino:

Calcular la distancia D_Y entre el origen y el punto de intersección $\mathbf{Y}$
Hallar la *Región inferior del camino* como los pixeles del camino que tengan su coordenada z en el mundo menor a $K_B \cdot D_Y$ ▷ Ver Fig. 4.7(c)
La *Región media del camino* se halla como el resto de la región del camino que no está ocupada por el resto de las subregiones. ▷ Ver Fig. 4.7(c)

Agregar las regiones de bordes correspondientes a cada subregión.

que $K_B \cdot D_Y$, donde D_Y es la distancia entre la cámara y el punto de intersección $\mathbf{Y}$ y K_B es tal que $0 \leq K_B \leq 1$. Una vez determinadas estas regiones en el mundo se proyectan inversamente hacia la imagen para generar las máscaras binarias que permitirán extraer la porción de la región del camino que les corresponde. La *Región media del camino* se obtiene como el área restante del camino que no es ocupada por el resto de las subregiones. Finalmente, la región del borde del camino que corresponde se agrega a cada subregión mediante los lineamientos descritos en la Sección 4.4.1.

4.5. Análisis de los resultados

Analizar el comportamiento de los ojos de una persona mientras realiza una tarea visual compleja presenta un gran desafío. En particular, los experimentos que aquí se llevaron a cabo muestran que las personas pueden resolver con éxito la tarea propuesta utilizando diferentes estrategias de inspección. Esto es, se pueden obtener resultados similares a pesar de administrar en forma desigual la distribución y el tiempo de las fijaciones. Ésto dificulta aún más el estudio y comparación de los resultados obtenidos con diferentes participantes y con diferentes escenas. Sin embargo, este trabajo aporta un enfoque innovador para el análisis de los resultados que facilitará esta comparación.

Esta sección presenta además un estudio acerca de las regiones de la imagen que son relevantes para la tarea y acerca de la influencia de las saliencias de la imagen en los patrones de fijación. También incluye un análisis dinámico de las fijaciones a través del índice de fijación, es decir, del momento en que la fijación se realizó dentro de la secuencia de fijaciones utilizadas en una imagen. Aunque no todas las personas utilizan la misma cantidad de fijaciones, el comportamiento promedio de todos muestra una dinámica aproximada de la estrategia de inspección. Hacia el final de la sección se enfatizan los principales resultados y sus implicancias.

4.5.1. Análisis general de las fijaciones: las estrategias de inspección

Dado que el tiempo que tienen los participantes para inspeccionar la imagen es muy limitado éstos deben administrar muy bien la cantidad de fijaciones y sus ubicaciones

de forma de poder adquirir la información mínima necesaria para resolver la tarea. Los participantes no conocen de antemano el tipo de camino que aparecerá en imagen ni el tipo de topología asociado por lo que deben desarrollar una estrategia de inspección a medida que observan la fotografía. El comportamiento resultante dependerá entonces de la información que se adquiera en las primeras fijaciones y de la experiencia que se ha ganado con las imágenes anteriores del mismo experimento.

En la Fig. 4.8 se incluye un histograma del número de fijaciones utilizados por imagen, independientemente del participante que las realizó o de la imagen donde se realizaron. Se puede observar allí una gran variabilidad que podría ser explicada por la variedad en las escenas y topologías observadas, o bien por los diferentes comportamientos de los voluntarios. Para evidenciar esto último, en la Fig. 4.9 se muestran los histogramas del número de fijaciones para seis participantes diferentes. Los tres histogramas que se encuentran en la columna izquierda pertenecen al grupo del 15 % de participantes que mejor rendimiento tuvieron en el experimento. Esto es, son tres de los voluntarios que mayor cantidad de respuestas correctas obtuvieron en la experiencia. En la columna derecha en cambio, se observan los histogramas para tres participantes que pertencen al 15 % que peor rendimiento mostraron. Se puede observar claramente que la cantidad de fijaciones no está directamente relacionada con el éxito en la tarea, ya que hay participantes que utilizan menos fijaciones que otros mostrando resultados similares.

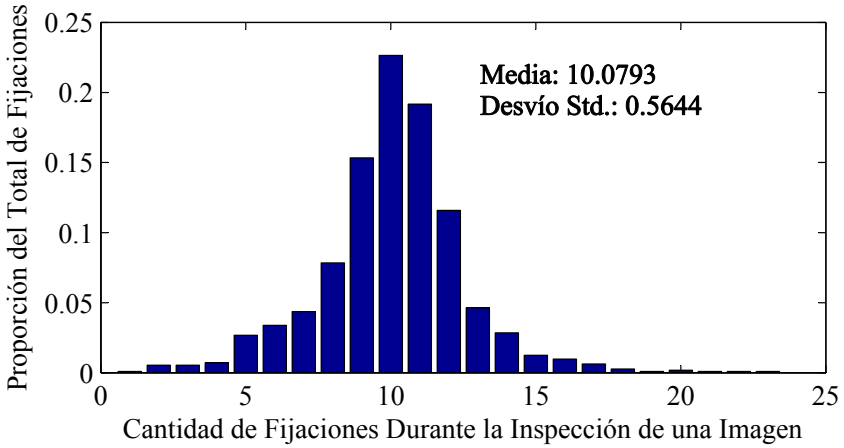


Figura 4.8: Histograma del número de fijaciones utilizadas para la inspección de las imágenes durante el experimento. También se incluyen el valor medio y el desvío estándar como referencia.

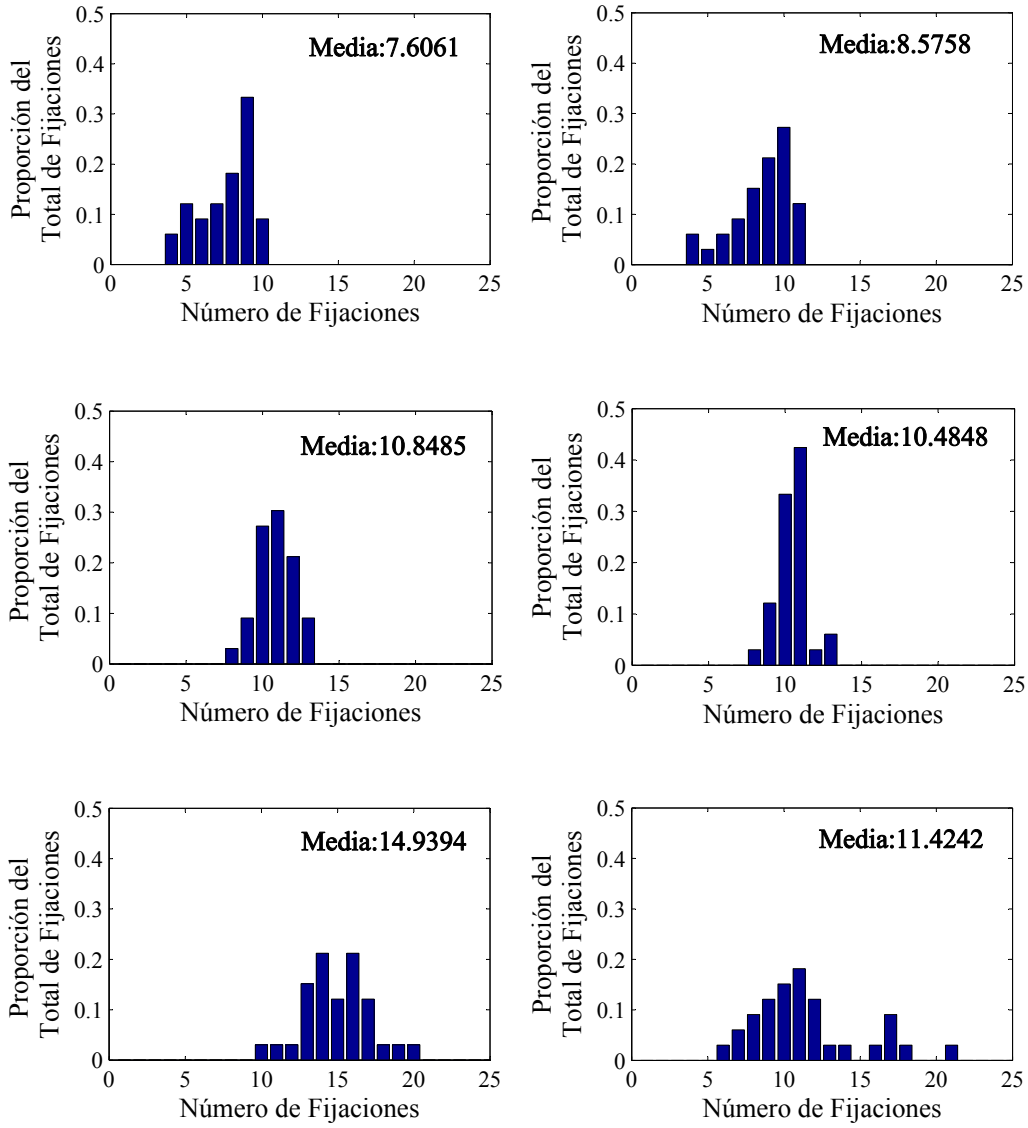


Figura 4.9: Histogramas de ejemplo junto a sus valores medios de fijaciones utilizadas por 6 participantes. Los histogramas de la columna de la izquierda son de 3 de los participantes que pertenecen al 15 % que mejor rendimiento tuvo en las experiencias, mientras que los histogramas de la columna derecha son de 3 voluntarios que están dentro del 15 % que peor lo hizo.

Si por otro lado se analiza cada una de las imágenes por separado, se podrá observar que la cantidad de fijaciones varía con cada una de las situaciones presentadas. La Fig. 4.10 muestra la cantidad promedio de fijaciones que los participantes realizaron en cada una de las fotografías (curva a trazos de color rojo) y el número mínimo de fijaciones utilizadas (curva punteada de color verde). La región sombreada expande el valor medio en un desvío estándar de cada lado para ilustrar la variabilidad de los datos. Las tres fotografías que se utilizaron para la fase de práctica (etiquetadas como A, B and C) se incluyeron para que sea posible comparar sus resultados con los del experimento en sí. Se puede ver que al finalizar el entrenamiento y comenzar con el test (imagen 1) los participantes prestan mayor atención a la inspección de la imagen, reflejándose esto en un inmediato aumento del promedio de fijaciones utilizados. La cantidad mínima de fijaciones también muestra un comportamiento muy similar. El desvío estándar presenta una tendencia a disminuir a lo largo del experimento, lo que sugiere que los participantes se vuelven más precisos en la inspección aprovechando el aprendizaje logrado con las imágenes previas. En general se observa que para las fotos que contienen escenas más complejas la cantidad promedio de fijaciones que se usa es mayor. Sin embargo, esto también se ve afectado por la distancia entre las regiones de interés, ya que cuanto más lejanas se encuentren en la imagen, mayor es la cantidad de fijaciones intermedias que se observan en los patrones resultantes.

La estrategia que se utiliza para inspeccionar la imagen puede estudiarse también desde el punto de vista de la amplitud de los sacádicos y la dispersión de la posición de las fijaciones. La amplitud de los sacádicos se define aquí como la distancia euclídea entre el origen del sacádico y su posición final, y se mide en en pixeles. Expresado matemáticamente sería:

$$A_{sac} = \sqrt{(x_{final} - x_{inicial})^2 + (y_{final} - y_{inicial})^2}. \tag{4.5}$$

Debido a que la agudeza visual disminuye dramáticamente con la excentricidad, no sería razonable esperar sacádicos de gran amplitud durante las primeras fijaciones, ya que la persona aún no ha obtenido suficientes detalles de la imagen. Si se ordenan todos los sacádicos según su índice, sin importar la persona o la imagen a las que pertenecen, se puede calcular una amplitud promedio de los sacádicos para cada índice (ver Fig. 4.11).

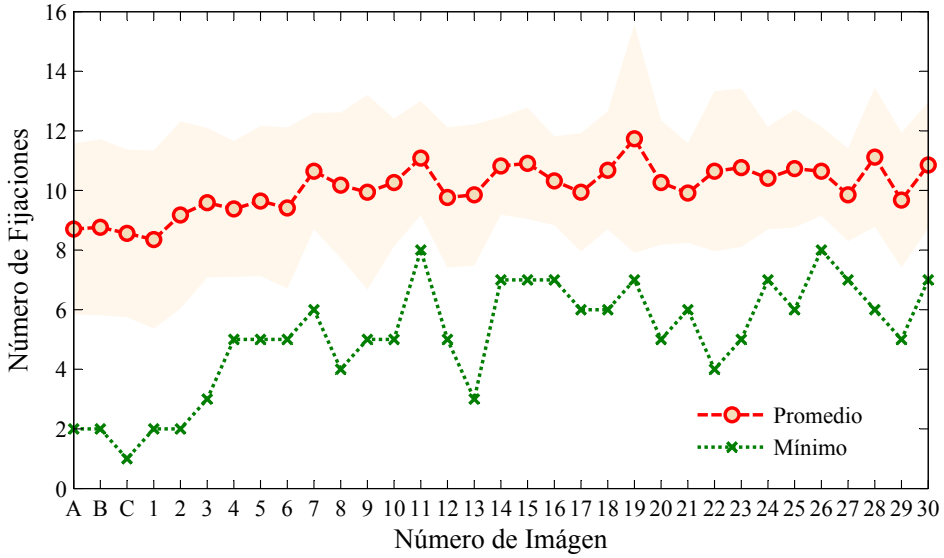


Figura 4.10: Número promedio de fijaciones utilizadas por los participantes en cada una de las fotografías inspeccionadas durante la actividad (curva a rayas de color rojo). La región sombreada indica un apartamiento de un desvío estándar a cada lado del valor promedio. La curva punteada de color verde indica el número mínimo de fijaciones utilizadas.

Se observa que los primeros sacádicos son en promedio más cortos, y que en 5 o 6 movimientos alcanzan su máxima amplitud. Esta última medida luego disminuye para finalmente volver a aumentar en las fijaciones finales. Esta disminución está acompañada por saltos con algo más de precisión que se infieren a partir del comportamiento del desvío estándar (ver región sombreada). Estos resultados sugieren que las personas se dedicarían a inspeccionar determinados lugares que ya observaron en los pasos previos. Aquí sería donde los procesos que involucran a la memoria entran en juego permitiendo una selección más precisa del próximo destino de un sacádico, aunque el estudio de esta relación no forma parte del objetivo de esta tesis.

La dispersión en la posición de las fijaciones es una medida de como se reparten espacialmente en la imagen. Una baja dispersión, por ejemplo, indicaría que el interés de las personas se focaliza solo en ciertas áreas de la imagen. Se define entonces la dispersión como el promedio de las distancias de cada fijación al centro de la imagen:

$$D_{fij} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\left(x_i - \frac{C}{2}\right)^2 + \left(y_i - \frac{F}{2}\right)^2}, \tag{4.6}$$

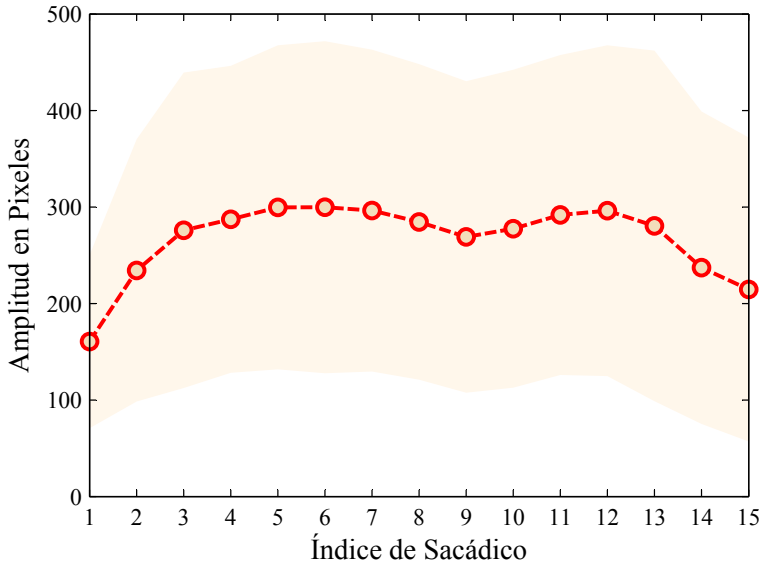


Figura 4.11: Amplitud media de los sacádicos en función del índice del sacádico. La región sombreada representa una extensión de un desvío estándar a cada lado del valor medio.

donde n es el número de fijaciones utilizadas para calcular D_{fij}. La Fig. 4.12 muestra la dispersión de todas las fijaciones del experimento ordenadas por su índice de fijación. Obviamente, la primer fijación tiene muy poca dispersión alrededor del centro de la imagen debido al proceso de recalibración que se da entre pantalla y pantalla. Hay que recordar que este paso implica que la persona mire un punto en el centro de la pantalla y presione una tecla para continuar. En las siguientes fijaciones la dispersión crece abruptamente alcanzando un pico máximo unas 10 veces mayor que el valor inicial. Este crecimiento indica claramente que los participantes comienzan a inspeccionar la imagen en diferentes direcciones basándose en la información detallada que adquieren en cada fijación, además de la información de baja resolución que obtienen con la visión parafoveal. Como ocurrió con el caso de la amplitud de sacádicos, existe un valle en la curva dinámica de dispersión que indica que existe interés en determinados lugares de la imagen que fueron observados anteriormente.

Los patrones de inspección observados sugieren la existencia de determinadas regiones de la imagen que atraen la atención más que otras para resolver la tarea. Los participantes se enfocan en dichos lugares utilizando múltiples fijaciones. Una de estas regiones recae en las zonas del borde del camino y se discutirá a continuación.

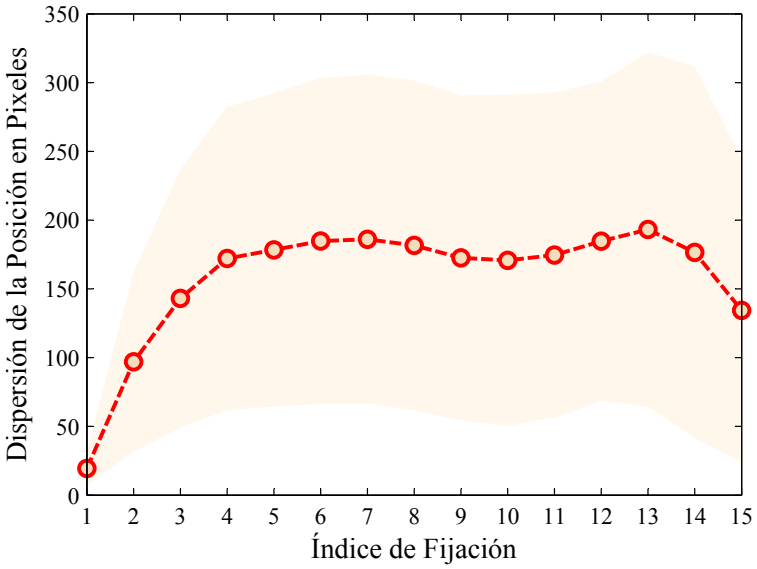


Figura 4.12: Dispersión promedio en la posición de las fijaciones ordenada en función del índice de fijación. La región sombreada representa una extensión de un desvío estándar a cada lado de la curva del valor medio.

4.5.2. La importancia de los bordes del camino

Para el siguiente análisis cada imagen se dividió automáticamente en 4 regiones: la región del *borde del camino*, la región *dentro del camino*, la región *fuera del camino*, y la región por *sobre el horizonte*. El perímetro del mapa binario obtenido con la región del camino se utilizó para hallar las zonas del borde como se explicó en las secciones previas. La región dentro del camino se define como la zona del mapa del camino que no corresponde al borde. La región fuera del camino es el área de la imagen que no pertenece ni al borde ni al camino, y que se encuentra por debajo de la línea del horizonte. Por último, la región por sobre el horizonte es justamente el resto de la imagen que no pertenece a las otras 3 regiones y que está por encima de esta línea. Esta separación permitirá agrupar y comparar las fijaciones espacialmente, sin importar quien las hizo o en que imagen particular se registraron. De las 30 imágenes del experimento, 4 de ellas (identificadas con los índices 5, 12, 16 y 21) no se utilizaron aquí debido a que en ellas el horizonte no está claramente definido.

Para cada una de las 26 imágenes se contó la cantidad de fijaciones N_i que se realizó en cada unas de las 4 regiones i y luego se las normalizó según el área en pixeles de la

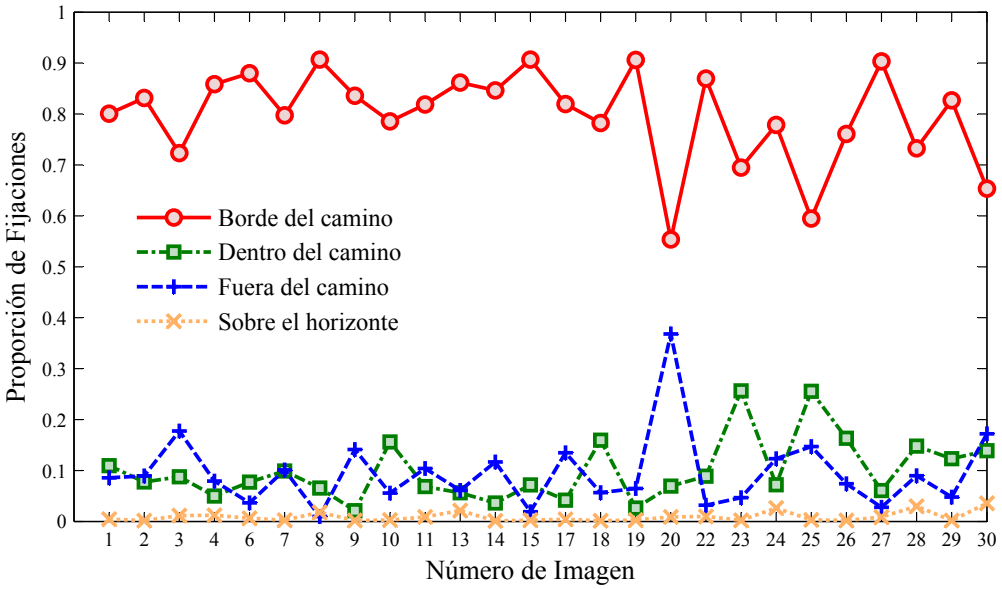


Figura 4.13: Proporción normalizada de fijaciones en el borde del camino (círculos rojos), dentro del camino (cuadrados verdes), fuera del camino (signos + azules) y por sobre el horizonte (signos × naranjas) en cada una de las imágenes del experimento.

región correspondiente A_i, es decir, $\frac{N_i}{A_i}$. Se utilizó la Ec. 4.7 para calcular la proporción de fijaciones en cada región de tal forma que el total para toda la imagen sume 1. Esta proporción se usa aquí como una medida de cuan atractiva resulta una región para los participantes. La normalización se efectúa para poder comparar resultados entre regiones de diferentes tamaños, y posteriormente entre diferentes imágenes.

$$f_i = \frac{\frac{N_i}{A_i}}{\sum_{j=1}^{4} \frac{N_j}{A_j}}. \tag{4.7}$$

Las proporciones calculadas para cada imagen se grafican en la Fig. 4.13. Esta gráfica solo considera aquellas fijaciones cuyo índice es mayor que 1 y menor o igual que 15[1]. Los resultados demuestran que la gran mayoría de las fijaciones se realizan en la región del borde, y que la tendencia es general para todas las fotografías y todas las topologías analizadas. Por otro lado, y de forma quizás contraria a lo que uno esperaría, se puede

[1]Recordar que la primera fijación siempre se encuentra alrededor del centro de la imagen por lo que considerarla sería un grave error. Por otro lado, se elige el límite superior de 15 ya que existen muy pocas fijaciones cuyo índice esté por encima de este valor, según la Fig. 4.8

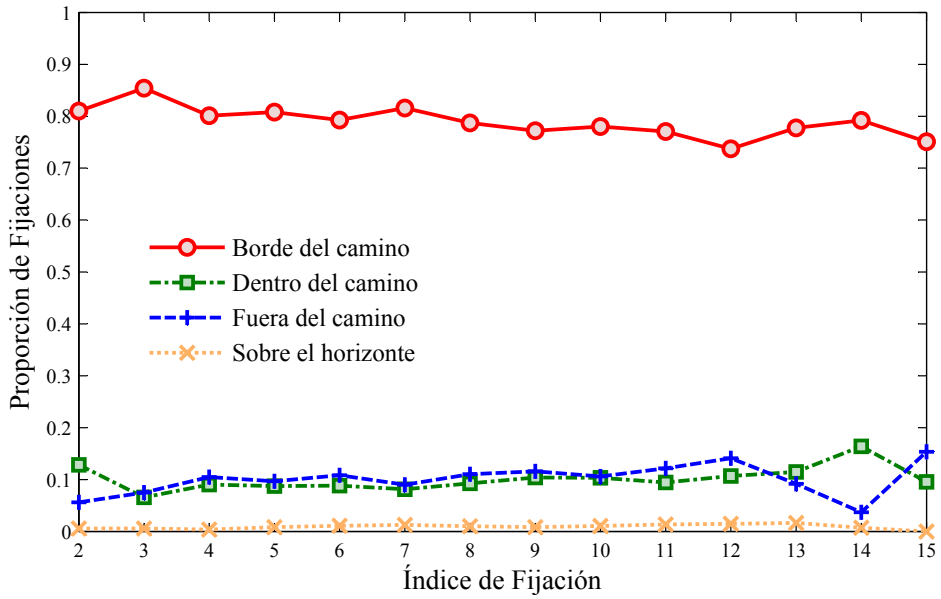


Figura 4.14: Proporción normalizada de fijaciones en el borde del camino (círculos rojos), dentro del camino (cuadrados verdes), fuera del camino (signos + azules) y por sobre el horizonte (signos x naranjas) ordenados por su índice de fijación.

observar que la cantidad de fijaciones dentro del camino no es mayor que la cantidad de fijaciones fuera del mismo, sugiriendo que ambas regiones tienen poca relevancia.

Si las fijaciones se ordenan por su índice en lugar de la imagen en la que se hicieron, se puede observar la evolución de dichas proporciones durante la inspección. Con este propósito, todas las fijaciones con índice j se contaron en cada una de las regiones i, y este valor se dividió por el área en pixeles de la región correspondiente, esto es, $\frac{N_i^j}{A_i}$. Luego se aplicó la siguiente normalización para asegurar que la suma total de las proporciones de la imagen sea 1:

$$f_i^j = \frac{\frac{N_i^j}{A_i}}{\sum_{k=1}^{4} \frac{N_k^j}{A_k}}. \tag{4.8}$$

Finalmente, se calculó el promedio $\bar{f}_i^j$ para todas las fotografías, el cual se grafica en la Fig. 4.14. Estas cuatro curvas muestran de alguna manera la dinámica de las fijaciones y la importancia de cada región en cada una de las etapas de la inspección de la imagen. Se puede ver que los bordes son mucho más relevantes en las primeras fijaciones, perdiendo gradualmente un pequeño porcentaje de fijaciones, para dar lugar a más fijaciones en el

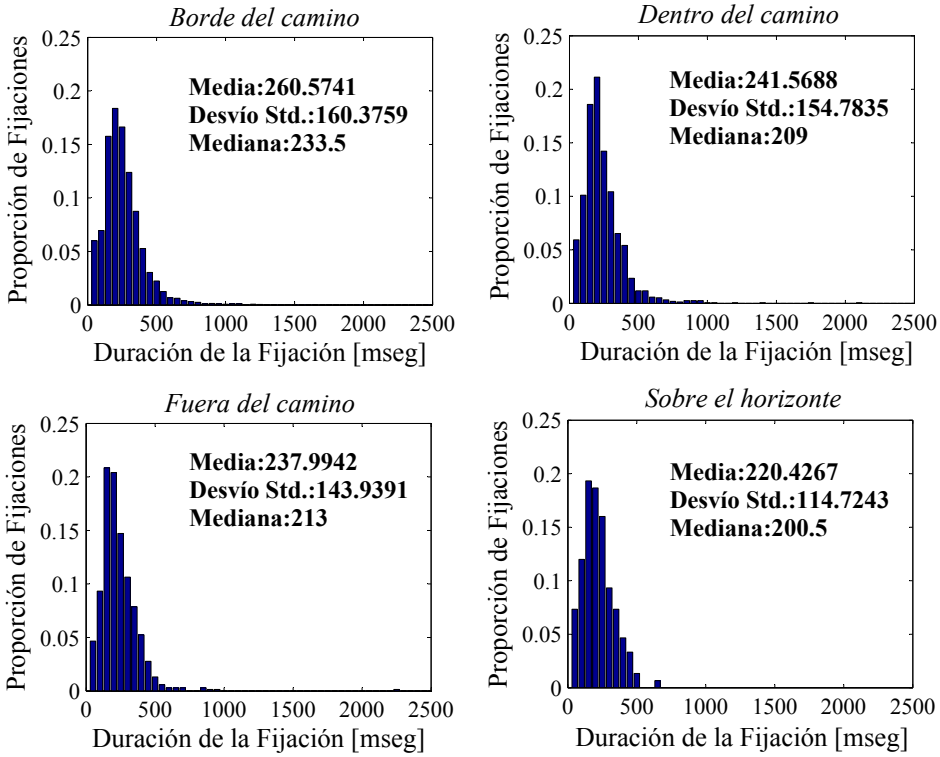


Figura 4.15: Histograma de la duración de las fijaciones en el borde del camino (arriba a la izquierda), dentro del camino (arriba a la derecha), fuera del camino (abajo a la izquierda) y por sobre el horizonte (abajo a la derecha). En cada uno de los casos se indica el valor medio, la mediana y el desvío estándar de la variable.

camino o fuera de él. A pesar de que las fijaciones por sobre el horizonte son marginales, éstas son algo más frecuentes para índices de fijación superiores, ya que para esa etapa de la inspección el participante generalmente ya adquirió la suficiente información para resolver la tarea y la probabilidad que se distraiga con detalles poco relevantes es mayor.

Por otra parte, la Fig. 4.15 incluye los histogramas de la duración de las fijaciones en cada una de las 4 regiones. Si bien no hay un contraste bien marcado entre dichas distribuciones, se pueden marcar algunas diferencias. Principalmente para el borde pero también para dentro del camino la distribución tiene una cola con duraciones más largas, mientras las otras dos distribuciones no la tienen. Esto muestra que las fijaciones más largas se dan con más frecuencia en la región del borde sugiriendo que las personas encuentran allí información de mayor relevancia. Junto a cada histograma se indican el valor medio de la distribución, la mediana y el desvío estándar. Las fijaciones en el borde son en promedio más largas y su variabilidad también es la más alta. Las fijaciones

que se dan por sobre el horizonte son las más cortas dado que en general se deben a distracciones.

Esta sección ha reportado clara evidencia de que las personas encuentran muy atractivas aquellas zonas en el límite entre lo que es transitable y lo que no lo es, para poder resolver la tarea que se le plantea. Este resultado es consistente con la idea intuitiva de que la geometría o la topología del camino está definida por sus bordes. Se muestra aquí que aproximadamente el 90 % de todas las fijaciones del experimento se concentran entre el borde y el camino. Además, en estas regiones las fijaciones son en promedio más duraderas. En la próxima sección solo se considerarán las fijaciones realizadas en estas dos regiones para analizar los patrones visuales que se ven para diferentes topologías.

4.5.3. La influencia de la topología en la distribución de las fijaciones

En la Sección 4.4.2 se discutió con detalle el procedimiento propuesto para separar la región transitable en diferentes subregiones con quizás un significado semántico diferente. Se debe recordar que para el caso considerado de una bifurcación de dos vías estas regiones son los comienzos del camino, los fines de camino, la intersección, la región intermedia y el camino cercano al vehículo. Las fotografías del experimento que en realidad pueden modelarse con una bifurcación de doble vía son 15 del total de 30. Los parámetros del algoritmo que se utilizaron para subdividir el camino son $K_L = 0.85$, $K_I = 0.6$, y $K_B = 0.45$.

Una vez que estas regiones se encuentran definidas para todas las imágenes es posible encontrar la cantidad de fijaciones en cada una de ellas, sin importar quien las hizo o en que imagen en particular se registraron. Se puede calcular la proporción normalizada de fijaciones al dividir el número de fijaciones por el correspondiente tamaño de la región y aplicando nuevamente la Ec. 4.7. La proporción de fijaciones en cada región para las 15 imágenes se grafica en la Fig. 4.16. Se debe aclarar primero que las regiones que corresponden a los finales de camino se separaron en dos grupos: los finales de camino que se pierden en el horizonte en un punto de fuga, y aquellos que salen del campo visual de la cámara. Esto se hizo debido a la notable diferencia entre la cantidad de fijaciones de una y otra, que se discutirá en detalle en la Sección 4.5.4. La región llamada *Comienzo*

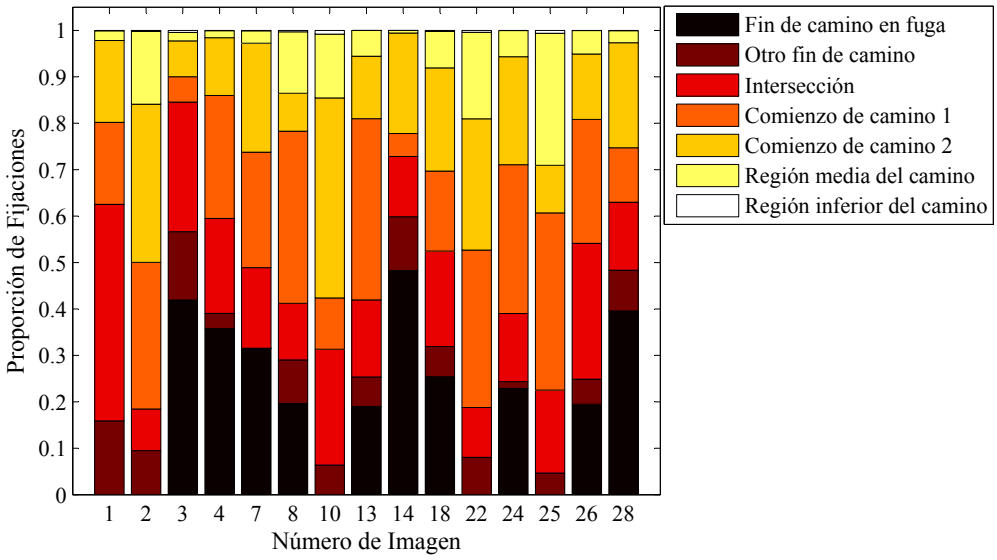
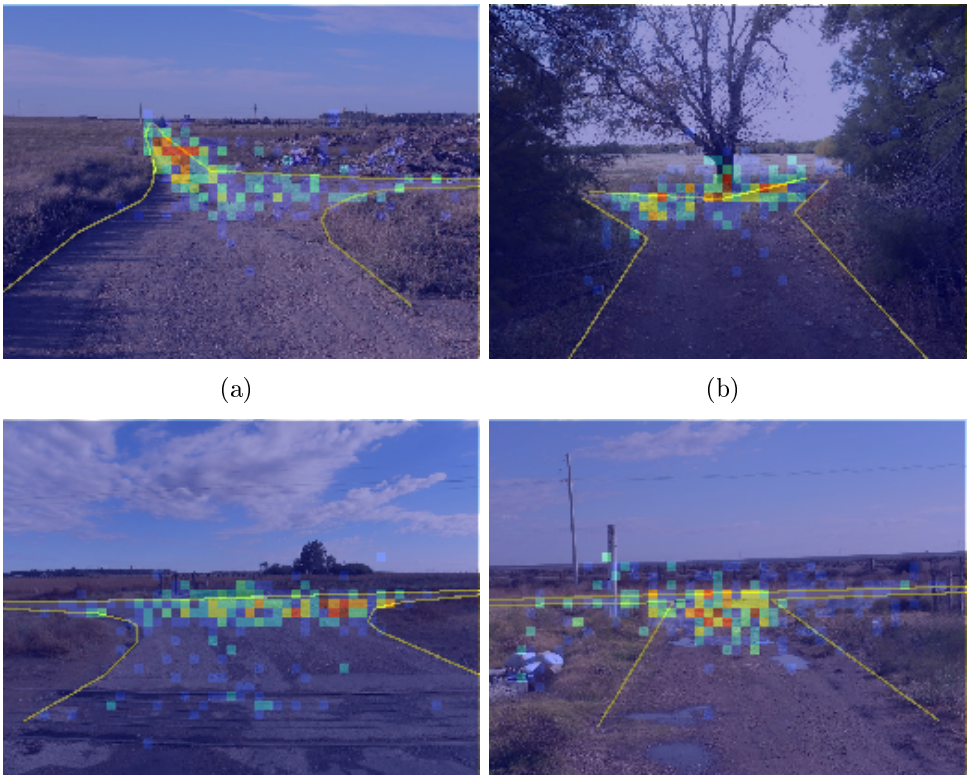


Figura 4.16: Proporción normalizada de fijaciones realizadas en cada una de las subregiones del camino. Los parámetros del algoritmo que se utilizaron para realizar la división son $K_L = 0.85$, $K_I = 0.6$, y $K_B = 0.45$.

1 corresponde al comienzo de aquel camino que fuga hacia el horizonte.

Se puede observar que los resultados no muestran fuertes tendencias para todas las imágenes como fue en el caso de los bordes de camino, en la Fig. 4.13. Ciertamente, la forma del camino tiene un gran impacto en como se distribuyen las fijaciones dentro de él. Aunque todas estas escenas han sido modeladas de forma similar, en realidad no es lo mismo una intersección en $\top$ que una bifurcación en $\curlyvee$ o una intersección en $\vdash$.

Puede verse en el gráfico de barras que las imágenes con mayor proporción de fijaciones en los fines de camino (principalmente los que se fugan en el horizonte) son las indizadas como 3, 4, 14 y 28. En todos esos casos la topología es similar: el camino principal por el que se transita se pierde en el horizonte, mientras que un camino secundario se une en forma casi perpendicular al primero. En este sentido, uno esperaría que las imágenes 13, 24 y 26 mostraran proporciones similares debido justamente a su parecido en la topología, pero sin embargo la proporción de fijaciones en los fines de camino es relativamente baja. Si se las compara con las 4 imágenes mencionadas anteriormente se puede ver que la proporción de la región *Comienzo 1* es significativamente alta. Esto significa que muchas de las fijaciones fueron realizadas en el comienzo del camino en lugar del final ya que dichas regiones son muy próximas entre sí. Esto puede confirmarse a



(a) (b)

(c) (d)

Figura 4.17: El correspondiente histograma 2D de la posición de las fijaciones se superpone sobre (a) la imagen 26, (b) la imagen 1, (c) la imagen 10 y (d) la imagen 25. El tamaño de los bines utilizados para construir el histograma es de 20 pixeles. El color rojo más oscuro se corresponde con el número más alto de fijaciones mientras que el color azul más oscuro se corresponde con el número más bajo de fijaciones.

través del histograma 2D de la posición de las fijaciones que se muestra en la Fig. 4.17(a). Por lo tanto, si se comprende que el límite entre el fin del camino y su comienzo es en realidad difuso se puede concluir que las fijaciones de los participantes se comportan de forma similar con este tipo de topologías.

Respecto a la intersección, las imágenes que mayor proporción de fijación han tenido en dicha región son las etiquetadas como 1, 3 y 26. Estas fotografías tienen en común que la intersección interrumpe la trayectoria natural que propone el camino por el que se transita, y uno debería necesariamente optar por una u otra dirección si se estuviera realmente manejando. Es diferente del caso discutido en el párrafo anterior, donde se podría seguir la trayectoria propuesta por el camino sin ningún tipo de desviación. La Fig. 4.17(b) muestra el ejemplo de la imagen 1.

Las fotografías en las que la región media recibió más fijaciones son las denominadas 2, 8, 10, 22 y 25. Las imágenes 2 y 25 representan intersecciones del tipo $\top$ mientras que las imágenes 8, 10 y 22 muestran curvas a las cuales se une un camino secundario, que resultan ser similares a una topología $\top$. La unión entre la curva y el otro comienzo de camino también son visitados frecuentemente. La atención del participante podría ser atraída hacia allí por los cambios en la orientación de la textura debido a marcas de vehículos en ambos sentidos. El histograma de la posición de las fijaciones para la imagen 10 se muestra en la Fig. 4.17(c).

Para aquellos casos en los que la topología es prácticamente simétrica, es decir que no hay diferencia entre las direcciones izquierda y derecha, se puede observar que los comienzos de camino tienen la misma cantidad de fijaciones, lo que implica que ninguno de los dos es más informativo que el otro. Aquí se refiere a las imágenes 1, 2, 7 y 22. Sin embargo, aunque la imagen 25 tiene una topología simétrica (una intersección en $\top$), el comienzo de la izquierda tiene en proporción muchas más fijaciones. La razón para ésto es que existe un poste de luz que representa una gran distracción para el observador (ver Fig. 4.17(d)). Entonces, pareciera que ante una igualdad de condiciones de ambas regiones para la tarea, las personas prefieren visitar aquellos lugares que maximicen otro tipo de funciones objetivo como por ejemplo la inspección de lugares salientes.

Todo este análisis sugiere que existe una relación entre la topología del camino y la importancia relativa de las diferentes subregiones que la conforman. En la próxima sección se muestra que ciertos lugares de la imagen son más relevantes que otros cuando se inspecciona la imagen.

4.5.4. Los lugares más visitados

En la gran mayoría de las imágenes presentadas durante el experimento se observa que la proporción de fijaciones en ambos fines de camino es dominante. Sin embargo, existe una diferencia significativa en la cantidad de fijaciones que se hacen en el fin de camino que se fuga en el horizonte, respecto al que no lo hace. En este sentido, se estudiaron las fijaciones realizadas en los fines de camino mediante un análisis ANOVA con un solo factor determinado por el tipo de fin de camino (ver Fig. 4.18). Considerando una distribución F con 1 grado de libertad en el numerador y 16 grados de libertad en el denominador, el cociente de variabilidad observado es mucho mayor que el valor crítico

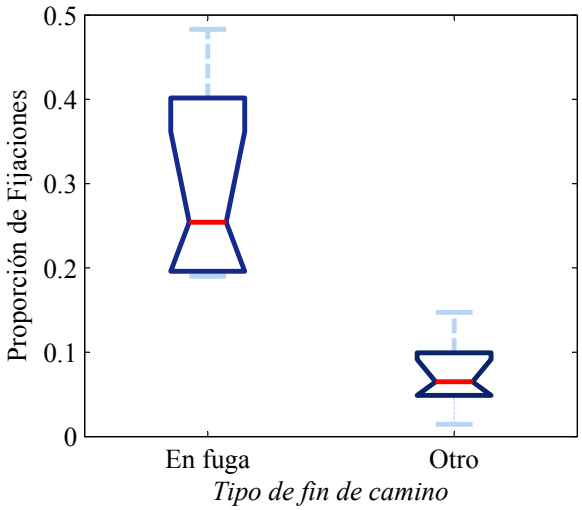


Figura 4.18: La proporción de fijaciones en los fines de camino se estudiaron mediante un análisis ANOVA, resultando en una clara evidencia de que las personas son mucho más atraídas por los caminos que fugan hacia el horizonte.

calculado. Esto es:

$$F\,(1, 17) = 32.45 > 4.494 \tag{4.9}$$

Por lo tanto, la hipótesis nula que considera que ambos datos provienen de una misma distribución puede ser rechazada fuertemente ($p < 3.306e^{-5}$). Puede concluirse entonces que las personas hallan los caminos que fugan hacia el horizonte mucho más atractivos que aquellos se pierden fuera del campo visual de la cámara.

Por otra parte, si las fijaciones se ordenan por su índice se puede analizar su comportamiento dinámico aproximado. La Fig. 4.19 muestra la evolución de la proporción de las fijaciones en cada una de las regiones durante la inspección de la imagen. Estas proporciones se calcularon también haciendo uso de la Ec. 4.8. El *Comienzo 1* junto a la *Intersección* son las regiones más importantes en las fijaciones iniciales. Como se mencionó antes, la primera es el comienzo del camino que se pierde en el horizonte, que se mostró recientemente como el más visitado. La segunda resulta ser un lugar estratégico para extraer información acerca de la orientación de la textura y las direcciones de ambos caminos. Se observa que el interés en el *Comienzo 2* crece cuando éste decrece en el otro comienzo. Aunque el *otro fin de camino* recibe luego más fijaciones, el crecimiento es menor que el que muestra su correspondiente comienzo de camino. Si las fijaciones

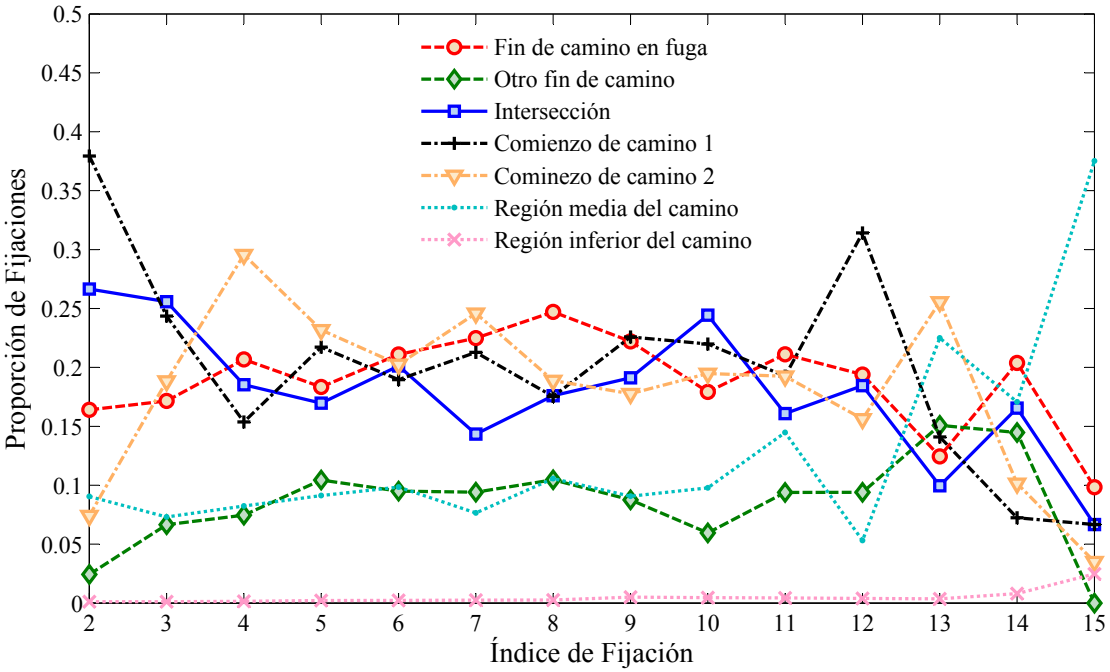


Figura 4.19: Proporción de fijaciones en cada región del camino como función del índice de fijación.

se analizan una a una para cada participante se puede ver que no todas las personas deciden profundizar en el fin del camino una vez que han visitado el comienzo. De hecho, solo alcanza con percibir el comienzo del camino para comprender que allí existe un camino. El nivel de detalle con que cada participante observa la imagen estaría influenciado por la estrategia utilizada, el nivel de concentración y el tipo de personalidad de cada uno, pero el estudio de estos efectos no está considerado dentro de los objetivos de este análisis. La región *media* del camino y el *otro fin de camino* tienen una importancia relativamente baja, mostrando un incremento hacia el final de la inspección. Por último, se muestra que la región *inferior* no tiene ninguna relevancia.

También es posible analizar la importancia relativa de las regiones del camino calculando los histogramas de duración de las fijaciones en cada una de ellas (Fig. 4.20). La región del *fin de camino en fuga* se lleva las fijaciones en promedio más largas ($279.76 msegs$) aunque su distribución tiene una mediana similar a la del *otro fin de camino* y a la del *comienzo 2*. El distribución de la duración en la *intersección* tiene la particularidad de tener la cola más extensa. El desvío estándar es además el más grande de todos. Ambas medidas indican que la mayoría de las fijaciones más largas se realizan

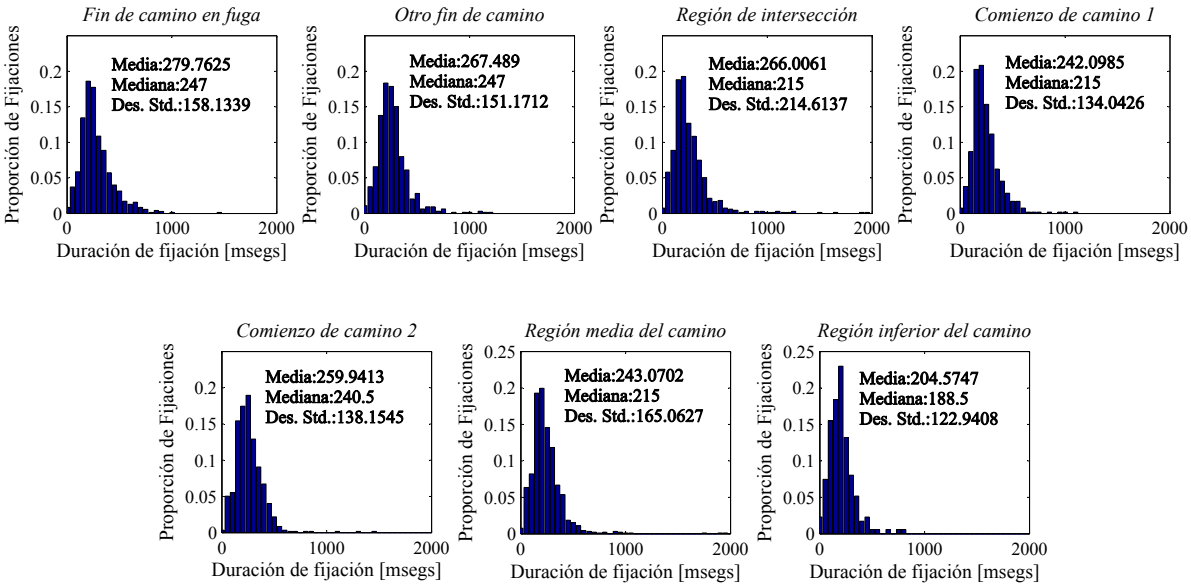


Figura 4.20: Histograma de la duración de las fijaciones hechas en cada una de las regiones del camino. También se indican el valor medio, el desvío estándar y la mediana de la distribución.

en la intersección. Las fijaciones en el *Comienzo 1* y en la *región media* son un poco más cortas que en las regiones anteriores y en general son pasos intermedios para alcanzar las otras regiones importantes. Finalmente, las fijaciones en la *región inferior* del camino son las más rápidas, confirmando que ésta es la región menos relevante para la tarea.

4.5.5. El rol de la saliencia

Los humanos combinan constantemente estrategias del tipo *bottom-up* y *top-down*, resultando sumamente complejo aislar un comportamiento particular, especialmente si los estímulos son generados con fotografías de ambientes reales. Sin embargo, el objetivo aquí es estudiar de forma aproximada como influyen los procesos del tipo *bottom-up* en la tarea visual propuesta, sin ignorar por supuesto su interacción con otros procesos de más alto nivel cognitivo. Con este propósito se estudió el modelo computacional propuesto por [Itti et al. (1998)] para generar mapas de saliencias de una imagen. Se utilizó el *Saliency Toolbox* de Matlab [Walther & Koch (2006)] para calcular los mapas de saliencia de todas las fotografías del experimento. Para esto se utilizaron los parámetros predefinidos de dicho software, excepto por el número de orientaciones de la textura, que

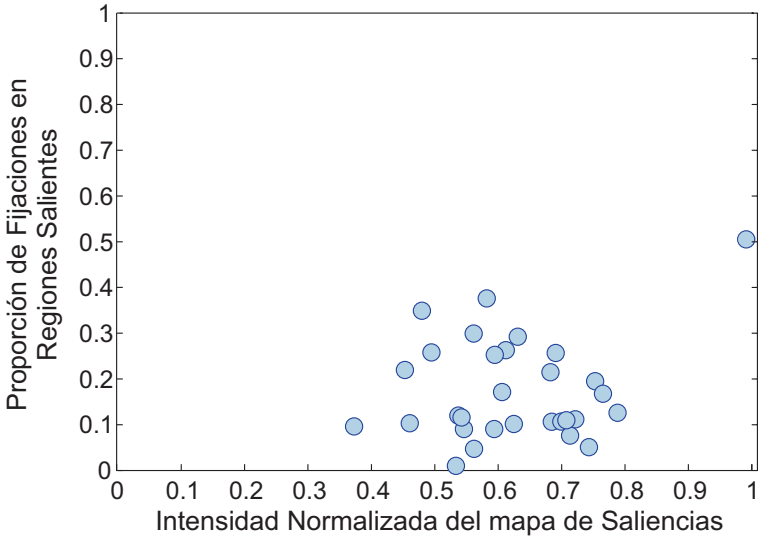


Figura 4.21: Proporción de fijaciones realizadas en las regiones salientes de la imagen como una función de la intensidad del mapa de saliencias. Cada marcador corresponde a una imagen diferente del experimento.

se definió en 8 en lugar de 4. La resolución predefinida del mapa de salida S_{Map} es de 37×50 por lo que cada "pixel" de este mapa representa un grupo de 20×20 pixeles en la imagen original de resolución 1024×768.

Un pixel p_{ij} se define como un pixel saliente si su valor dentro del mapa S_{Map} no es nulo, esto es, $S_{Map}(p_{ij}) > 0$. El área saliente total $A_{S_{Map}}$ de una imagen es la proporción de pixeles salientes en el mapa. La intensidad total del mapa puede calcularse como $I_{S_{Map}} = \sum_{ij} \frac{S_{Map}(p_{ij})}{A_{S_{Map}}}$, para todo pixel p_{ij} perteneciente al mapa.

Para obtener una medida de como los estímulos salientes atraen la atención de los participantes durante la tarea se contó la cantidad de fijaciones en las regiones salientes mencionadas en el párrafo anterior. En la Fig. 4.21 se grafica la proporción de fijaciones en función de la intensidad del mapa $I_{S_{Map}}$ para cada una de las 30 imágenes de la experiencia. La intensidad se normalizó a valores entre 0 y 1, tomando el valor 1 la imagen más saliente de las 30. En primer lugar, se observa que los voluntarios no miran los lugares salientes con mayor frecuencia que aquellos que no lo son, tomando la proporción el valor 0.5 en el caso más extremo. Si las saliencias tuvieran un rol importante para la tarea se esperaría que la proporción de fijaciones en dichos lugares se incrementara a medida que la intensidad del mapa es mayor. Sin embargo, la evidencia muestra aquí que no hay una relación directa entre ellos. El coeficiente de correlación para ajuste

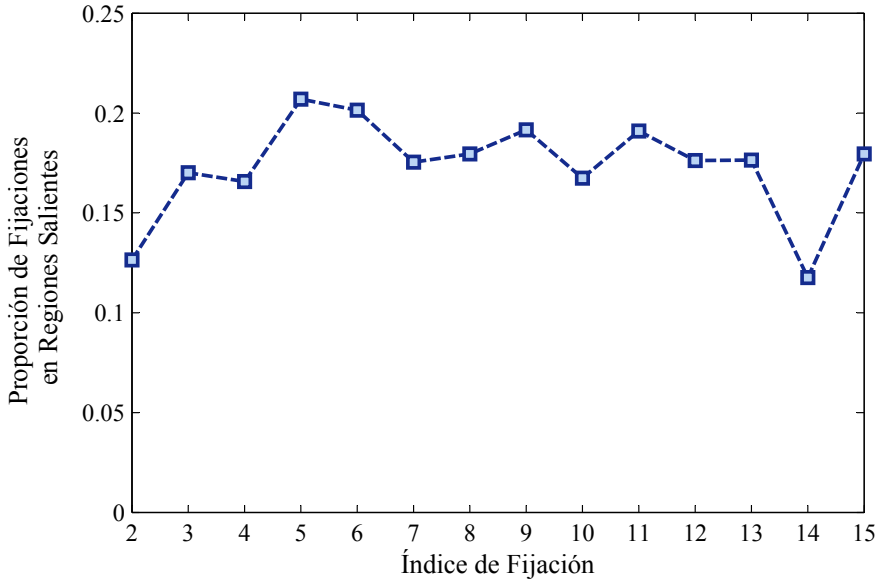


Figura 4.22: Proporción de fijaciones registradas en las regiones salientes de la imagen en función de su índice de fijación.

mediante regresión lineal es $\rho = 0.2239$.

Con respecto a la duración de las fijaciones, no es posible marcar diferencias ya que las distribuciones tanto para las regiones salientes como para las no salientes son muy similares. Los valores medios son $255.1 msegs$ para los lugares salientes y $253.8 msegs$ para el resto de las fijaciones. Por lo tanto, no es posible concluir aquí que las regiones salientes sean en promedio más informativas que el resto de la imagen.

Por otro lado, si se ordenan las fijaciones por su índice, se pueden contar la cantidad de fijaciones en las saliencias N_s^j para cada índice j. La proporción de fijaciones se puede calcular de la siguiente manera dividiendo por el total de fijaciones que tengan el mismo índice j (N^j):

$$p_s^j = \frac{N_s^j}{N_j} = \frac{\sum_{i=1}^{30} N_{s_i}^j}{\sum_{i=1}^{30} N_i^j}, \tag{4.10}$$

donde i representa el índice de la imagen. En la Fig. 4.22 se demuestra que los lugares salientes son en promedio menos interesantes para las personas al comienzo de la inspección de la imagen. Teniendo en cuenta que las primeras fijaciones son aquellas que proveen a la persona la mayor cantidad de información nueva acerca de la imagen, estos

resultados refuerzan la idea de que las saliencias no tienen aquí un rol importante.

4.6. La implicancia de los resultados

Los resultados del experimento han confirmado que las personas utilizan ciertas características que se encuentran en lugares específicos de la imagen para poder luego determinar el tipo de topología que presenta la escena. Esto es otro logro importante de esta tesis, y la comprobación de esto fue una de las principales motivaciones para el diseño y el desarrollo de las actividades experimentales con el equipo para el seguimiento de ojos.

También es interesante enfatizar que las personas pueden claramente resolver la misma tarea visual mediante diferentes estrategias. Existen elementos que demuestran que se pueden obtener resultados similares utilizando diferente cantidad de fijaciones, dando evidencia de que un número mayor de éstas no asegura un buen rendimiento en la experiencia (ver Fig. 4.9).

Todas estas evidencias confirman la existencia de ciertas regiones de la imagen que atraen la atención de los participantes. Se puede observar que los bordes del camino recibieron una gran proporción de fijaciones, sin importar la fotografía o la topología analizada. Además, estas fijaciones fueron en promedio las más largas en tiempo. Todo esto indica que las personas consideran muy importantes para detectar la topología todas aquellas zonas en el límite entre lo transitable y lo que no lo es. Desde un punto de vista computacional, el agrupamiento de zonas de similar apariencia en una imagen y la detección de regiones de alto contraste de color, intensidad o textura pueden ser útiles para identificar aquellas regiones de borde donde extraer la información de más alto nivel que interesa.

Por otro lado, una vez que se dividió el camino en regiones se encontró que los caminos que fugan hacia el horizonte atraen mucho la atención de los participantes. Se mostró también que los patrones que muestran las fijaciones resultan similares cuando la topología del camino en escena es similar. Por estos motivos, tanto el estudio de las orientaciones de la textura como el cómputo de los puntos de fuga serán considerados a la hora de diseñar los algoritmos para la detección automática de topologías.

A partir de un análisis dinámico de las fijaciones se encontró evidencia de que ciertas

regiones como los comienzos de los caminos que fugan hacia el horizonte y las regiones de intersección de los caminos son las regiones más visitadas en las primeras fijaciones. En [Tatler et al. (2005)] se asigna un valor fundamental a las primeras fijaciones respecto de la percepción y la comprensión de la información visual. Aquí también jugarían un papel muy importante las orientaciones de la textura para determinar aquellas regiones donde existe evidencia de cambios de dirección en los caminos.

Por otra parte, en la literatura del tema se han propuesto varias aproximaciones para el modelado de la atención visual. Uno de los objetivos de este estudio era obtener un mejor entendimiento de como los procesos del tipo *bottom-up* influyen en esta tarea visual particular, y como se vinculan con procesos de más alto nivel cognitivo. Del análisis realizado utilizando los mapas de saliencia obtenidos con el modelo computacional de [Itti et al. (1998)] se demostró que las regiones salientes de la imagen no tienen relevancia alguna para la tarea, además de que no se encontró una relación aparente entre la intensidad del mapa y la cantidad de fijaciones en pixeles salientes. También se probó que las saliencias son menos visitadas en las fijaciones iniciales. Todo esto sugiere dos posibles conclusiones: o bien las estrategias de inspección utilizadas son principalmente un reflejo de procesos del tipo *top-down*, o los estímulos visuales que el sistema visual humano utiliza para esta tarea no pueden ser detectados por el modelo computacional utilizado. Para concluir definitivamente que las saliencias no tienen influencia en la tarea debería realizarse un análisis exhaustivo de los resultados con otros modelos de la atención visual que consideren otro tipo de características de la imagen o bien incluyan construcciones de más alto nivel como por ejemplo la detección de líneas o esquinas. Este último tipo de estímulos salientes alineados suele estar presente en los límites de los caminos por lo que serán también considerados en el diseño de los algoritmos correspondientes.

Para finalizar, además de los elementos cuantitativos que se han reportado es interesante remarcar algunos resultados cualitativos aunque es válido aclarar que para su correcta justificación sería necesario realizar nuevos experimentos. En primer lugar, existe una aparente relación entre la personalidad de un individuo y su comportamiento durante la inspección. Aquellos participantes que son meticulosos y detallistas típicamente realizan un número mayor de fijaciones, por ejemplo al profundizar en los fines de camino. Sin embargo, otras personas prefieren fijar durante mayor tiempo en lugares

estratégicos como las intersecciones y los comienzos de camino. Por otro lado, las personas suelen adaptar el número de fijaciones a la complejidad de la fotografía, aunque ésto también depende de la experiencia que tuvieron con la imagen inmediatamente anterior. Esto es, si el participante comprende que el número de fijaciones no fue suficiente para poder elegir certeramente una topología, es muy probable que en la siguiente imagen adopte una estrategia de inspección más agresiva incrementando el número de fijaciones. Por último, en el caso de las topologías que son simétricas, se observó que cuando dos regiones parecen ser igualmente atractivas para la tarea, las personas suelen visitar aquella que le permita alcanzar otros objetivos secundarios como por por ejemplo la inspección de pixeles salientes. Éste es un tema a considerar en futuras investigaciones.

4.7. Resumen

En este capítulo se han descrito en profundidad las actividades experimentales llevadas a cabo para el estudio de los patrones de atención visual de las personas cuando identifican una topología de camino en una imagen estática. Se detallaron todas las consideraciones tenidas en cuenta para el diseño y la interpretación de los datos estadísticos generados. De los resultados analizados se desprende la existencia de determinados elementos de la imagen que tienen mayor relevancia para las personas al momento de identificar la topología correspondiente. Por último, se analizaron las implicancias de dichos resultados a la hora de diseñar un algoritmo para la detección automática de topologías a partir de una imagen.

A continuación se exponen las conclusiones finales y se especifican los trabajos a futuro que se desprenden de esta tesis de investigación. Finalmente, se realiza una valoración personal acerca del estado del arte de los sistemas de percepción en la actualidad y cuál será el rol que tendrá la visión artificial en los futuros vehículos autónomos.

Capítulo 5

Conclusiones y perspectivas a futuro

La aparición de los vehículos que se conducen solos por las calles de una ciudad ya dejó de ser parte de la ciencia ficción para ser una realidad. Ya sea por iniciativas militares o por el anhelo de reducir el peligro y los costos que implica hoy en día conducir un vehículo, los investigadores se han enfocado durante años hacia el desarrollo de sistemas capaces de participar del tránsito con una mínima o nula intervención de un humano. Existen muchos proyectos en vigencia, impulsados por centros de investigación o la misma industria automotriz, cuyos prototipos han demostrado experimentalmente diferentes grados de autonomía en la conducción. Sin embargo, los costos que en general implica acondicionar un vehículo para que se comporte de forma autónoma son aún prohibitivos a la hora de pensar en un producto final para un usuario común y corriente.

Hoy en día el diseño y el desarrollo de estos vehículos está enmarcado bajo la premisa de que deben estar preparados para interactuar con otros vehículos conducidos por humanos y en las mismas condiciones que lo hacen ellos. Dado que no es posible a mediano plazo contar con una infraestructura suficientemente inteligente para guiar a los vehículos, es necesario que éstos tengan la capacidad de percibir su entorno, identificando tanto el terreno sobre el que pueden transitar como los posibles obstáculos. Con la asistencia de mapas digitales de alta precisión, sensores LIDAR de alto costo y GPS, los proyectos más maduros han experimentado con éxito la navegación autónoma por ambientes poco estructurados y también en ambientes complejos como los urbanos. Sin embargo, existen cuestiones que tienen que ver con la verdadera comprensión de la situación que presenta la escena frente al vehículo que aún no han sido resueltas. Para estos problemas la utilización de sistemas de visión parece ser la herramienta adecuada, sumando el hecho de que todo tipo de estructura existente alrededor de un camino está particularmente diseñada para ser vista por los conductores.

La detección de caminos es una de las aristas de la comprensión de escena que ha

sido afrontada principalmente como un problema de visión monocular. Aunque la extracción de información de alto nivel a partir de una imagen no es sencilla, la utilización de técnicas del *machine learning* y el filtrado estadístico, sumada a la disponibilidad de sistemas computacionales cada vez más potentes han permitido un gran avance en la búsqueda de soluciones para este problema. A lo largo de esta tesis se han estudiado en profundidad los diferentes enfoques que existen en la literatura, lográndose identificar y analizar los componentes principales de un sistema genérico para la detección de caminos en imágenes. Se estudiaron las ventajas y desventajas de cada alternativa y se evaluó su comportamiento ante las diferentes problemáticas que deben enfrentar estos sistemas. Dichas complicaciones son derivadas de los cambios bruscos en las condiciones de iluminación, climáticas y en el tipo de camino y/o ambiente. Se marcó además la necesidad de nuevos bloques funcionales dentro del sistema genérico.

Los sistemas que han demostrado una mayor robustez ante los diferentes cambios de iluminación y del clima se basan principalmente en un modelo geométrico para el camino, representado por algún tipo de curva. La estimación y el seguimiento de los parámetros de la curva a través del filtrado bayesiano permite considerar formalmente el error en dichas estimaciones y agiliza los cálculos de los parámetros al integrar temporalmente la información a lo largo de una secuencia de imágenes. El principal inconveniente que padece este enfoque aparece cuando la forma del camino cambia de tal manera que ya no es posible modelarla con el tipo de curva considerado. Esto es muy común en las intersecciones, bifurcaciones y rotondas, que no pueden modelarse con una única curva y que se ha denominado a lo largo de la tesis como caminos que tienen una *topología no lineal*. Este tipo de situaciones ha sido poco estudiada en la bibliografía por lo que casi no existen propuestas de algoritmos capaces de detectarlas mediante el procesamiento de una imagen.

Dado que las personas resuelven a menudo situaciones similares, esta tesis focalizó muchos de sus esfuerzos en comprender cómo lo hacen a partir del estudio de los patrones de atención visual. Con este objetivo se diseñó y llevó a cabo un experimento único en su tipo, que permitió obtener evidencias acerca de los elementos visuales que mayor relevancia tienen para la persona mientras intenta reconocer la topología de un camino en una imagen. El análisis de los datos registrados no fue una tarea sencilla ya que se observó una notoria variabilidad en los patrones visuales resultantes, que no solo

depende de la variación natural que se da cuando se consideran participantes y fotografías distintas, sino que también incluye cuestiones como la estrategia de inspección, el nivel de concentración, y la personalidad de cada individuo. Para poder comparar las fijaciones de las personas en diferentes imágenes con diferentes topologías de caminos y además de diferentes dimensiones, se propusieron dos métodos para separar cada una de las imágenes en regiones que tuvieran el mismo significado para todas. Esto resultó fundamental para poder comparar las coordenadas de las fijaciones y establecer tendencias a través de indicadores estadísticos, sin depender de qué persona las realizó o en qué imagen se hicieron.

En primer lugar, se encontró que para todos los casos una mayoría contundente de fijaciones se realizó en la zona de los bordes del camino, que se definió como una región en el límite entre el camino y lo que no lo es. Además, las fijaciones en dicha zona son más duraderas, dando mayor soporte a la idea de que esta región tiene mucha información relevante para determinar la topología. El borde se caracteriza por algún tipo de contraste entre la región del camino y sus alrededores, ya sea de colores, texturas, materiales o de la altura que se percibe. Aunque este contraste puede ser bien marcado o suave el sistema visual humano puede percibirlo fácilmente.

Luego, al dividir el camino de la imagen en subregiones se hallaron determinados lugares que son más vistos que otros. Es el caso de los fines de camino que fugan hacia el horizonte, que reciben muchas más fijaciones que aquellos fines de camino que salen del campo visual de la imagen. Se ha visto en los capítulos introductorios que las personas pueden percibir distancias a partir de la información de la perspectiva de la imagen, entre otras cosas. En tal sentido, los resultados mostrados aquí sugieren que los participantes utilizarían indicios de convergencia de la textura para proyectar posibles trayectorias hacia adelante. Cuando esta trayectoria virtual se ve interrumpida a causa de una intersección o una bifurcación, el ojo suele detenerse en dichos lugares con mayor frecuencia, tal como también se ha demostrado. En los lugares de intersección los diferentes patrones de la orientación de la textura podrían ser utilizados para determinar la dirección de los diferentes caminos.

Tal como se mencionó en el párrafo anterior, los resultados muestran que los patrones de atención visual están influenciados por el tipo de topología presente en la imagen.

Además se encontraron patrones de fijaciones similares entre imágenes que tienen justamente caminos con topologías similares. Esto sugiere que la estrategia de inspección estaría fuertemente afectada por procesos del tipo *top-down* en los que prevalece el conocimiento previo acerca la forma y aspecto que suelen tener los caminos.

Cuando se analizaron las fijaciones desde un punto de vista dinámico, se demostró que los patrones tienen un primera etapa donde se inspecciona la imagen, y luego se realizan refijaciones en aquellos lugares previamente observados y que tienen mayor relevancia. En esta última etapa es donde la memoria de corto plazo podría tener mucha más participación. Estos resultados se han obtenido mediante el estudio de la amplitud de los sacádicos y de la dispersión en la posición de las fijaciones. Respecto de las subregiones del camino, se observó que las intersecciones y los comienzos de camino reciben la atención en primer lugar. En particular, el comienzo del camino que fuga al horizonte recibe muchas más fijaciones que el otro comienzo de camino, sugiriendo que desde los primeros instantes y mediante la visión periférica, la persona percibe cierta convergencia de las líneas de la textura.

Por otro lado, se hizo también hincapié en estudiar cual es el efecto de las regiones salientes de una fotografía en los patrones de movimiento de los ojos. Los grupos de pixeles que sobresalen de la imagen se hallaron a partir de uno de los modelos computacionales más reconocidos en la literatura para la detección de estímulos del tipo *bottom-up*. Al analizar los datos experimentales respecto de dichas saliencias se observa claramente que no tienen influencia alguna en el comportamiento de los participantes, reforzando la hipótesis de que los procesos *top-down* son dominantes para la tarea.

5.1. Trabajos futuros

A lo largo de esta tesis se han estudiado en profundidad los sistemas basados en visión para la detección de caminos. Debido a las limitaciones que aún muestran estos sistemas se justificó la necesidad de nuevas estrategias de alto nivel que permitan obtener una mayor comprensión de la situación presente en la imagen. En particular, uno de los problemas a resolver es la detección de la topología de un camino cuando no se dispone de un mapa preciso del lugar. Con este objetivo se planteó entonces utilizar las técnicas de seguimiento ocular para estudiar y comprender un poco más acerca de como las

personas resuelven esta tarea.

La posibilidad y la disponibilidad de usar un equipo de estas características ofrece nuevas chances para continuar investigando las diferentes capacidades de la visión humana en pos del diseño de sistemas de percepción cada vez más robustos. Dado que éste fue el primer experimento de este tipo llevado a cabo dentro del grupo es posible encontrar muchas oportunidades para las mejoras a corto plazo. Si bien no resultó sencillo alcanzar los resultados que aquí se muestran, este proceso permitió generar y acumular mucha experiencia tanto en la realización del experimento como en el posterior análisis de los datos. El aprendizaje alcanzado abre la puerta hacia nuevas actividades experimentales que permitan reforzar aún más los logros aquí obtenidos.

Una de las próximas experiencias incluirá un procedimiento similar al que aquí se ha propuesto pero solo se estudiará un conjunto finito de topologías, representadas cada una de ellas por muchas más fotografías. Esto permitirá estudiar con mayor detalle cada caso, reforzando estadísticamente algunas tendencias encontradas. Otra de las actividades planeadas pretende extender el estudio del impacto de los estímulos del tipo *bottom-up* para incluir construcciones como segmentos alineados o líneas, que no están contempladas en el modelo computacional utilizado y que si podrían ser importantes para el observador al momento de determinar una topología. En tercer lugar, y siempre bajo el mismo contexto, se diseñará un nuevo experimento para estudiar en detalle como influye la orientación de la textura de la imagen en los sacádicos generados durante la inspección de imágenes y video.

Por otra parte, una de las motivaciones principales para la realización del experimento reportado era encontrar evidencias que fueran útiles e inspiradoras para el delineamiento de un sistema capaz de detectar de forma automática la topología de un camino a partir de la imagen. Por lo tanto, los pasos naturales a seguir involucran el diseño y la implementación de un algoritmo de procesamiento de imágenes que permita estimarla.

Las bases de este algoritmo estarán sustentadas en dos fuentes principales de información como son el contraste y la textura. Dicha información se obtendrá a través de procesamientos tanto de forma localizada o en regiones de interés como de forma global para toda la imagen, tal como lo hace el sistema visual humano. El contraste de apariencia estará definido en función del color y de la varianza en las direcciones de la textura de los pixeles. Esto tiene relación con el hecho de que en un ambiente poco estructurado

la dirección de las líneas de la textura del camino suelen ser mucho más uniformes que fuera de él. Luego, un análisis de la orientación de la textura en diferentes regiones de interés de la imagen permitirá estimar las zonas más transitadas que se utilizarán como indicio para determinar las diferentes trayectorias que los vehículos siguen en tal lugar.

El algoritmo estará enfocado a detectar los indicios en la imagen que sugieran un cambio en la forma de los caminos. Se diseñará como un sistema complementario o supervisor del sistema base para la detección de caminos, de tal manera que proponga los cambios necesarios en el modelo geométrico considerado a una frecuencia suficientemente rápida que permita planificar con tiempo la correspondiente maniobra. La información provista por el algoritmo de detección del camino se utilizará para restringir las regiones a procesar y para agilizar los tiempos de cómputo involucrados.

5.2. Comentarios finales

En este punto es importante destacar el gran avance que ha tenido el desarrollo de los vehículos autónomos durante la última década. Las demostraciones exitosas que se han realizado en ambientes sumamente complejos como los urbanos no dejan de sorprender al mundo e invitan a pensar en un futuro no muy lejano en el que este tipo de vehículos sean participes de la vida cotidiana de las personas. Sin embargo, bajo una mirada quizás algo escéptica, uno podría decir que pasarán muchos años más hasta que alguno de estos proyectos pueda transicionar de la etapa del prototipado y la investigación hacia un producto comercial al alcance de todos.

La robustez y la confiabilidad son esenciales para una implementación comercial. Pero uno de los principales enemigos de estas iniciativas sigue siendo el costo de los sensores que impide que estos sistemas puedan generar un impacto real en la sociedad. Si bien el avance tecnológico permitirá una reducción del valor de sensores como el RADAR y el LIDAR, y un aumento en la precisión de los mismos, el costo de un sistema de cámaras de video es un par de ordenes de magnitud más bajo que un sensor LIDAR 3D por ejemplo. Además tiene la ventaja de ser un sensor totalmente pasivo que consume muy poca energía y no es potencialmente peligroso, como es el caso de los otros dos en modos de alta potencia. Sumando ésto al resto de las ventajas mencionadas durante este informe, se podría pensar en la visión como una de las principales herramientas a utilizar

en los futuros sistemas de percepción y comprensión autónoma del área transitable y su contexto. Mientras tanto la aplicación de visión monocular y estéreo ya se considera una práctica estándar en la asistencia al conductor de vehículos de mediana y alta gama [Gat et al. (2005)]. También existen desarrollos de nuevas modalidades de sensado como los sonares acústicos de fenómeno *micro-doppler*, que permiten realizar un análisis acústico de la escena y reconocer diferentes objetos como animales, personas, vehículos, etc., a partir de sus patrones de movimiento [Zhang et al. (2007), Zhang & Andreou (2008)].

Como se ha visto en el capítulo introductorio, el sistema de percepción de un vehículo autónomo debe ser capaz, entre otras cosas, de: estimar la posición y la velocidad propia, detectar y seguir a los demás participantes del tráfico, detectar obstáculos, estimar la forma o la geometría del camino, y localizarse dentro de un mapa. Para muchos de estos problemas la visión ya proporciona soluciones relativamente maduras, entre las que se puede mencionar los sistemas para la detección de carril, detección de señalización, detección de luces de tránsito, mejoras en la visibilidad, detección y seguimiento de otros vehículos, y la detección y seguimiento de peatones. Si bien la detección y seguimiento de la forma y la topología de los caminos aún no está totalmente resuelta, a partir de lo aquí estudiado se puede decir que dichos sistemas ya se encuentran muy avanzados.

La comprensión completa de una escena mediante una imagen implica el diseño de sistemas robustos y eficientes que incluyan no solo todos los módulos que se mencionaron en el párrafo anterior, sino también los mecanismos que definan como es su interacción. También serán necesarias implementaciones muy eficientes junto a sistemas poderosos de procesamiento de tal forma de asegurar que los tiempos de respuesta sean los mínimos indispensables según los correspondientes estándares de seguridad vial. Por último, será imprescindible trabajar en pos de garantizar la máxima confiabilidad posible de éstos sistemas para que sean totalmente seguros para el usuario. Para esto, la emulación del comportamiento del ser humano seguirá siendo primordial.

Apéndice A

Material fotográfico y patrones de fijación asociados

Para referencia del lector se incluyen las 33 fotografías de la experiencia (incluyendo las 3 del entrenamiento) y las pantallas de opciones de respuesta asociadas. Las regiones más relevantes de la imagen se muestran a través del histograma de la posición de las fijaciones superpuesto sobre cada una de las fotografías. El tamaño de los bines del histograma es de 20x20 pixeles. El color rojo más oscuro indica el valor más alto de fijaciones mientras que el color azul más oscuro indica el valor mas bajo de fijaciones contadas. La primer fijación se descarta por razones descritas en la sección 4.3.4.

A.1. Imágenes para el entrenamiento

A.2. Imágenes para el experimento

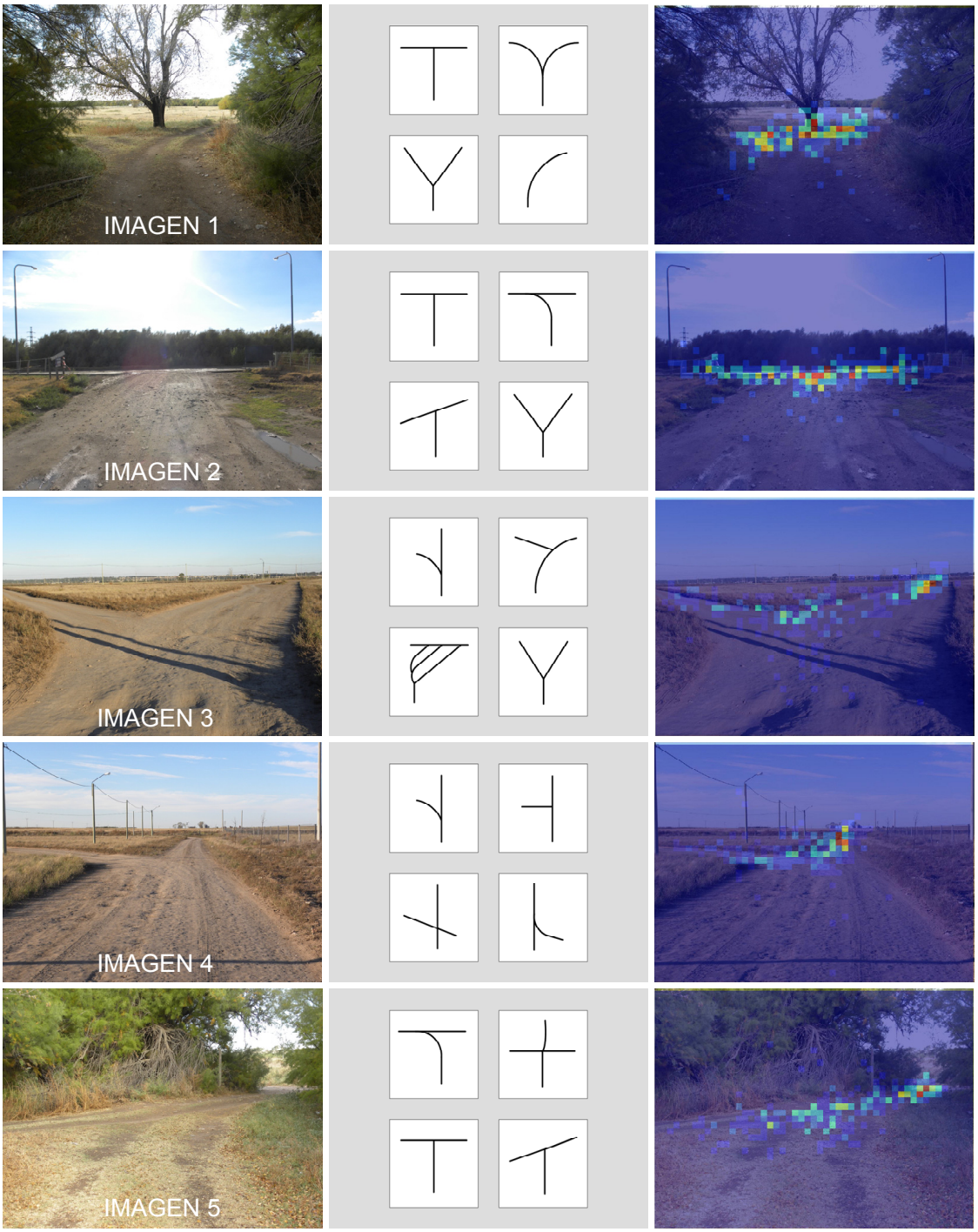



IMAGEN 6
IMAGEN 7
IMAGEN 8
IMAGEN 9
IMAGEN 10


IMAGEN 11
IMAGEN 12
IMAGEN 13
IMAGEN 14
IMAGEN 15


IMAGEN 16
IMAGEN 17
IMAGEN 18
IMAGEN 19
IMAGEN 20


IMAGEN 21
IMAGEN 22
IMAGEN 23
IMAGEN 24
IMAGEN 25

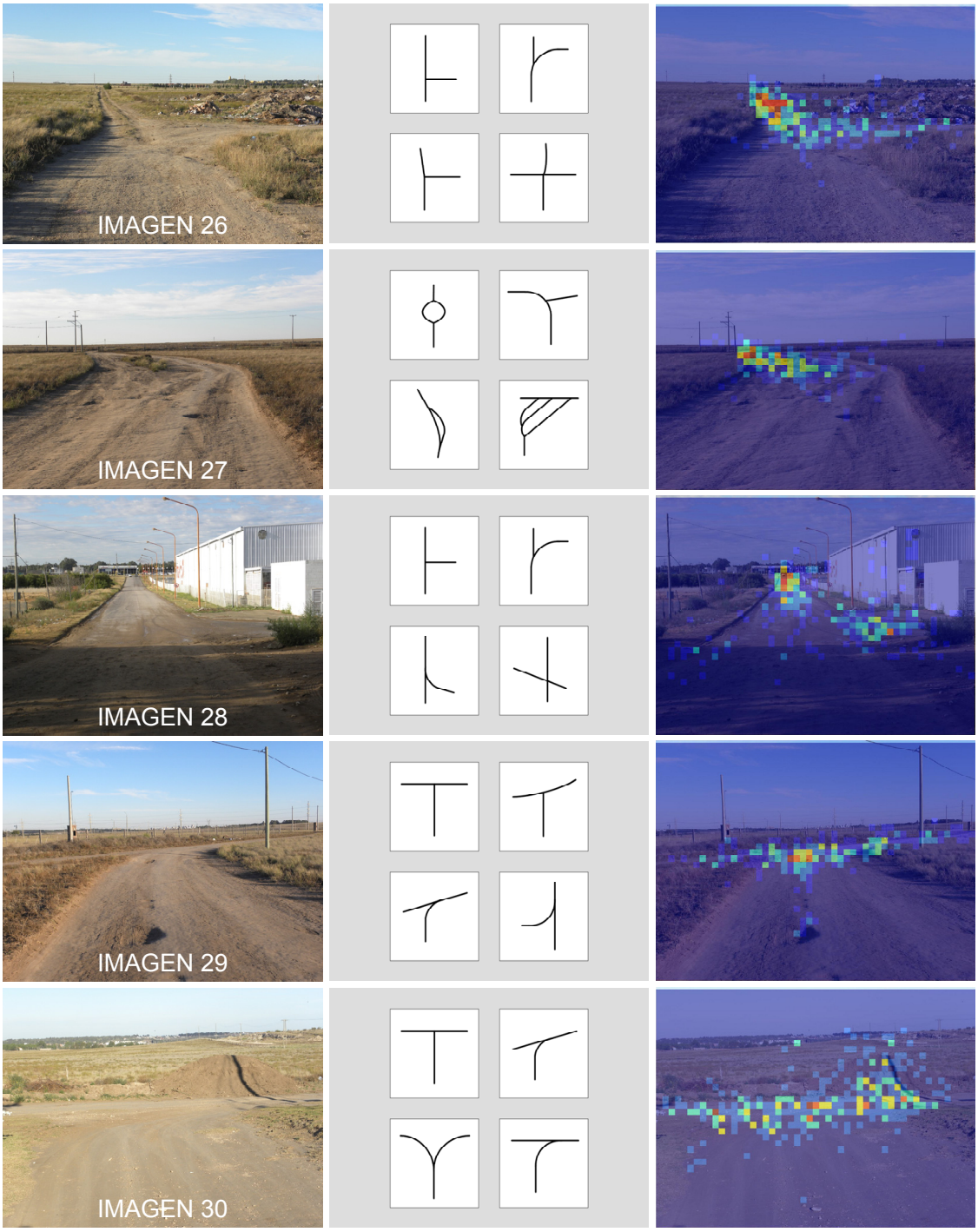


IMAGEN 26
IMAGEN 27
IMAGEN 28
IMAGEN 29
IMAGEN 30

Bibliografía

Ackerman, E. (2013). UK unveils 'affordable' self-driving RobotCar. Publicado 19 Feb. 2013, en IEEE Spectrum Automaton Blog. Disponible en http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/uk-affordable-self-driving-robotcar/?utm_source=roboticsnews&utm_medium=email&utm_campaign=021913.

Alon, Y., Ferencz, A., & Shashua, A. (2006). Off-road path following using region classification and geometric projection constraints. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* (pp. 689–696).

Alvarez, J., Lopez, A., & Baldrich, R. (2008). Illuminant-invariant model-based road segmentation. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 1175–1180).

Apostoloff, N. & Zelinsky, A. (2003). Robust vision based lane tracking using multiple cues and particle filtering. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 558–563).

Bai, L., Wang, Y., & Fairhurst, M. (2010). Multiple condensation filters for road detection and tracking. *Pattern Analysis and Applications*, Vol. 13 (Num. 3), (pp. 251–262).

Ballard, D. H. & Brown, C. M. (1982). *Computer Vision*. Prentice-Hall.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, Vol. 15 (Num. 4), (pp. 600–609).

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, Vol. 29 (Num. 2), (pp. 529–535).

Bar Hillel, A., Lerner, R., Levi, D., & Raz, G. (2012). Recent progress in road and lane detection: a survey. *Machine Vision and Applications*, (pp. 1–19).

Bertozzi, M., Bombini, L., Broggi, A., Buzzoni, M., Cardarelli, E., Cattani, S., Cerri, P., Coati, A., Debattisti, S., Falzoni, A., Fedriga, R., Felisa, M., Gatti, L., Giacomazzo, A., Grisleri, P., Laghi, M., Mazzei, L., Medici, P., Panciroli, M., Porta, P., Zani, P., & Versari, P. (2011). VIAC: An out of ordinary experiment. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 175–180).

Bertozzi, M., Broggi, A., & Fascioli, A. (2000). Vision-based intelligent vehicles: State of the art and perspectives. *Robotics and Autonomous Systems*, Vol. 32 (Num. 1), (pp. 1–16).

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer Science+Business Media.

Broggi, A. & Cattani, S. (2006). An agent based evolutionary approach to path detection for off-road vehicle guidance. *Pattern Recognition Letters*, Vol. 27 (Num. 11), (pp. 1164–1173).

Caraffi, C., Cattani, S., & Grisleri, P. (2007). Off-road path and obstacle detection using decision networks and stereo vision. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8 (Num. 4), (pp. 607–618).

Cerri, P., Soprani, G., Zani, P., Choi, J., Lee, J., Kim, D., Yi, K., & Broggi, A. (2011). Computer vision at the hyundai autonomous challenge. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems,* (pp. 777–783).

Chapuis, R., Aufrere, R., & Chausse, F. (2002). Accurate road following and reconstruction by computer vision. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3 (Num. 4), (pp. 261–270).

Chen, Q. & Wang, H. (2006). A real-time lane detection algorithm based on a hyperbola-pair model. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 510–515).

Chen, S. (2012). Kalman filter for robot vision: A survey. *IEEE Transactions on Industrial Electronics*, Vol. 59 (Num. 11), (pp. 4409–4420).

Chiku, T. & Miura, J. (2012). On-line road boundary estimation by switching multiple road models using visual features from a stereo camera. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems,* (pp. 4939–4944).

Crisman, J. & Thorpe, C. (1991). UNSCARF-a color vision system for the detection of unstructured roads. In *Proceedings of the IEEE International Conference on Robotics and Automation (Vol. 3),* (pp. 2496–2501).

Crisman, J. & Thorpe, C. (1993). SCARF: a color vision system that tracks roads and intersections. *IEEE Transactions on Robotics and Automation*, Vol. 9 (Num. 1), (pp. 49–58).

Dahlkamp, H., Kaehler, A., Stavens, D., Thrun, S., & Bradski, G. R. (2006). Self-supervised monocular road detection in desert terrain. *Robotics Science and Systems*, Vol. 38 .

Desouza, G. & Kak, A. (2002). Vision for mobile robot navigation: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24 (Num. 2), (pp. 237–267).

Dickmanns, E. (2002). The development of machine vision for road vehicles in the last decade. In *Proceedings of th IEEE Intelligent Vehicle Symposium (Vol. 1),* (pp. 268–281).

Dickmanns, E. & Zapp, A. (1986). A curvature-based scheme for improving road vehicle guidance by computer vision. In *Proceedings of International Society for optics and photonics (Vol. 727),* (pp. 161–168).

Dickmanns, E. D. & Zapp, A. (1988). Autonomous high speed road vehicle guidance by computer vision. In *Proceedings of the International Federation of Automatic Control World Congress (Vol. 1),*.

Ekinci, M. & Thomas, B. (1996). Road junction recognition and turn-offs for autonomous road vehicle navigation. In *Proceedings of the International Conference on Pattern Recognition (Vol. 3),* (pp. 318–322).

Espasa-Calpe (1992). *Diccionario Enciclopédico Espasa 1 (7ma. Edición).* Espasa-Calpe S.A., Madrid.

Forsyth, D. A. & Ponce, J. (2002). *Computer Vision: A modern Approach.* Prentice-Hall.

Franks, U., Loose, H., & Knoppel, C. (2007). Lane recognition on country roads. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 99–104).

Frintrop, S., Rome, E., & Christensen, H. I. (2010). Computational visual attention systems and their cognitive foundations: A survey. *ACM Trans. Appl. Percept.*, Vol. 7 (Num. 1), (pp. 6:1–6:39).

Funke, J., Theodosis, P., Hindiyeh, R., Stanek, G., Kritatakirana, K., Gerdes, C., Langer, D., Hernandez, M., Muller-Bessler, B., & Huhnke, B. (2012). Up to the limits: Autonomous Audi TTS. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 541–547).

Gat, I., Benady, M., & Shashua, A. (2005). A monocular vision advance warning system for the automotive aftermarket. In *SAE World Congress & Exhibition (Vol. 2005),*.

Gerlach, C., Aaside, C., Humphreys, G., Gade, A., Paulson, O., & Law, I. (2002). Brain activity related to integrative processes in visual object recognition: bottom-up integration and the modulatory influence of stored knowledge. *Neuropsychologia*, Vol. 40 (Num. 8), (pp. 1254–1267).

Gerónimo, D., López, A., Sappa, A., & Graf, T. (2010). Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32 (Num. 7), (pp. 1239–1258).

Gevers, T., Smeulders, W., et al. (1999). Color based object recognition. *Pattern recognition*, Vol. 32 (Num. 3), (pp. 453–464).

Ghurchian, R. & Hashino, S. (2005). Shadow compensation in color images for unstructured road segmentation. In *Proceedings of the IAPR Conference on Machine Vision Applications,* (pp. 598–601).

Gilbert, C. D. & Sigman, M. (2007). Brain states: Topdown influences in sensory processing. *Neuron*, Vol. 54 , (pp. 677–696).

Gonzalez, R. C. & Woods, R. E. (2001). *Digital Image Processing.* John Wiley & Sons, Inc.

Gregory, R. (1965). Seeing in depth. *Nature*, Vol. 207 (Num. 4992), (pp. 116–117).

Guizzo, E. (2013). Toyota's semi-autonomous car will keep you safe. Publicado 8 Ene. 2013, en IEEE Spectrum Automaton Blog.
Disponible en http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/toyota-semi-autonomous-lexus-car-will-keep-you-safe/?utm_source=roboticsnews&utm_medium=email&utm_campaign=012213.

Guo, C. & Mita, S. (2009). Drivable road region detection based on homography estimation with road appearance and driving state models. In *Proceedings of the International Conference on Autonomous Robots and Agents,* (pp. 204–209).

Hautière, N., Tarel, J. P., & Aubert, D. (2010). Mitigation of visibility loss for advanced camera-based driver assistance. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11 (Num. 2), (pp. 474–484).

Hayhoe, M. & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, Vol. 9 (Num. 4), (pp. 188–194).

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, Vol. 3 (Num. 1).

He, Y., Wang, H., & Zhang, B. (2004). Color-based road detection in urban traffic scenes. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5 (Num. 4), (pp. 309–318).

Heimes, F. & Nagel, H.-H. (1998). Real-time tracking of intersections in image sequences of a moving camera. *Engineering Applications of Artificial Intelligence*, Vol. 11 (Num. 2), (pp. 215–227).

Henson, D. B. (1993). *Visual fields.* Oxford: Oxford University Press.

Hoffman, J. E. (1998). Visual attention and eye movements. *Attention*, (pp. 119–153).

Hoffman, J. E. & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, Vol. 57 , (pp. 787–795).

Howard, I. P. (2012). *Perceiving in Depth, Volume 1: Basic Mechanisms*, volume 29. Oxford University Press, USA.

Hummel, B., Kammel, S., Dang, T., Duchow, C., & Stiller, C. (2006). Vision-based path-planning in unstructured environments. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 176–181).

Hummel, B., Thiemann, W., & Lulcheva, I. (2008). Scene understanding of urban road intersections with description logic. In A. G. Cohn, D. C. Hogg, R. Möller, & B. Neumann (Eds.), *Logic and Probability for Scene Interpretation (Num. 08091),*, Dagstuhl Seminar Proceedings Dagstuhl, Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

Iagnemma, K. & Buehler, M. (2006a). (Eds.). Special issue on the DARPA Grand Challenge, Part 1 [Special issue]. *Journal of Field Robotics*, Vol. 23 (Num. 8), (pp. 461–652).

Iagnemma, K. & Buehler, M. (2006b). (Eds.). Special issue on the DARPA Grand Challenge, Part 2 [Special issue]. *Journal of Field Robotics*, Vol. 23 , (pp. 655–835).

Itti, L. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, Vol. 40 , (pp. 1489–1506).

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 (Num. 11), (pp. 1254–1259).

Jochem, T., Pomerleau, D., & Thorpe, C. (1995). Vision-based neural network road and intersection detection and traversal. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems 'Human Robot Interaction and Cooperative Robots' (Vol. 3),* (pp. 344–349).

Jochem, T., Pomerleau, D., & Thorpe, C. (1996). Vision based intersection navigation. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 391–396).

Kang, D. J., Choi, J. W., & Kweon, I. S. (1996). Finding and tracking road lanes using "line-snakes". In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 189–194).

Kong, H., Audibert, J.-Y., & Ponce, J. (2010). General road detection from a single image. *IEEE Transactions on Image Processing*, Vol. 19 (Num. 8), (pp. 2211–2220).

Land, M. & Lee, D. (1994). Where we look when we steer. *Nature*, Vol. 369 , (pp. 742–744).

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, Vol. 28 (Num. 11), (pp. 1131–1328).

Land, M. F. (1997). The knowledge base of oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences.*, Vol. 352 (Num. 1358), (pp. 1231–1239).

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18 (Num. 10), (pp. 959–971).

Lehtonen, E., Lappi, O., & Summala, H. (2012). Anticipatory eye movements when approaching a curve on a rural road depend on working memory load. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 15 (Num. 3), (pp. 369–377).

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, Vol. 7 (Num. 6), (pp. 651–657).

Lombardi, P., Zanin, M., & Messelodi, S. (2005). Switching models for vision-based on-board road detection. In *Proceedings of the IEEE Intelligent Transportation Systems,* (pp. 67–72).

Loose, H., Franke, U., & Stiller, C. (2009). Kalman particle filter for lane recognition on rural roads. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 60–65).

Lucchese, L. & Mitra, S. K. (2001). Color image segmentation: A state-of-art survey. In *Proceedings of the Indian National Science Academy (Vol. 67-A),* (pp. 207–221). New Delhi,India.

Luettel, T., Himmelsbach, M., & Wuensche, H.-J. (2012). Autonomous ground vehicles - concepts and a path to the future -. *Proceedings of the IEEE*, Vol. 100 (Special Centennial Issue), (pp. 1831–1839).

Lutzeler, M. & Dickmanns, E. (2000). EMS-vision: recognition of intersections on unmarked road networks. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 302–307).

MacCormick, J. & Isard, M. (2000). Partitioned sampling, articulated objects, and interface-quality hand tracking. In D. Vernon (Ed.), *Computer Vision*, Lecture Notes in Computer Science (Vol. 1843) (pp. 3–19). Springer Berlin Heidelberg.

Manz, M., Himmelsbach, M., Luettel, T., & Wuensche, H. (2011). Detection and tracking of road networks in rural terrain by fusing vision and lidar. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 4562–4568).

Manz, M., von Hundelshausen, F., & Wuensche, H.-J. (2010). A hybrid estimation approach for autonomous dirt road following using multiple clothoid segments. In *Proceedings of the IEEE International Conference on Robotics and Automation,* (pp. 2410–2415).

Markoff, J. (2010). Google cars drive themselves, in traffic. *The New York Times*, Vol. 10 (Oct. 2010).

Markoff, J. (2011). Google lobbies nevada to allow self-driving cars. *The New York Times*, Vol. 10 (May 2011).

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, Vol. 58 , (pp. 25–45).

McCall, J. & Trivedi, M. (2006). Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7 (Num. 1), (pp. 20–37).

Miksik, O., Petyovsky, P., Zalud, L., & Jura, P. (2011). Robust detection of shady and highlighted roads for monocular camera based navigation of ugv. In *Proceedings of the IEEE International Conference on Robotics and Automation,* (pp. 64–71).

Montemerlo, M., Becker, J., Bhat, S., Dahlkamp, H., Dolgov, D., Ettinger, S., Haehnel, D., Hilden, T., Hoffmann, G., Huhnke, B., Johnston, D., Klumpp, S., Langer, D., Levandowski, A., Levinson, J., Marcil, J., Orenstein, D., Paefgen, J., Penny, I., Petrovskaya, A., Pflueger, M., Stanek, G., Stavens, D., Vogt, A., & Thrun, S. (2008). Junior: The stanford entry in the urban challenge. *Journal of Field Robotics*, Vol. 25 (Num. 9), (pp. 569–597).

Moreyra, M. L. & Masson, F. R. (2010). Combinación de rango y visión monocular para la identificación de zonas transitables. In *Actas de las VI Jornadas Argentinas de Robótica (JAR)* (Disponible en formato digital).

Moreyra, M. L. & Masson, F. R. (2011). Detección del terreno transitable con imágenes monoculares. In *Actas de la XIV Reunión de Trabajo en Procesamiento de la Información y Control (RPIC)* (Disponible en formato digital).

Moreyra, M. L. & Masson, F. R. (2012). Visual attention dataset: Non-linear road topologies perception in unstructured scenes (available in http://lcr.uns.edu.ar/pictv/eyetrackerdata.html).

Mourant, R. R. & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 12 (Num. 1), (pp. 81–87).

Mourant, R. R. & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 14(Num. 4), (pp. 325–335).

Nefian, A. & Bradski, G. (2006). Detection of drivable corridors for off-road autonomous navigation. In *Proceedings of the IEEE International Conference on Image Processing,* (pp. 3025–3028).

Niebur, E. & Koch, C. (1998). *Computational Architectures for Attention.* R. Parasuraman, ed. *The Attentive Brain*, pp. 163 - 186. Cambridge, Mass.: MIT Press.

Peynot, T., Underwood, J., & Scheding, S. (2009). Towards reliable perception for unmanned ground vehicles in challenging conditions. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 1170–1176).

Pomerleau, D. (1995). RALPH: rapidly adapting lateral position handler. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 506–511).

Pomerleau, D. A. (1992). *Neural network perception for mobile robot guidance.* Technical report, DTIC Document.

Rasmussen, C. (2002). Combining laser range, color, and texture cues for autonomous road following. In *Proceedings of the IEEE International Conference on Robotics and Automation (Vol. 4),* (pp. 4320–4325).

Rasmussen, C. (2003). Road shape classification for detecting and negotiating intersections. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 422–427).

Rasmussen, C. (2004a). Grouping dominant orientations for ill-structured road following. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol. 1),* (pp. 470–477).

Rasmussen, C. (2004b). Texture-based vanishing point voting for road shape estimation. In *Proceedings of the British machine vision conference,* (pp. 470–477).

Rasmussen, C. (2006). A hybrid vision + ladar rural road follower. In *Proceedings of the IEEE International Conference on Robotics and Automation,* (pp. 156–161).

Rasmussen, C., Lu, Y., & Kocamaz, M. (2009). Appearance contrast for fast, robust trail-following. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 3505–3512).

Rasmussen, C. & Scott, D. (2008). Shape-guided superpixel grouping for trail detection and tracking. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 4092–4097).

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, Vol. 124 (Num. 3), (pp. 372–422).

Robinson, T., Chan, E., & Coelingh, E. (2010). Operating platoons on public motorways: an introduction to the sartre platooning programme. In *Proceedings of the 17th ITS World Congress (Busan)*.

Roorda, A. (2002). *Human Visual System-Image Formation. Encyclopedia of Imaging Science and Technology*. John Wiley & Sons, Inc.

Shinar, D., Mcdowell, E. D., & Rockwell, T. H. (1977). Eye movements in curve negotiation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 19 (Num. 1), (pp. 63–71).

Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., & Gilbert, C. D. (2005). Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron*, Vol. 46 (Num. 5), (pp. 823–835).

Sotelo, M. A., Rodriguez, F. J., Magdalena, L., Bergasa, L. M., & Boquete, L. (2004). A color vision-based lane tracking system for autonomous driving on unmarked roads. *Autonomous Robots*, Vol. 16 (Num. 1), (pp. 95–116).

Southall, B. & Taylor, C. (2001). Stochastic road shape estimation. In *Proceedings of the IEEE International Conference on Computer Vision (Vol. 1),* (pp. 205–212).

Szeliski, R. (2010). *Computer Vision: Algorithms and Applications*. Springer.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research*, Vol. 45 (Num. 5), (pp. 643–659).

Thorpe, C., Hebert, M., Kanade, T., & Shafer, S. (1988). Vision and navigation for the carnegie-mellon navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10 (Num. 3), (pp. 362–373).

Tue-Cuong, D.-S., Dong, G., Hwang, Y. C., & Heng, O. S. (2008). Robust extraction of shady roads for vision-based ugv navigation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 3140–3145).

Turk, M., Marietta, M., Morgenthaler, D., Gremban, K., & Marra, M. (1988). Vits-a vision system for autonomous land vehicle navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 10 (Num. 3), (pp. 342–361).

Underwood, G., Chapman, P., Crundall, D., Cooper, S., & Wallen, R. (1999). The visual control of steering and driving: Where do we look when negotiating curves? *Vision in Vehicles*, Vol. VII, (pp. 245–252).

Urmson, C., Anhalt, J., Bagnell, D., Baker, C., Bittner, R., Clark, M. N., Dolan, J., Duggins, D., Galatali, T., Geyer, C., Gittleman, M., Harbaugh, S., Hebert, M., Howard, T. M., Kolski, S., Kelly, A., Likhachev, M., McNaughton, M., Miller, N., Peterson, K., Pilnick, B., Rajkumar, R., Rybski, P., Salesky, B., Seo, Y.-W., Singh, S., Snider, J., Stentz, A., Whittaker, W. R., Wolkowicki, Z., Ziglar, J., Bae, H., Brown, T., Demitrish, D., Litkouhi, B., Nickolaou, J., Sadekar, V., Zhang, W., Struble, J., Taylor, M., Darms, M., & Ferguson, D. (2008). Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics*, Vol. 25 (Num. 8), (pp. 425–466).

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, Vol. 19 , (pp. 1395–1407).

Wang, Y., Bai, L., & Fairhurst, M. (2008). Robust road modeling and tracking using condensation. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9 (Num. 4), (pp. 570–579).

Wang, Y., Teoh, E. K., & Shen, D. (2004). Lane detection and tracking using B-Snake. *Image and Vision Computing*, Vol. 22 (Num. 4), (pp. 269–280).

Westheimer, G. (1987). *Visual Acuity.* Moses, R. A. and Hart, W. M., ed. *Adler's Physiology of the eye, Clinical Application* (cap. 17). The C.V. Mosby Company.

Wille, J., Saust, F., & Maurer, M. (2010). Stadtpilot: Driving autonomously on Braunschweig's inner ring road. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 506–511).

Wu, Q., Zhang, W., Chen, T., & Kumar, B. (2010). Prior-based vanishing point estimation through global perspective structure matching. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing,* (pp. 2110–2113).

Yarbus, A. (1967). *Eye movements and vision.* Plenum, New York.

Zhang, G., Zheng, N., Cui, C., Yan, Y., & Yuan, Z. (2009). An efficient road detection method in noisy urban environment. In *Proceedings of IEEE Intelligent Vehicles Symposium* (pp. 556–561).

Zhang, J. & Nagel, H.-H. (1994). Texture-based segmentation of road images. In *Proceedings of the IEEE Intelligent Vehicles Symposium,* (pp. 260–265).

Zhang, Z. & Andreou, A. (2008). Human identification experiments using acoustic micro-doppler signatures. In *Proceedings of the Argentine School of Micro-Nanoelectronics, Technology and Applications,* (pp. 81–86).

Zhang, Z., Pouliquen, P., Waxman, A., & Andreou, A. (2007). Acoustic micro-doppler gait signatures of humans and animals. In *Proceedings of the Annual Conference on Information Sciences and Systems,* (pp. 627–630).