

Reconocimiento de objetos en imágenes Su aplicación a la identificación de especies de serpientes

AUTOR
Luciano La Frazia

DIRECTORA
Dra. Laura Lanzarini

CO-DIRECTOR
Lic. César Estrebou

ASESOR PROFESIONAL
Lic. Luis Giambelluca (CEPAVE)



Tesina de Licenciatura en Informática

FACULTAD DE INFORMATICA
UNIVERSIDAD NACIONAL DE LA PLATA

26 de septiembre de 2018

Índice general

1. Introducción	5
1.1. Introducción al procesamiento digital de imágenes	5
1.2. Organización de la tesina	9
2. Serpientes	11
2.1. Serpientes de la región	11
2.2. Características de las serpientes	12
2.3. Conclusiones	19
3. SIFT (Scale Invariant Feature Transform)	21
3.1. Introducción	21
3.2. Descripción general del algoritmo	22
3.2.1. Detección de extremos en cuanto a espacio-escala	23
3.2.2. Detección de extremos locales	26
3.3. Extracción de puntos claves	26
3.3.1. Eliminación de bordes	27
3.4. Asignación de la orientación de los puntos claves	28
3.5. Creación de descriptores	29
3.6. Otros métodos similares	32
3.7. Conclusiones	33
4. Definición del modelo de datos	34
4.1. Introducción	34
4.2. Descripción del modelo	35
4.3. Construcción de la base de datos	38
4.4. Selección de imágenes	38
4.5. Generación de descriptores	41
4.5.1. Análisis de descriptores	44
4.6. Validación del modelo	45
4.7. Tipos de validación cruzada	46

4.7.1. Validación cruzada dejando uno fuera	46
4.7.2. Validación cruzada de K iteraciones	47
4.8. Resultados obtenidos	49
4.9. Conclusiones	54
5. Reconocimiento de Especies	56
5.1. Introducción	56
5.2. Método Propuesto	56
5.3. Resultados Obtenidos	57
5.4. Conclusiones	61
6. Conclusiones Generales	62
6.1. Conclusiones	62
6.2. Trabajos Futuros	63
Bibliografía	69

Agradecimientos

En primer lugar, quisiera agradecer a Laura y César por permitirme trabajar con ellos, por su paciencia y su cariño a lo largo de todo el tiempo que tomó este trabajo. Gracias a su guía y sus críticas pude llegar a esta instancia. Los errores, por supuesto, me pertenecen.

A Luis, por su acompañamiento y su disposición para atender mis dudas.

A Diego, por su amistad y sus consejos a lo largo de toda la carrera.

A mi familia, en especial a mis padres por apoyar mis decisiones de forma incondicional.

A mis amigos y amigas, por estar conmigo siempre y acompañarme en todo momento.

A Corita, por aparecer en el momento justo de mi vida.

Y por último, a Cate, la persona que más quiero en el mundo.

—

Resumen

En el área de procesamiento de imágenes, la detección de puntos característicos para la posterior clasificación y/o reconocimiento de objetos, es uno de los casos de estudio vigentes en la actualidad. Si bien no existe un algoritmo genérico que funcione de manera óptima en todos los casos de estudio, el objetivo general de este trabajo es estudiar y analizar distintos mecanismos para el reconocimiento de objetos presentes en una imagen.

El objetivo particular es utilizar el resultado de esta investigación para desarrollar e implementar un algoritmo capaz de identificar la especie a la que pertenece una serpiente a partir de la imagen de un ejemplar. Para llevar adelante esta tarea se cuenta con asesoramiento profesional especializado.

Este problema posee distintos niveles de complejidad que dependen de la cantidad de especies a reconocer y de la facilidad con la que las partes distintivas del ejemplar puedan ser ubicadas dentro de la imagen.

En esta tesina, se trabajará con las especies de serpientes de la provincia de Buenos Aires y en particular con las del partido de La Plata. Esto, si bien reduce considerablemente el número de opciones, no invalida la importancia y aplicabilidad del trabajo a realizar.

Este trabajo de investigación cubre varios aspectos de la solución de este problema:

- En primer lugar se presenta una introducción sobre el tema a nivel teórico, donde se detallan características de las distintas especies, además de la importancia de su conservación para el equilibrio del medio ambiente.
- Luego se describe la construcción de la base de datos contra la que se compararán los ejemplares desconocidos, la selección y segmentación de imágenes y varias estadísticas que clarifican su armado.
- Después se describe el método utilizado para generar los vectores de características que representan a una especie en particular, donde además, se presentan ejemplos de aquellas partes distintivas de cada especie.
- Por último, se describe y detalla el funcionamiento de la aplicación, junto con su respuesta

indicando la especie a la que pertenece la imagen de entrada, su grado de peligrosidad y confiabilidad de respuesta suministrada.

Palabras Claves: Procesamiento de Imágenes, Visión Artificial, Reconocimiento de objetos, Descriptores de puntos de interés. Ofidios. Herpetología.

Capítulo 1

Introducción

1.1. Introducción al procesamiento digital de imágenes

Una imagen puede ser definida matemáticamente como una función bidimensional $f(x, y)$, donde x e y representan las coordenadas espaciales y el valor de la función en un punto cualquiera (x, y) es la intensidad o nivel de gris de la imagen en esa coordenada [Gonzalez and Woods, 2006].

El término niveles de grises se utiliza para referir la intensidad monocromática de las imágenes. Las imágenes a color son formadas por una combinación individual de imágenes 2-D, por ejemplo en el sistema de color RGB un color consiste en tres componentes individuales de una imagen que son rojo, verde y azul respectivamente.

La visión artificial (o visión por computador) es una disciplina compleja que involucra otras ciencias e incluye estudios de física, matemáticas, ingeniería eléctrica, computación, entre otras. El continuo desarrollo de nuevos algoritmos, funciones y aplicaciones, hacen de esta disciplina y del procesamiento digital de imágenes un área en constante evolución. La visión artificial intenta emular la capacidad que tienen algunos seres vivos para ver una escena y entenderla. Los procesos del procesamiento de imágenes pueden ser subdivididos en cinco áreas principales [Gonzalez and Woods, 2006], como se detalla en la figura 1.1 y están agrupados de acuerdo a la complejidad y delicadeza que lleva su implementación. Estas son:

- **Captura y adquisición:** es el proceso a través del cual se obtiene una imagen digital utilizando un dispositivo de captura como una cámara digital, cámara de video, escáner, satélite, etc.
- **Preprocesamiento:** es el proceso que busca mejorar la imagen de forma que aumenten las posibilidades de éxito de los procesos posteriores. El preprocesamiento incluye técnicas tales como la reducción del ruido, mejora del brillo, contraste, etc.

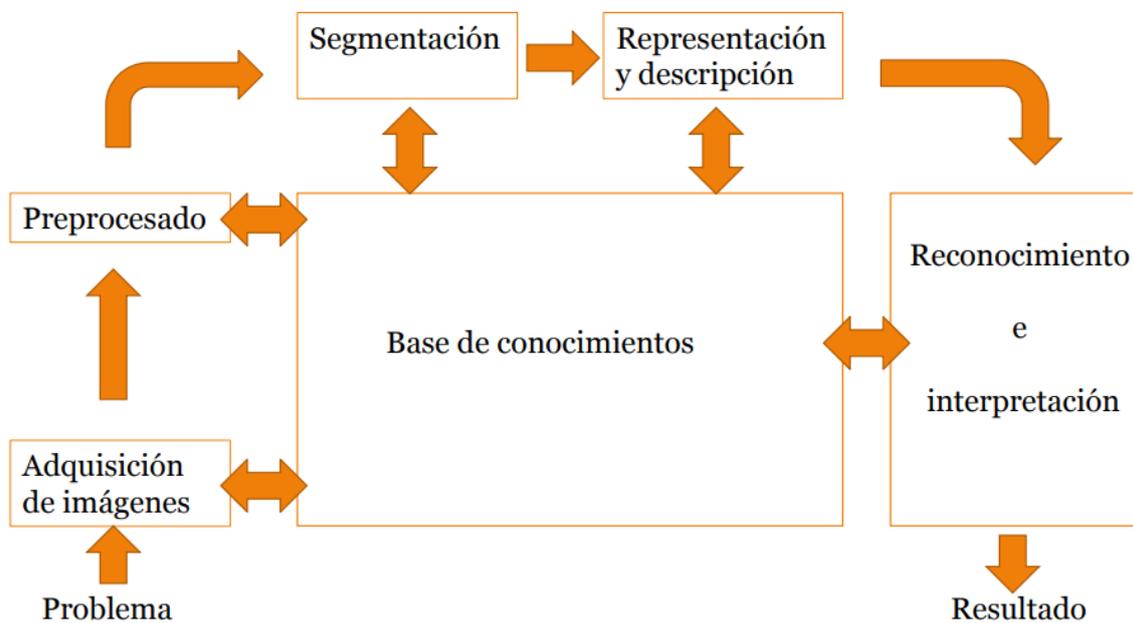


Figura 1.1: Detalle de los procesos que componen el procesamiento de imágenes

- **Segmentación:** es el proceso que divide una imagen en objetos que sean de interés de estudio. Es uno de los pasos más críticos, ya que de esta etapa dependerá que la probabilidad de éxito en las etapas finales sea mayor. La figura 1.2 ejemplifica los pasos a seguir para identificar la zona donde se encuentra el objeto a reconocer.
- **Descripción:** es el proceso que obtiene características convenientes para diferenciar un tipo de objeto de otro, es decir, extraer rasgos con alguna información cuantitativa de interés o que sean fundamentales para diferenciar una clase de objetos de otra. Este tema se analizará con mayor detalle en el capítulo 3.
- **Reconocimiento e interpretación:** es el proceso que identifica los objetos caracterizados en el paso anterior. Para ello requiere de información previa referida a los distintos objetos a reconocer. Esta información está almacenada en la base de conocimiento.

No todas las aplicaciones requieren de todos los procesos y depende de la complejidad del problema que se va a resolver. Los resultados obtenidos en este tipo de aplicaciones dependen de la calidad de la imagen original, por lo que se deben tomar todas las precauciones necesarias para tener una iluminación adecuada y uniforme en el momento de su adquisición.

Se pueden clasificar a los procesos de captura y adquisición y preprocesamiento, como funciones de visión de bajo nivel. La segmentación, descripción y reconocimiento como funciones de visión de nivel medio y por último, la interpretación como función de visión de alto nivel.

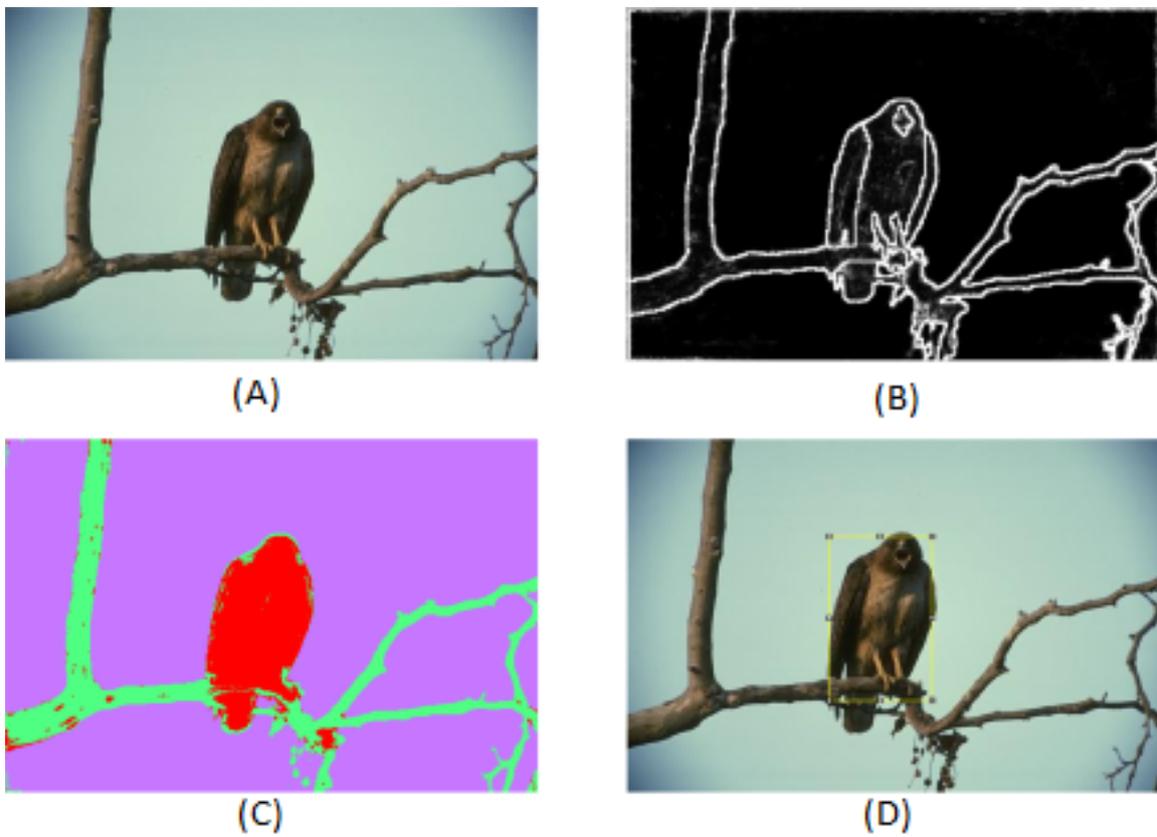


Figura 1.2: (A) Imagen original (B) Detección de bordes (C) Segmentación semántica (D) Detección del objeto de interés

Tabla 1.1: Entradas y salidas de los distintos niveles en los que pueden agruparse los métodos del procesamiento digital de imágenes

Nivel	Procesos	Entrada	Salida
Bajo	Reducción de ruido Realce de contraste Realce de características	Imagen	Imagen
Medio	Segmentación (regiones, objetos) Descripción de objetos Clasificación o Reconocimiento	Imagen	Atributos de objetos: bordes, contornos, áreas, descripción de objetos individuales
Alto	Interpretación Análisis de la imagen Funciones cognitivas	Objetos encontrados	Análisis de la imagen (información, clasificación de los objetos.)

A su vez, se pueden agrupar los métodos del procesamiento de imágenes en tres niveles, mencionados en la tabla 1.1:

Como se mencionó, no todas las aplicaciones relacionadas con el procesamiento digital de imágenes requieren de todos los procesos descritos anteriormente. Por lo general, mientras la complejidad del problema a resolver crece, el número de procesos requeridos también lo hace. [Santillán, 2008]

El reconocimiento de objetos en una imagen es una de las tareas centrales de la visión por computadora y tiene por finalidad encontrar e identificar los elementos presentes en la misma. Existen múltiples áreas donde se aplica el procesamiento digital de imágenes para la resolución de distintos problemas, entre ellas se pueden mencionar:

- Imágenes médicas [Zhang et al., 2013].
- Observaciones de imágenes satelitales [Gill and Sharma, 2017].
- Identificación de billetes [Delgado, 2015].
- Reconocimiento de caracteres o símbolos (patentes de vehículos) [Patel et al., 2013].
- Estudios de agricultura para detectar plagas y deficiencias de nutrientes [Janwale, 2015].
- Reconocimiento de rostros [Lanzarini et al., 2013, K.V.Bhuvaneshwari, 2017].
- Tracking de videos [Cohen and Medioni, 1998].
- Reconocimiento de gestos [Lanzarini et al., 2016], etcétera.

Los seres humanos son capaces de reconocer objetos en imágenes con suma facilidad incluso aún cuando la imagen se encuentra distorsionada o es captada desde diferentes puntos de vista.

Sin embargo, esta tarea es todavía un desafío para los sistemas de visión por computadoras. Existen en la literatura numerosas soluciones que han buscado dar respuesta a este problema sin lograr aún resolverlo por completo.

Si bien los métodos a utilizar son diversos, las dos grandes tareas a resolver son: la segmentación del objeto y posteriormente, su clasificación y/o reconocimiento. La primera tiene que ver con ubicar la zona de interés separando el objeto del fondo de la imagen, mientras que la segunda requiere disponer de información de un conjunto de objetos conocidos contra los que se comparará el resultado de la segmentación. Habitualmente el proceso de comparación no se realiza pixel a pixel sino que se caracterizan los objetos y se utiliza algún criterio de similitud. De esta forma, se evitan los problemas relacionados con la escala, la rotación y el punto de vista, entre otros.

Teniendo en cuenta que no existe un método genérico el cual funcione para cualquier tipo de problema, la elección de las tecnologías para describir y caracterizar imágenes o videos continúa siendo una incógnita, ya que existen algoritmos que se ajustan mejor a determinados ambientes y contextos.

1.2. Organización de la tesina

A continuación se detalla como se encuentra organizado el resto del documento:

En el capítulo 2 se describe una introducción sobre las serpientes que habitan la zona, los distintos estados y contextos en los cuales pueden ser halladas, algunas características que resultan interesantes y como esto impacta en esta tesina. El capítulo concluye remarcando la importancia de conservar a las especies y como este desarrollo puede ayudar a conseguirlo.

En el capítulo 3 se desarrollan distintas técnicas para la generación de descriptores utilizadas en esta tesina. Aquí, se explican con detalle los diferentes pasos que conlleva la representación de una imagen a un descriptor. Finalmente, se comentan de forma breve distintas variantes que existen en la literatura indicando donde tienen mejor desempeño.

En el capítulo 4 se define la base de datos que se implementó para la clasificación de los distintos ejemplares del partido de La Plata. Se resume la selección de imágenes justificando la elección de las mismas y múltiples pruebas que se realizaron para probar la robustez del modelo. Además se implementó un algoritmo para definir el umbral óptimo a utilizar y se presentaron distintas tasas de acierto obtenidas a partir de validaciones cruzadas.

El capítulo 5 detalla los experimentos llevados a cabo para mostrar la eficacia del modelo presentado. Las pruebas se realizaron con imágenes nuevas y se implementó una interfaz que facilita este proceso. Finalmente, la salida del aplicativo detalla información de la especie ganadora junto con un gráfico que informa la precisión de la respuesta.

Por último, el capítulo 6 presenta las conclusiones de la tesina junto con las posibles líneas de trabajos futuros.

Capítulo 2

Serpientes

2.1. Serpientes de la región

Los objetos de estudio en esta tesina son las distintas especies de serpientes de la provincia de Buenos Aires y en especial las correspondientes al partido de La Plata.

Las serpientes constituyen uno de los grupos más llamativos para el hombre, siendo para algunas culturas deidades y para otras sinónimo del mal. Es muy difícil encontrar personas a las que le sea indiferente. La mayoría las ama o las odia, cosa que con otros grupos (aves, tortugas, etc.) pueden tener una tercera postura, que es "me son indiferentes".

La presencia de veneno en algunas especies les ha dado una importancia indiscutida en cuanto al interés sanitario. Mucha gente considera que las serpientes deben ser matadas con celeridad ya que sólo pueden causar problemas o aprensión. Afortunadamente, en la actualidad se empezó a rever esa postura, dándole a las serpientes un rol mucho más racional, considerándolas como un grupo más junto con otras especies que conforman a un ecosistema. Esta última tendencia a dejar mitos y miedos a un lado y ver a las serpientes con otros ojos, permitió que las mismas tengan el lugar que se merecen, como parte importante en el equilibrio de las especies.

Analizando los pros y contras, puede decirse que dentro de las contras, algunas especies pueden causar un accidente que requiera internación y colocación de suero antiofídico y que en poquísimos casos en Argentina, llegan a la muerte. En la actualidad, no constituyen la principal causa de muerte. Esto se debe a que en la provincia de Buenos Aires, sólo hay dos especies con potencialidad en sus venenos para causar un accidente de esta gravedad, mientras que las 29 restantes no revisten gran interés sanitario; además existe un antiveneno efectivo para cada especie. Por otro lado, analizando los pros, debe remarcarse que ayudan a mantener controladas algunas poblaciones de roedores que sí pueden causar daños a la salud (como enfermedades zoonóticas) o daños a las cosechas, los cuales pueden ser cuantiosos. También hay que considerar que las serpientes son a su vez el sustento de algunos depredadores, los

cuales encuentran en ellas una fuente de alimentación.

La ofidiofauna de la provincia de Buenos Aires, comprende cuatro familias, dos de las cuales el común de la gente no las reconoce como serpientes, ya que su aspecto de cordel sumado a su pequeño tamaño, forma de vida hipógea y su baja frecuencia de encuentro las coloca dentro de los ofidios ignorados por la mayoría de las personas y en caso de encontrarlas pueden confundirlas con lombrices o cualquier otro animal. Las otras dos familias son bien reconocibles como serpientes, siendo una la más diversa, la cual incluye a las culebras (Familia Dipsadidae), mientras que la otra incluye a las comúnmente llamadas Yarará (Familia Viperidae). Dentro de las culebras, se pueden encontrar ejemplares con distintos diseños de dibujos y coloración o totalmente lisas, con pupilas circulares o elípticas, con cabezas apenas distinguibles de su cuerpo hasta llegar a las que pueden mostrar una cabeza notoria respecto a su cuerpo. También la sección del cuerpo puede ser muy variable, con especies más chatas o más circulares y también la cola puede ser de distintas longitudes. Esta diferencia de formas y coloración se debe a sus distintos hábitat y hábitos, donde pueden encontrárselas en lugares secos, húmedos y hasta acuáticos, como así también fosoriales. Pero a igual hábitat o hábitos distintas especies pueden tornarse similares en su forma y color, como sucede con las falsas yarará o falsas corales, respecto a las verdaderas. En la zona de estudio, se eligieron las especies más frecuentes, las cuales son las que se grafican en la figura 2.1, y pertenecen a las dos familias más reconocidas antes mencionadas. Dentro de estas especies se encuentran aquellas que no poseen veneno y las venenosas, las cuales no causan accidentes mortales en el ser humano. En cuanto a la familia Viperidae, la misma tiene, en la provincia, dos especies, las cuales son fácilmente diferenciables, ya que una tiene hocico respingado y cuerpo manchado Yarará Ñata (*Bothrops ammodytoides*), mientras la otra especie es la Yarará grande o de la Cruz (*Bothrops alternatus*) que posee inconfundibles dibujos laterales, tal como se observa en la figura 2.1. En la zona a tratar, solo aparece la Yarará Grande. Esta familia tiene los representantes que pueden causar accidentes graves, los cuales necesitan la aplicación de antisuecos.

2.2. Características de las serpientes

Las serpientes son un grupo muy homogéneo, ya que todas tienen un cuerpo muy largo, cubierto de escamas, sin patas (salvo los rudimentos que tienen algunos grupos), lengua bífida, son todas carnívoras, tienen los párpados fijos y transparentes. Hay otras características que no dejan de ser interesantes, algunas compartidas con los demás reptiles [Giambelluca, 2015].

Comenzando por su posición sistemática, puede decirse que son **vertebrados**, es decir que tienen vértebras como las que se observan en la figura 2.2. Estas vértebras protegen la médula espinal al rodearla y son el sostén del cuerpo. Las costillas van a tener una función de almacén del cuerpo, protección de los órganos y también de locomoción, al ser la fijación de músculos



Figura 2.1: (A) *Bothrops Alternatus* (B) *Paraphimophis Rusticus* (C) *Philodryas Patagoniensis* (D) *Tomodon Ocellatus* (E) *Xenodon Dorbignyi*

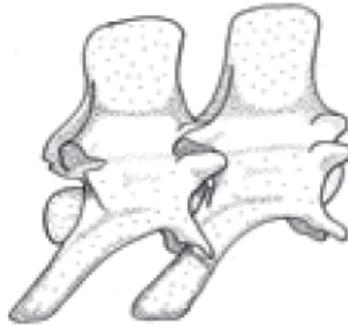


Figura 2.2: Vértebras de una serpiente. ([Giambelluca, 2015], página 12 - reimpresso con permiso)

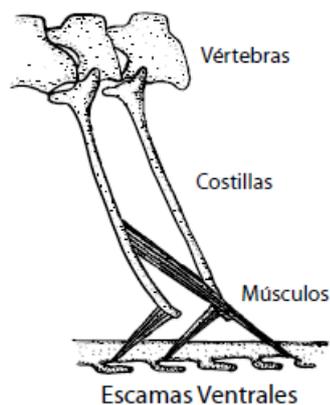


Figura 2.3: Escamas ventrales de una serpiente. ([Giambelluca, 2015], página 12 - reimpresso con permiso)

que se conectan a las escamas ventrales y que al contraerse en forma rítmica producen el desplazamiento del ofidio como si tuviera tantas patas como escamas ventrales, como se observa en la figura 2.3.

Son **Amniotas**, esta condición las separa de los peces y anfibios y la que comparten con las aves y mamíferos. El *amnio*, es una membrana que está dentro del huevo y envuelve al embrión, dándole un ambiente acuoso. Esta es la gran adaptación de algunos vertebrados para independizarse de los cuerpos de agua y poder conquistar los ambientes secos. El embrión también cuenta con una membrana llamada *alantoides*, la cual recibe los desechos metabólicos durante su vida en el huevo como se observa en la figura 2.4.

En el caso general de los ofidios, ponen huevos que tienen una cáscara apergaminada. Como se observa en las figuras 2.5 y en 2.6, generalmente son elípticos en su contorno que después de



Figura 2.4: Escamas ventrales de una serpiente. ([Giambelluca, 2015], página 13 - reimpresso con permiso)

puestos toman turgencia, es decir, se hinchan debido a la presión ejercida por los fluidos y por el contenido celular sobre las paredes de la célula, dejando su aspecto arrugado. Las puestas son generalmente en oquedades que pueden estar bajo tierra, en un árbol o en el interior de hormigueros. En este último caso reciben humedad, temperatura y cuidados de las hormigas podadoras, porque las mismas son cultivadoras de hongos y cuidan que sus hongueras no seas invadidas por hongos que no sean los que ellas cultivan (los cuales no atacan a los huevos). Pero no todas las serpientes tienen puestas, algunas retienen sus huevos y los viboreznos eclosionan en el momento de expulsarlos. Esta particularidad produce huevos cuya cáscara es muy débil y da la impresión, al nacer, de estar ante un animal vivíparo (envueltos en la placenta).

Son **Reptiles**. Puede decirse que los mismos, comparten características fisiológicas como el hecho de no producir temperatura corporal (término que se lo puede encontrar como poiquilotermo, heterotermo o de "sangre fría"). Es así que los reptiles regulan la temperatura mediante su comportamiento, asoleándose o colocándose a la sombra según la temperatura que han adquirido o tomando el calor por contacto con objetos calientes.

Esta dependencia de la temperatura ambiental trae como consecuencia que si las temperaturas se tornan muy frías (ejemplo otoño avanzado, invierno y principios de primavera) el reptil se retira a un refugio y entra en letargo. A este estado se lo llama hibernación. Por el contrario, en una temporada muy calurosa (verano muy tórrido) tiene el riesgo de entrar a una hipertermia, porque al no transpirar como los mamíferos no puede bajar su temperatura y el calor lo puede matar. A este estado se lo conoce como estivación. Hay reptiles que no hibernan ni estivan ya que el ambiente mantiene todo el año una temperatura acorde a sus necesidades.

Una particularidad de las serpientes es que el párpado es fijo y transparente, por esta razón es posible encontrarlo en la muda. La muda es el cambio completo de toda la piel del ofidio.



Figura 2.5: Textura apergaminada de un huevo de un ofidio.[Ark, 2016]



Figura 2.6: Huevos apergaminados. [Vig, 2013]



Figura 2.7: Víbora ratonera (*Pantherophis obsoletus*) a punto de mudar de piel, momento en que la escama ocular se ve opaca. [War, 2016]

Cuando está por suceder este cambio de piel, se junta un líquido entre la piel nueva y la vieja, que hace que el ejemplar tenga los colores más apagados y el ojo con un color celeste grisáceo, como se muestra en la figura 2.7. Finalmente llega el momento de mudar la piel y fregando la punta de la cabeza y luego mediante movimientos del cuerpo la piel se va dando vuelta como una media, hasta que queda libre de esa piel vieja y se ve al ofidio con colores mucho más brillantes, como en la figura 2.8.

Las **Serpientes** tienen muchas otras características, algunas compartidas con otros escamosos o con otros reptiles. Si se empieza por la cabeza, en ella pueden encontrarse los ojos, los cuales muchas veces tienen un rol muy importante para la captura de las presas (especies arborícolas, cazadoras de aves) o un rol secundario ya que la caza se basa en la detección de sustancias aromáticas (olfato).

Si se hace foco en la parte ósea de la cabeza se dice que los ofidios tienen los huesos que rodean al cerebro bien firmes y todos los demás muy móviles, salvo aquellos grupos hipogeos (que se desarrollan bajo el suelo), que tienen mandíbulas más rígidas que los demás. Pero la mayoría de los ofidios tienen los huesos con uniones muy laxas, lo cual les permite hacer ingestas de grandes presas en comparación al tamaño de la boca.

Las mandíbulas consisten en varios huesos (si bien hay una fuerte reducción en las serpientes respecto a los lagartos) y la derecha y la izquierda no están soldadas en el mentón, como los mamíferos o las aves. Así ambas mandíbulas se pueden separar al momento de la ingesta. Por otra parte, la unión de ella al cráneo se hace por medio de dos huesos, el cuadrado y el escamoso, produciendo una apertura bucal desmesurada.



Figura 2.8: Momento exacto en que el ofidio cambia su piel. Se puede apreciar el brillo en su color. [Vel, 2008]

Los dientes tienen mucha importancia en la dieta, los cuales se encuentran en los maxilares, en el premaxilar (en algunos), en el dentario de la mandíbula y en los huesos palatinos y pterigoides (o sea en el paladar). Todos son curvados hacia atrás, como se muestra en la figura 2.9, para que la presa no retroceda en su viaje al tracto digestivo.

Las serpientes que se consideran venenosas son las que tienen dientes modificados con canal o perforados y una glándula de veneno asociada (opistoglifas, proteroglifas y solenoglifas). La composición de los venenos varía entre especies, pero también en el mismo individuo a lo largo de su vida. Esto muchas veces se debe a que un juvenil no come lo mismo que un adulto (por ejemplo: la dieta del juvenil se basa en anfibios y de roedores cuando es adulto).

El caso de las proteroglifas (Corales), que en la provincia de Buenos Aires se las ha mencionado y posiblemente podrían llegar con inundaciones del Paraná, o estar en el noreste de la provincia, son poseedoras de venenos que pueden ser mortales para el hombre.

El caso de las solenoglifas, son las que más causan accidentes ofídicos, los cuales pueden llegar a la muerte o dejar terribles secuelas. Hay que mencionar que este aparato picador formidable que han desarrollado los Vipéridos (Yararaes y Cascabel en Argentina) son muy eficientes para inocular veneno, sumado a que muchas veces el tamaño de los ejemplares les permite producir mucho veneno, tienen un temperamento irritable y se defienden atacando. Igualmente, salvo casos muy puntuales, al igual que las corales, con la aplicación de antiveneno la muerte puede evitarse en un gran número de casos.

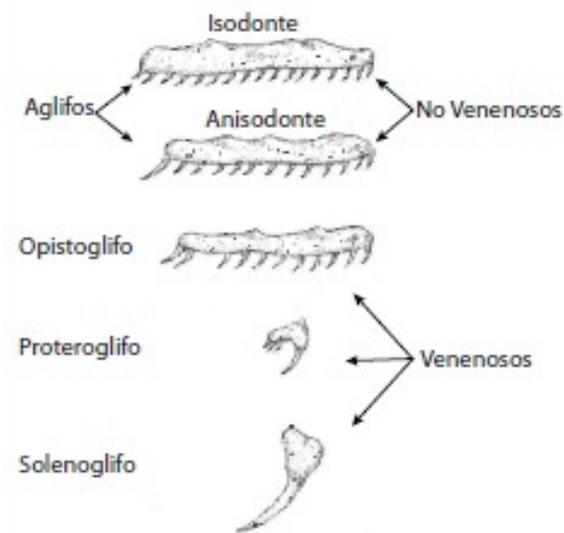


Figura 2.9: Morfología de los dientes de distintas especies de serpientes. ([Giambelluca, 2015], página 13 - reimpresso con permiso)

2.3. Conclusiones

En este capítulo se ha brindado una introducción general de las serpientes, tratando temas como su comportamiento, morfología y otras características generales que resultan llamativas. Además, se ha descrito el problema que actualmente enfrentan las distintas especies de serpientes en la provincia de Bs.As.

El desarrollo de esta tesina busca combinar ambas temáticas a través del desarrollo de una aplicación que será de utilidad para el público en general, ya que ingresando una imagen de determinado ejemplar se podrá consultar el grado de peligrosidad del animal así como otros datos de sus hábitos y distribución.

Preservarlas es de interés público ya que constituyen un fuente de control natural de los roedores. Estos últimos se encuentran cada vez más presentes en distintos ámbitos constituyendo una fuente importante de transmisión de enfermedades.

El primer tema que debe enfrentarse, para el desarrollo de la aplicación mencionada, es la construcción de la base de datos, que será la fuente de consulta contra la que se compararán los diferentes ejemplares a reconocer. Su conformación es crucial para el buen desempeño de la aplicación. Dicha base de datos se encuentra formada por imágenes de sectores representativos de los cuerpos de las serpientes. No se trata de imágenes del animal completo sino sólo los sectores que se consideran necesarios para efectuar el reconocimiento. Para la selección de estos sectores, se debe tener en cuenta el brillo, contraste, nitidez, ángulo y definición de las mismas entre tantas otras características. En el capítulo siguiente se presentan distintos métodos que

permiten representar imágenes en forma de descriptores. Éstos se utilizarán para la construcción de la BDD y el posterior reconocimiento.

Capítulo 3

SIFT (Scale Invariant Feature Transform)

3.1. Introducción

La detección automática de objetos o patrones en imágenes tiene aplicaciones en temáticas tan diversas como la industria, seguridad, medicina, robótica y demás.

El reconocimiento de objetos o patrones es una técnica que se utiliza para lograr individualizar, reconocer y clasificar objetos, formas o estructuras en las imágenes analizadas.

Las investigaciones en visión por computadora abordaron el problema del reconocimiento de objetos basándose, en un principio, en la geometría de los mismos [Hu, 1962] [Harris and Stephens, 1988] [Mundy, 2006].

A lo largo de la última década, los algoritmos más exitosos para la resolución de problemas de visión por computadoras fueron basados en descripciones locales invariantes de las imágenes. Como muchas aproximaciones en el procesamiento de imágenes, una popular y eficiente metodología consiste en extraer y comparar partes locales de diferentes imágenes. Sin embargo, para diseñar algoritmos rápidos y, a partir de ellos, obtener representaciones invariantes, locales y compactas, se deben realizar procedimientos de selección y normalización. Una descripción genérica de una imagen es necesaria para evitar comparaciones que requerirán de un alto costo computacional. Los principales desafíos son, por lo tanto, mantener las características más significativas de las imágenes (como bordes y esquinas) y construir una descripción local, a partir de estas características, que sea invariante a distintos problemas como el ruido presente en cualquier imagen, cambios fotométricos y geográficos, puntos de vista, brillo, etcétera.

Para abordar estos problemas, puede consultarse el famoso método detector de esquinas, Stephen-Harris [Harris and Stephens, 1988] y la investigación de Lindeberg en la detección de características en multiescalas [Lindeberg, 1994]. La descripción local invariante de una imagen a partir del análisis multiescala, es un tema reciente.

En la actualidad, se utilizan principalmente regiones concretas de la imagen en las que se considera que existe información suficientemente útil y repetible como para caracterizar un objeto. Esta información se almacena en vectores de características denominados descriptores. La estrategia para llevar a cabo el reconocimiento de objetos, consiste en aplicar a una imagen un método automático de detección y extracción de características con el fin de obtener un conjunto de descriptores. Luego, este mismo método se aplica en diversas imágenes para detectar si contienen un conjunto de descriptores similares al del objeto que se quiere reconocer. Este algoritmo, es capaz de entrenar un sistema para que interprete imágenes y determine el contenido.

El proceso de reconocimiento consiste en cuatro etapas:

1. Detección y obtención de los descriptores de la imagen que contiene el objeto a reconocer.
2. Detección y obtención de los descriptores de una o varias imágenes correspondientes al objeto de interés.
3. Comparación y cálculo de correspondencias entre los descriptores de la primera imagen contra los descriptores del resto de las imágenes.
4. Resultados de correspondencias entre las imágenes y el objeto a reconocer.

En este capítulo se presenta detalladamente SIFT (Scale Invariant Feature Transform), algoritmo que detecta características distintivas de un objeto presente en una imagen. Fue desarrollado por David Lowe en 1999 y patentado en 2004 [Lowe, 1999] [Lowe, 2004].

Las características obtenidas con SIFT son invariantes a la escala y a la rotación de la imagen. Además, son robustas en lo que respecta a cambios de iluminación, oclusión parcial, ruidos y pequeños cambios del punto de vista. Son altamente distintivas, fáciles de extraer y permiten detectar correctamente objetos con un bajo margen de error.

3.2. Descripción general del algoritmo

Para generar un conjunto de descriptores SIFT de una imagen, es necesario, en primera instancia, obtener los puntos característicos y luego, calcular sus vectores descriptores de características a partir de la información de los puntos que lo rodean. SIFT utiliza imágenes en escala de grises, por lo que cada descriptor, contiene información sobre la distribución de los niveles de intensidad alrededor de cada punto de interés previamente obtenido.

El algoritmo SIFT sigue las siguientes fases o etapas:

1. Detección de extremos en cuanto a espacio-escala: Se encuentran puntos de interés o características que son invariantes con respecto a la escala y la orientación usando la diferencia de Gaussianas (DoG).
2. Extracción de puntos claves: Para cada característica candidata se determina la localización y la escala. Una vez que los potenciales puntos claves han sido localizados, se filtran con el objetivo de obtener mejores resultados. Finalmente, se eliminan aquellos que están próximos a los bordes y/o tienen poco contraste.
3. Asignación de la orientación de los puntos claves: Para obtener invarianza con respecto a la rotación de la imagen, se asigna una orientación a cada punto clave detectado basándose en el vector gradiente. Esta orientación mide hacia donde se produce más variación. Todas las operaciones que se realicen sobre la imagen, se harán en base a la orientación, escala y localización de los puntos claves, brindando invarianza a transformaciones.
4. Creación de descriptores: Se crea el descriptor de cada punto clave. Para ello, se genera una matriz cuadrada de 16×16 alrededor de cada punto clave. A continuación, cada bloque se divide en un subbloque de 4×4 . Para cada uno de ellos se crea un histograma de orientación. Finalmente, se genera un descriptor de 128 posiciones.

El algoritmo analiza una imagen a partir del espacio-escala, aplicando sucesivos filtros gaussianos hasta formar una pirámide Gaussiana. A partir de las diferencias de Gaussianas (DoG), se genera una pirámide Laplaciana, de la cual se obtendrán puntos máximos y mínimos que serán los potenciales puntos claves. De todos los puntos obtenidos, se descartan aquellos que son inestables y se genera un descriptor para cada uno de los restantes. Los descriptores son la caracterización alrededor de un punto clave, a su vez, son invariantes a escala, rotación y pequeños cambios de iluminación [Forsyth and Ponce, 2012].

A continuación se explican las etapas con mayor nivel de detalle:

3.2.1. Detección de extremos en cuanto a espacio-escala

La distancia a la que están ubicados los objetos condiciona la percepción de los mismos. Un objeto que se encuentra lejos, presenta un tamaño menor, mientras que si está más cerca, su tamaño será más grande. Estas variantes producen que el mismo objeto pueda aparecer en una imagen con diferentes dimensiones a partir de la distancia de quien observa. Para obtener una descripción robusta de estos objetos se utiliza el espacio-escala [Witkin, 1983].

En [Koenderink, 1984] y [Lindeberg, 1994] se ha demostrado que bajo razonables suposiciones, el único kernel posible para el espacio-escala es la función gaussiana. Por lo tanto, el espacio-escala de una imagen, $I(x, y)$, consiste en un conjunto de imágenes derivadas, $L(x, y, \sigma)$, que se

obtienen a partir de la convolución de una gaussiana de desviación σ y escala variable $G(x, y, \sigma)$, con la imagen de entrada $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (3.1)$$

donde $*$ denota la operación de convolución entre x e y , y donde

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3.2)$$

Un claro ejemplo de espacio-escala se detalla en la figura 3.1, donde diferentes valores de la desviación de la gaussiana producen imágenes con menor nivel de detalle.

Para detectar de forma más eficiente las ubicaciones de los puntos característicos, en [Lowe, 1999] se propuso utilizar los valores extremos del espacio escala de la diferencia de Gaussianas de la imagen, $D(x, y, \sigma)$, que se puede calcular como la diferencia de dos escalas consecutivas separadas por un factor constante k :

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.3)$$

Hay numerosas razones por las cuales se eligió esta función, las más importantes son:

- Es eficiente de calcular ya que $D(x, y, \sigma)$ puede obtenerse restando las imágenes anteriores
- Según se demostró en [Mikolajczyk and Schmid, 2004], los máximos y mínimos de la aproximación al Laplaciano, producen características de imágenes más estables, al compararlas con otras funciones de detección de esquinas como Harris, Hessiano o Gradiente.

Además de construir el espacio escala de cada imagen, mediante la convolución con diferentes gaussianas, se crean varios espacios reduciendo el tamaño de la imagen original llamados octavas. En este procedimiento se obtienen los puntos característicos a partir de los extremos del espacio-escala generado de la diferencia de Gaussianas (DoG). Una pirámide Gaussiana se construye suavizando y reduciendo repetidamente la dimensión de la imagen original. La pirámide de diferencias de Gaussianas se calcula a partir de las diferencias entre los niveles adyacentes de la pirámide de Gaussiana. En la figura 3.2 se detalla lo explicado.

En resumen, la imagen original se parte y se generan imágenes suavizadas a través de la convolución con una gaussiana incrementando su desviación. La relación entre las desviaciones viene determinada por una constante k , de forma tal que si a la primera imagen se le aplica σ , a la segunda se le aplicará $k * \sigma$. Al conjunto de todas las imágenes a las que se les aplica el

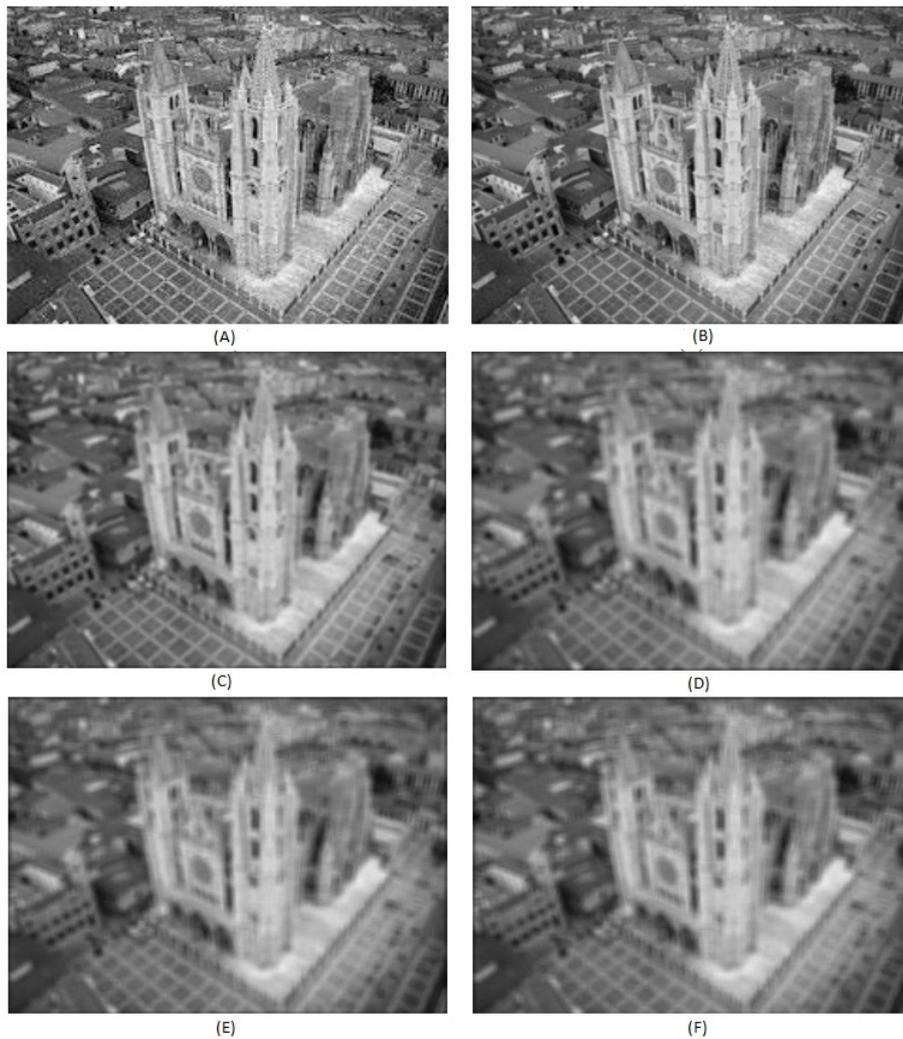


Figura 3.1: Resultado de la convolución de una imagen con una función gaussiana para distintos valores de su desviación σ . (A) Imagen original. (B) $\sigma = 1$. (C) $\sigma = 2$. (D) $\sigma = 4$. (E) $\sigma = 8$. (F) $\sigma = 16$

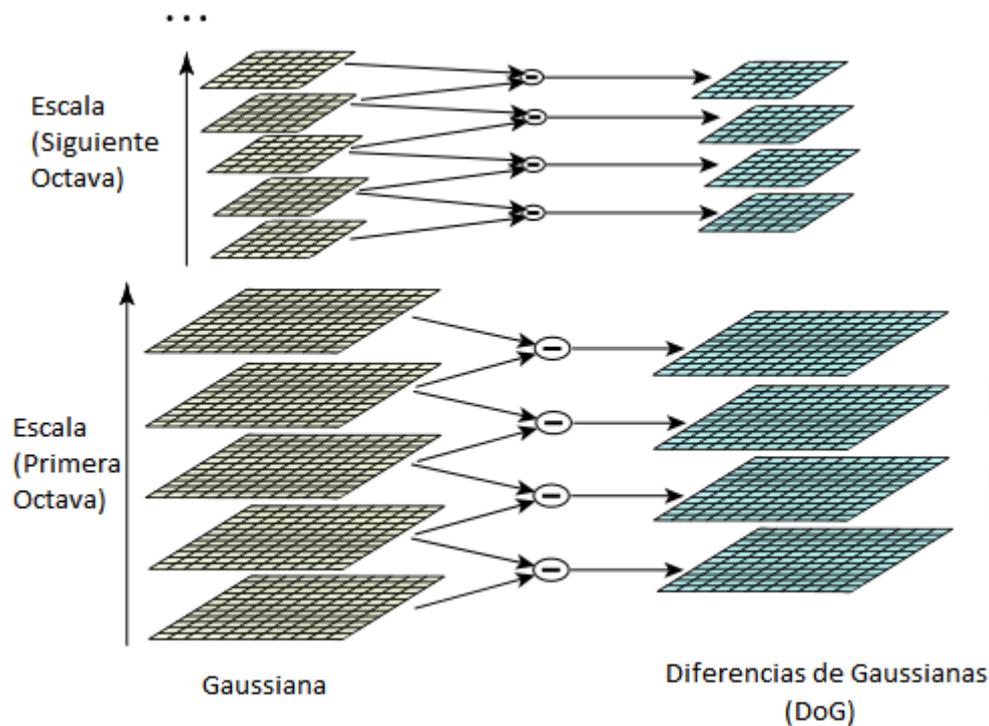


Figura 3.2: A izquierda, se representa el espacio escala de dos octavas. Cada una de ellas tiene 5 escalas. A la derecha, se obtiene el espacio escala de las diferencias de gaussianas.

suavizado, pertenecen a la misma octava. A medida que se divide la imagen a la mitad, aparecerá la siguiente octava y así sucesivamente.

3.2.2. Detección de extremos locales

Para detectar máximos y mínimos locales de la función $D(x, y, \sigma)$, cada punto de la muestra se compara contra sus ocho vecinos de la imagen en la que él se encuentra y también con los nueve vecinos de la escala superior e inferior como se observa en la figura 3.3. El punto se selecciona como posible extremo solo si el valor de D es mayor o menor al de sus 26 vecinos.

3.3. Extracción de puntos claves

En [Brown and Lowe, 2002] se desarrolló un método para ajustar la función cuadrática 3D a los puntos de una muestra local de manera de poder determinar la localización interpolada del máximo. Sus estudios demostraron que este cambio proporciona una mejora al momento de buscar correspondencias y es más estable, ya que se eliminan los puntos de bajo contraste y

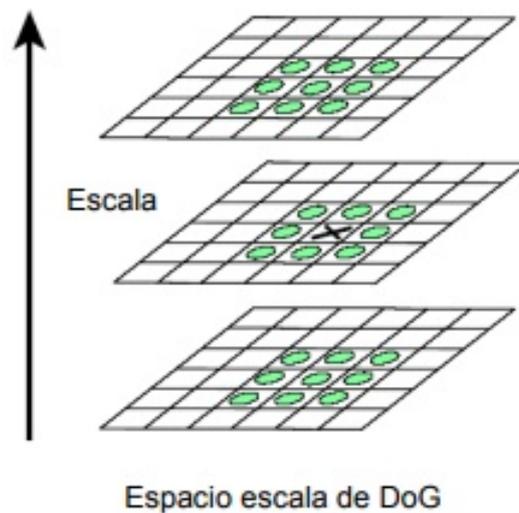


Figura 3.3: Los círculos verdes representan los píxeles vecinos de un punto de interés, que son analizados para determinar si el pixel central es un máximo o mínimo local.

aquellos puntos que puedan estar a lo largo de un borde.

La propuesta de Brown utiliza lo realizado por Taylor hasta los términos cuadráticos de la función espacio-escala, $D(x, y, \sigma)$, desplazada de forma que el punto de origen esté en el punto de la región. Brown sugirió aproximar el Hessiano y la derivada de D utilizando diferencias entre vecinos del punto de la muestra. El valor de la función en el extremo, $D(\hat{x})$, es útil para rechazar extremos con bajo contraste en la siguiente función:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad (3.4)$$

3.3.1. Eliminación de bordes

Los puntos característicos detectados a lo largo de los bordes son puntos menos estables y pueden ser similares a pequeñas cantidades de ruido. Por este motivo, se los debe eliminar.

Como la diferencia de Gaussianas es una función equivalente al Laplaciano de la Gaussiana, la respuesta que tiene sobre los bordes es muy fuerte. Un punto ubicado en un borde y mal definido en la función de diferencia de Gaussianas presentará una gran curvatura a lo largo del borde, pero una pequeña curvatura en la dirección principal. Las curvaturas principales pueden

ser calculadas, a una determinada escala, a partir del Hessiano de 2×2 , H :

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \quad (3.5)$$

Donde D_{xx} es la derivada segunda en x , D_{xy} es la derivada segunda en x e y . Las derivadas se calculan a partir de las diferencias de los vecinos del punto de muestreo. Los valores propios de H son proporcionales a las curvas principales. Siendo α el valor propio con mayor magnitud y β el de menor, la suma (3.6) y el producto (3.7) de ambos puede obtenerse de la siguiente manera:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (3.6)$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (3.7)$$

En casos en los que el determinante sea negativo el punto también se descarta como posible extremo ya que las curvaturas tienen diferentes signos. Sea r la relación entre el valor propio de la magnitud más grande (α) y la más pequeña (β), tal que $\alpha = r\beta$, se puede expresar esta relación en función a la traza (3.6) y al determinante del Hessiano (3.7), de la siguiente manera:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r} \quad (3.8)$$

Cuando ambos valores propios son iguales, el resultado de $(r + 1)^2/r$ será mínimo y se irá incrementando con r . Por lo tanto, para comprobar que la relación de las curvaturas principales es inferior a un umbral r se debe cumplir la siguiente ecuación:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r + 1)^2}{r} \quad (3.9)$$

Se utiliza un valor de $r = 10$ para eliminar puntos claves con una relación entre las curvaturas principales con valores mayores de 10.

3.4. Asignación de la orientación de los puntos claves

Con el objetivo de conseguir invarianza a la rotación, a cada descriptor se le asigna una orientación en función de las propiedades locales de la imagen.

Para llegar a esto, se selecciona la imagen Gaussiana suavizada, L , con la escala más próxima a la del punto característico del que se va a obtener el descriptor. Por cada región de la imagen

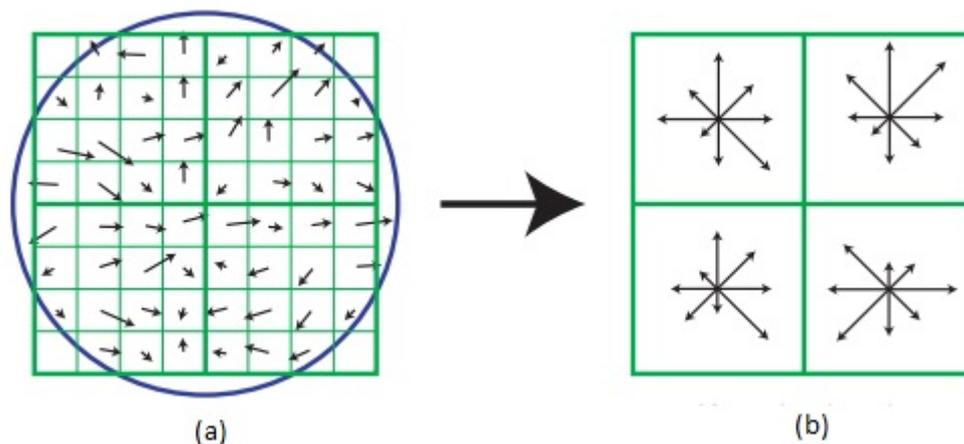


Figura 3.4: (a) Gradientes de la imagen. (b) Descriptor del punto clave

en una escala específica, $L(x, y)$, se calcula la magnitud del gradiente, $m(x, y)$, y su orientación, $\theta(x, y)$, utilizando diferencias entre los píxeles vecinos, como en las ecuaciones 3.10 y 3.11.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + ((L(x, y+1) - (L(x, y-1)))^2} \quad (3.10)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right) \quad (3.11)$$

A partir de las orientaciones obtenidas de la región alrededor del punto característico, se calcula un histograma de 36 niveles cubriendo todas las posibles orientaciones (360°).

Las direcciones dominantes de los gradientes locales se corresponderán con los picos del histograma. Se detecta el pico más alto y para él y todos aquellos con un valor superior al 80 % de ese pico más alto, se creará un punto característico en la misma posición pero con las correspondientes orientaciones dominantes. Aunque los puntos con múltiples orientaciones son únicamente alrededor del 15 % según Lowe [Lowe, 2004], éstos contribuyen mucho a que la búsqueda de correspondencias sea estable.

3.5. Creación de descriptores

Mediante las operaciones explicadas en las secciones previas, a cada uno de los puntos característicos obtenidos, se les ha asignado una posición en la imagen, una escala y orientación. El último paso del método consiste en calcular un descriptor que sea altamente característico y al mismo tiempo lo más invariante posible a variaciones restantes, como cambios en iluminación y puntos de vista 3D. La figura 3.4 describe el cálculo del descriptor de puntos claves:

En forma de resumen, el cálculo del descriptor SIFT se lleva a cabo en los siguientes pasos:

1. Se obtiene la magnitud y orientación del gradiente en una ventana de 16×16 píxeles centradas en el punto característico.
2. Se calcula un histograma de 8 niveles con las orientaciones del gradiente para cada una de las 16 regiones que se obtienen al dividir la ventana inicial en regiones menores de 4×4 . Cada nivel se corresponde con 45° , el primero de 0 a 44° , el segundo de 45° a 89° y así sucesivamente. El histograma contendrá las orientaciones de los gradientes que aparecen en esa ventana.
3. Se normaliza el vector de 128 elementos que se obtiene al concatenar los histogramas de 8 valores para cada una de las 16 divisiones que se han realizado en la región.

Estos pasos se detallan a continuación:

1. **Obtención de los gradientes en una ventana de 16×16 muestras alrededor del punto característico.** Para conseguir estos valores, se realizan las siguientes operaciones:
 - a) Se calculan a la escala apropiada, seleccionando la imagen con nivel de suavizado Gaussiano que se corresponde con la escala del punto característico.
 - b) Se obtienen con invarianza a la orientación, rotando las coordenadas del descriptor y las orientaciones del gradiente en relación con la orientación del punto característico.
 - c) Se obtienen de manera eficiente, calculándolos previamente para todos los niveles de la pirámide de Gaussianas. El cálculo es el mismo realizado para asignar la orientación al descriptor, explicado en secciones anteriores.
 - d) Se obtienen valores ponderados por una gaussiana para evitar cambios bruscos en el descriptor ante pequeñas modificaciones en la posición de la ventana, y para limitar la influencia de los gradientes alejados del centro del descriptor. Para conseguir esto, se asigna un peso a la magnitud de cada punto mediante una función gaussiana con σ igual a la mitad del ancho de la ventana del descriptor.
2. **Histograma de orientaciones del gradiente para subregiones de 4×4 muestras**
 - a) Se obtiene el histograma de 8 direcciones para cada subregión.
Para cada una de las 16 subregiones de 4×4 se calcula un histograma de 8 intervalos, que son las orientaciones que resultan al subdividir el espacio en regiones de 45° . El valor que se le asigna a cada uno de los 8 intervalos del histograma, proviene de la suma de las magnitudes del gradiente de todas las muestras cuya orientación se corresponde con la subregión del histograma.

- b) Se distribuye el valor de cada gradiente en los intervalos adyacentes del histograma. Esto evita los efectos de los límites de manera que el descriptor pueda cambiar de forma abrupta cuando una muestra pase de un histograma a otro o de una orientación a otra. Cada una de las entradas a un intervalo del histograma se multiplica por un peso de $1 - d$ para cada dimensión, donde d es la distancia de la muestra del valor central del intervalo medida en unidades del espaciado entre los intervalos del histograma [Lowe, 2004].

Una de las ventajas de este método es que una muestra del gradiente puede desplazarse varias posiciones, dentro de las subregiones de 4×4 , y aún contribuir de la misma forma al histograma obtenido, haciendo al descriptor robusto ante desplazamientos locales de la posición de los puntos característicos.

El descriptor se forma concatenando los 16 histogramas de 8 direcciones de manera que el vector de características tendrá una longitud de $4 \times 4 \times 8 = 128$ elementos, para cada punto característico.

3. Normalización del vector de características

Como los valores de los gradientes se calculan a partir de diferencias entre niveles de gris de los píxeles, el descriptor que se obtiene es invariante a cambios en la iluminación. Más allá de esto, pueden ocurrir cambios no lineales en la iluminación debido a la saturación de la cámara o cambios en superficies 3D que presenten diferentes orientaciones. Para evitar estos problemas, se realiza la siguiente normalización:

- a) Se normaliza el vector a longitud unidad.

Dado el vector de 128 elementos obtenido para un punto característico, se obtiene la longitud de dicho vector, y luego se divide cada uno de sus elementos por la magnitud. De esta forma el vector pasará a tener longitud unidad:

$$\hat{u} = \frac{u}{\|u\|} \quad (3.12)$$

- b) Se aplica un umbral de 0,2 y se vuelve a normalizar el vector como en (a)

Esto implica aplicar una umbralización a cada elemento del vector de manera tal que ninguno tenga un valor superior a 0,2. Finalmente el vector se normaliza a la longitud unidad.

3.6. Otros métodos similares

Esta sección describe brevemente otros métodos que, al igual que SIFT, permiten reconocer objetos en una imagen por medio de descriptores. SIFT es el algoritmo raíz a partir del cual se implementaron todos los siguientes métodos.

- **SURF**: Al igual que SIFT, SURF [Bay et al., 2008] consta de las mismas etapas pero introduce modificaciones a la hora de realizar los cálculos en las mismas. Su objetivo es mejorar la velocidad de sus procesos con respecto a SIFT, reduciendo el tamaño del descriptor y utilizando filtros de caja e imágenes integrales. En la etapa de detección de puntos de interés, SURF utiliza el determinante de la matriz Hessiana para la localización y escala de los puntos. Esta matriz está formada por la convolución de la derivada parcial de segundo orden de la Gaussiana, calculada a partir de una imagen I , en un punto p , perteneciente a la escala σ . Para analizar el espacio-escala, SURF utiliza filtros de caja [Simard et al., 1999] [Viola and Jones, 2001], que aproximan las derivadas parciales de segundo orden de la gaussiana y se evalúan de manera rápida a través de imágenes integrales. Una imagen integral es una nueva imagen, donde para cada punto, se guarda la suma acumulada de todas las intensidades de la imagen desde el punto de origen hasta el punto de interés. Una vez que se obtienen los puntos de interés con su escala, se continúa con la generación del descriptor, donde el primer paso es el cálculo de la orientación. En esta etapa, SURF se comporta diferente a SIFT. Para conseguir un punto invariante a la rotación, se calculan las respuestas de Haar sobre las coordenadas x e y de un punto de interés. Una vez estimada la orientación dominante, se suman las respuestas de Haar junto con sus valores absolutos obteniendo un descriptor de 64 dimensiones.
- **PCA-SIFT**(Principal Component Analysis): Este método se comporta igual que SIFT, excepto en la etapa de generación de descriptores de los puntos de interés. Se obtiene por cada punto clave, un vector con una dimensión de 3042, que luego se reduce a 32 a través de PCA [Ke and Sukthankar, 2004].
- **C-SIFT**: Este método es más robusto que el método SIFT convencional con respecto al color y a las variaciones fotométricas. Se han realizado evaluaciones en [Abdel-Hakim and Farag, 2006] que certifican la afirmación. Su desventaja es que su costo computacional es mayor a SIFT.
- **RIFT** (Rotation Invariant Feature Transform): El descriptor es construido usando regiones circulares normalizadas divididas en anillos concéntricos de igual ancho, y con cada anillo se computa un histograma de orientación del gradiente [Lazebnik et al., 2005]. A diferencia de SIFT, la orientación está representada por 8 intervalos distintos, por lo que el descriptor final tiene un total de 24 entradas.

- **U-SURF**: Al igual que SURF, para la obtención de descriptores genera un vector de dimensión 64, donde se incluye la escala y la rotación pero no la posición. Resulta más rápido que SURF, ya que no realiza cálculos de orientación, es decir, que omite el proceso de asignación de orientación [Bay et al., 2006] .

3.7. Conclusiones

En el ámbito del procesamiento de imágenes, la búsqueda de correspondencias de características locales es la tendencia actual, tanto en imágenes como en videos. A pesar de que es un área que está siendo estudiada, no se ha establecido un método óptimo y genérico que sirva para resolver todos los problemas.

SIFT permite representar una imagen como un conjunto de vectores generados a partir de puntos de interés detectados a diferentes niveles de escala e intensidad. Estos vectores son lo suficientemente distintivos permitiendo realizar tareas de clasificación y reconocimiento. Si bien otros autores han desarrollado distintos métodos que buscan mejorar los resultados que brinda SIFT, éste continúa siendo el más utilizado en la práctica debido a su capacidad de caracterización de objetos, su fuerte invarianza a los cambios de escala y rotación y a su invarianza parcial frente a cambios de puntos de vista e iluminación.

En el capítulo siguiente, se presenta el modelo propuesto en esta tesina, detallando la selección de imágenes que lo componen, las distintas pruebas que se realizaron con SIFT y SURF que llevaron a decidir por uno de los dos métodos y otras misceláneas que resultan interesantes a tener en cuenta al momento de la construcción de una base de datos.

Capítulo 4

Definición del modelo de datos

4.1. Introducción

Para abordar este tipo de problemas es necesario contar con una base de datos que sirva como fuente de consulta. Por ejemplo, para enfrentar un problema clásico como el reconocimiento de rostros se pueden utilizar las bases de datos publicadas en [Mislav Grgic, 2007], que están conformadas con imágenes apropiadas para resolver estos casos. De la misma forma, es necesario contar con una BDD que contenga imágenes que representen a los ejemplares que se desean estudiar en esta tesina. La misma puede estar compuesta por cualquier imagen que se considere lo suficientemente distintiva como para resolver la problemática y que contenga el menor ruido posible. Estas imágenes pueden detallar el cuerpo entero de un ejemplar o bien, partes más puntuales como escamas, pupilas, dibujos en el cuerpo, etcétera.

Si bien existen publicaciones que dan soluciones a propuestas similares, en la actualidad no se han encontrado bases de datos que sirvan de ejemplo o referencia para tratar el problema a resolver en esta tesina.

En [Roly et al., 2015], los autores orientan la construcción de la base de datos en base al estudio de la composición del veneno de las serpientes y sus aplicaciones en la medicina en Bangladesh. Otra aproximación se puede encontrar en "The Reptile Database"[Uetz and Etzold, 1996], este sitio ofrece una base de datos pública, la cual cubre aspectos descriptivos y taxonómicos de diferentes especies del mundo. Más allá de esto, su contenido no satisface las necesidades para utilizarla como fuente de estudio para este desarrollo. Por último, el caso de estudio que más se acerca al que se propone en esta tesina, se detalla en [James, 2017], donde el autor investiga y estudia diferencias entre dos tipos de serpientes (Víbora de Russell y Cobra india), que habitan en el continente asiático. En esta investigación, se toman muchas características que si bien son distintivas, son difíciles de capturar al momento de tomar una fotografía, como por ejemplo las pupilas, partes frontales y parietales de la cabeza, partes nasales, preoculares, postoculares,

etcétera.

En este capítulo se presenta el modelo de datos propuesto en esta tesina para el reconocimiento de especies de serpientes. Debido a la natural complejidad del dominio, el modelo está compuesto por diversos módulos que abordan los diferentes problemas como la segmentación de patrones distintivos de los ejemplares, cálculo de descriptores y clasificación. Por otro lado, se detallan pruebas realizadas para la selección del criterio con el cual se detectaron los puntos de interés para la posterior generación de los descriptores de características.

4.2. Descripción del modelo

El modelo desarrollado en esta tesina consta de diversas etapas para, finalmente, lograr la clasificación de especies de serpientes. Si bien, las mismas tienen muchas partes distintivas ya que se pueden diferenciar por su color y la forma de su pupila, o en otros casos solamente con las escamas de sus cabezas, como en la figura 4.1, en esta investigación se tuvo en cuenta el contexto en el cual se usaría la aplicación (en pajonales, pastizales, arenales, arbustos, cuevas subterráneas, roquedales [Giambelluca, 2015]). Suponiendo un encuentro casual con un ejemplar, las probabilidades de obtener una foto de su pupila o de las escamas de su cabeza son casi nulas, teniendo en cuenta que una persona que no tenga conocimientos en el tema, consideraría un peligro acercarse al animal, por lo tanto se optó por los dibujos o texturas de la parte dorsal de los ofidios. Las características de reconocimiento mencionadas anteriormente se resumen en la figura 4.2. Si bien parece sencillo ejecutar el árbol de la figura 4.2, el usuario no siempre está en condiciones de responder muchas de las opciones que determinan la respuesta final. Por ejemplo, una persona sin conocimientos sobre serpientes, no sabe distinguir la diferencia entre lisas y listadas.

Por otro lado, resulta complejo obtener una foto de cuerpo entero de un ejemplar debido a que, como se mencionó, los lugares usuales de encuentro no son óptimos para conseguir una buena captura. Además, en ocasiones, las serpientes se encuentran enrolladas, ocultando partes de su cuerpo.

La construcción del modelo de datos se resume en los siguientes pasos:

1. Selección de partes/patrones distintivos de las especies a estudiar.
2. Obtención de los puntos de interés y sus descriptores correspondientes.

La base de datos presentada tiene dos objetivos: en primer lugar, al estar compuesta por recortes, facilita la tarea de segmentación ya que las partes no cuentan con un fondo ni objetos que generen oclusión. Por otro lado, el conjunto de datos propuesto pretende ser un primer paso en



Figura 4.1: A la izquierda, la cabeza de una Yará. Se aprecia como el dibujo de su cuerpo permanece continuo. A la derecha, la cabeza de una Culebra con sus placas distintivas.

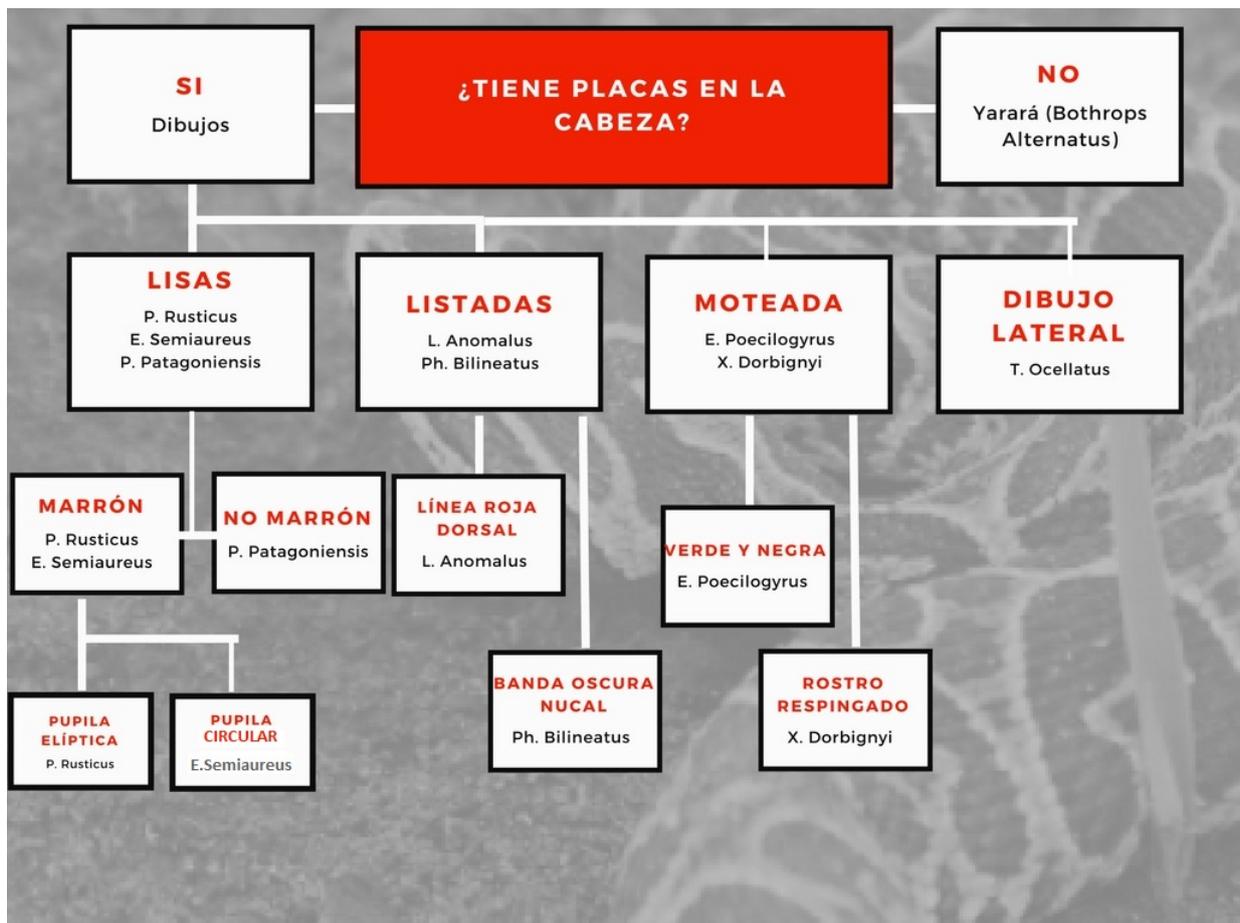


Figura 4.2: Árbol de decisión donde se detallan características de cada especie

la construcción de una base de datos que contemplen especies de la región, inexistente hasta el momento. A lo largo del capítulo se hace foco en cada etapa que conlleva a la construcción de la BDD.

4.3. Construcción de la base de datos

En esta sección se detallan características relevantes de la base de datos como la selección de imágenes y las dificultades que esto presenta, la representación de las mismas en forma de descriptores que conformarán la BDD y por último pruebas y medidas obtenidas que demuestran la robustez del modelo. El fin del mismo es utilizarlo como fuente de consulta para posteriormente, clasificar una imagen según las cinco especies estudiadas.

4.4. Selección de imágenes

Las imágenes se han obtenido, en su mayoría, en ambientes naturales. En su selección, se ha tenido consideración especial con distintas variantes como:

- El ciclo en que se encuentra el ejemplar: Como se mencionó en el capítulo 2, las serpientes mudan de piel, en esta etapa la misma luce más apagada, lo que implica cambios críticos al momento de obtener los descriptores de las imágenes.
- El contexto de hallazgo del ejemplar: Cabe destacar que dependiendo de la condición climática, la piel de animal se verá distinta, por ejemplo, en un lugar húmedo la piel será más brillante que en una zona rocosa, donde probablemente tenga vestigios de tierra encima de ella. En la figura 4.3 se observan las distintas tonalidades de la piel de los ofidios como consecuencia del contexto. Por otro lado, un ejemplar puede ser encontrado en pastizales, roquedas, etcétera. En estos casos, el entorno nunca es controlado, sino que en algunas situaciones puede ser muy complejo (por ejemplo en zonas acuáticas rodeadas de pastos) y provocar la oclusión de distintas partes del animal.
- La postura del ejemplar: No siempre se puede encontrar al animal estirado. En muchas situaciones, el mismo se encuentra enrollado en si mismo o en una rama por ejemplo.

Por los motivos mencionados, y siguiendo las recomendaciones del asesor profesional de esta tesina, se seleccionaron aquellas partes distintivas de las especies a estudiar, y se realizaron los recortes correspondientes de forma manual como se observan en la figura 4.4. De esta manera, se descartó el fondo y el ruido de cada segmento. Es importante resaltar que en todas las imágenes se manejó un nivel de escala que permite distinguir los dibujos característicos de cada especie.



Figura 4.3: En las figuras A y C, se puede observar la piel de los ofidios con colores más vívidos debido a su contacto con el agua. En las figuras B y D, se perciben los colores más apagados debido a su contacto con la tierra y el polvo proveniente de las rocas, además se encuentran enrollados en ambos casos.



Figura 4.4: Imágenes de distintos recortes de cuerpos de serpientes. En cada imagen, se aprecian los patrones que caracterizan a cada ofidio

4.5. Generación de descriptores

La detección de puntos de interés para la posterior generación de descriptores, es uno de los pasos críticos para la construcción de este modelo. Para aplicar distintas técnicas de clasificación o reconocimiento, es necesario representar las imágenes seleccionadas como descriptores. Para abordar este problema, se utilizaron los algoritmos SIFT y SURF presentados en el capítulo anterior para generar los descriptores correspondientes a cada imagen. Una vez obtenidos, se realizaron pruebas donde se analizó que tan representativos son los descriptores generados por ambos métodos. Para esto, es necesario la selección de un umbral óptimo con el cual se clasificarán las imágenes. En esta sección, se detallan resultados obtenidos que justifican la elección del método junto su umbral.

Por otro lado, se realizaron pruebas con detectores y descriptores populares en la literatura, como MSER (Maximally Stable Extremal Regions) [Matas et al., 2002] y LBP (Local Binary Pattern) [Ahonen et al., 2006]. Viendo que su desempeño fue inferior al obtenido con SIFT y SURF se decidió no incluir las mediciones correspondientes.

Para la generación de los vectores correspondientes a distintas regiones de una imagen, se necesita que las imágenes elegidas no posean ruido que pueda causar problemas en la etapa del reconocimiento. Cada fragmento seleccionado es convertido en su conjunto de vectores SIFT. Cada vector corresponde a un punto característico de la imagen y contiene información de la región que lo rodea. El conjunto de descriptores obtenido conforma la base de datos. El algoritmo 1 expresa lo antes expuesto.

Algoritmo 1: Pseudocódigo del método para seleccionar los vectores SIFT que conforman la base de datos

imList ← imágenes seleccionadas

for *i* = 1 to *Cant. de imágenes(imList)* **do**

 leer la *i*-ésima imagen de la lista

 Detectar los puntos de interés a partir de la DoG

 Determinar la escala y localización de los puntos obtenidos en el paso anterior

 Eliminar puntos ubicados en los bordes y aquellos con bajo contraste

 Asignar orientación a los puntos clave

 Generar los descriptores correspondientes a cada punto clave

 Guardar los descriptores en la BDD

Fin del for ← Retorna la base de datos

Las versiones de SIFT y SURF utilizadas, generan descriptores a partir de imágenes en tonos de grises. Por este motivo se analizaron distintas funciones de conversión del espacio de color RGB a tonos de grises y se desarrollaron pruebas para determinar cual producía mejores resultados. Finalmente, como estas pruebas no produjeron mejoras significativas se optó por los valores

RGB estándar: $0,2989 * R + 0,5870 * G + 0,1140 * B$. Una vez que se obtuvieron los descriptores, se continuó por la selección del umbral a partir del cual se clasificarán las distintas especies de serpientes. Para esto se calcularon las distancias euclídeas entre cada descriptor de cada imagen contra los restantes. A partir de estas distancias, se seleccionaron las dos mejores para obtener los cocientes correspondientes. Finalmente para encontrar el umbral óptimo se realizó una prueba donde se analizaron distintos valores que se compararon contra los cocientes generados anteriormente. Aquellos que superaron el valor del umbral se consideraron aciertos. Los siguientes pasos resumen lo explicado:

1. Por cada descriptor de cada imagen, se calculó la distancia euclídea contra todos los descriptores de todas las imágenes.
2. Una vez obtenidas las distancias, se calcularon los cocientes entre las dos mejores por cada descriptor. Estos cocientes se utilizan en el algoritmo 3 para determinar el umbral óptimo.

Los pasos explicados anteriormente se resumen en el algoritmo 2.

Algoritmo 2: Pseudocódigo del cálculo de las distancias euclídeas y generación de cocientes.

```

for  $i = 1$  to Cant. de imágenes de la BDD do
  for  $j = 1$  to Cant. de imágenes de la BDD do
    if  $i <> j$  then
      for  $d = 1$  to Cant. descriptores de la imagen  $i$  do
        Calcular distancia entre descriptor de la imagen  $i$  y
        cada uno de los descriptores de la imagen  $j$ 
        Cociente( $i,j,d$ ) =  $min2/min1$  siendo  $min1$  y  $min2$  las dos distancias
        mínimas obtenidas en el paso anterior, tal que  $min1 < min2$ 

```

Para medir las coincidencias en [Vedaldi and Fulkerson, 2008], los autores recomiendan utilizar un umbral de 1,5. Para definir el umbral óptimo para este problema en particular, se generó un algoritmo que itera sobre un rango de valores (posibles umbrales) y compara a cada uno contra los cocientes generados en el algoritmo 2. El mejor umbral se decide en base a la mejor tasa de acierto.

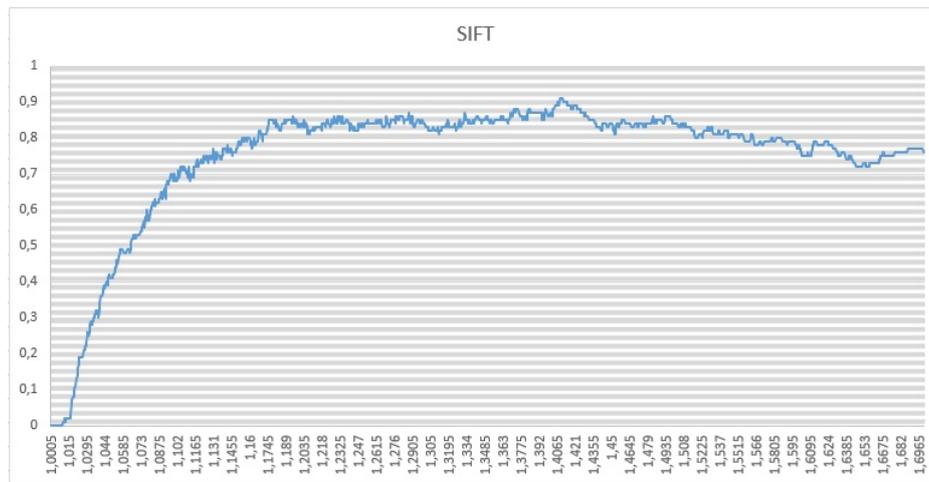


Figura 4.5: Mejor tasa de acierto obtenida con SIFT(91 %) se detectó con un umbral de 1.4080

El algoritmo 3 resume lo antes expuesto.

Algoritmo 3: Pseudocódigo del cálculo del mejor umbral junto con su tasa de acierto

umbrales ← lista de valores de umbrales a verificar

aciertosPorEspecie ← contendrá el total de aciertos por especie para cada valor del umbral

for *i* = 1 to Cant. de imágenes de la BDD **do**

for *u* = 1 to Cant. de umbrales **do**

for *j* = 1 to Cant. de imágenes de la BDD **do**

if *i* <> *j* **then**

for *d* = 1 to Cant. descriptores imagen *i* **do**

if *cocientes*(*i*, *j*, *d*) > *umbrales*(*u*) **then**

aciertosPorEspecie(*k*) ++

 siendo *k* la especie a la que pertenece la imagen *j*

if imagen(*i*) fue correctamente clasificada **then**

aciertosPorUmbral(*u*,*i*) ++

 Fin del for ← Retornar el mejor umbral según la tasa de acierto

Los resultados de la salida del algoritmo 3, para SIFT y SURF, se reflejan en las figuras 4.5 y 4.6, respectivamente. El umbral seleccionado se utilizó para hacer las pruebas que se explican en el capítulo 5.

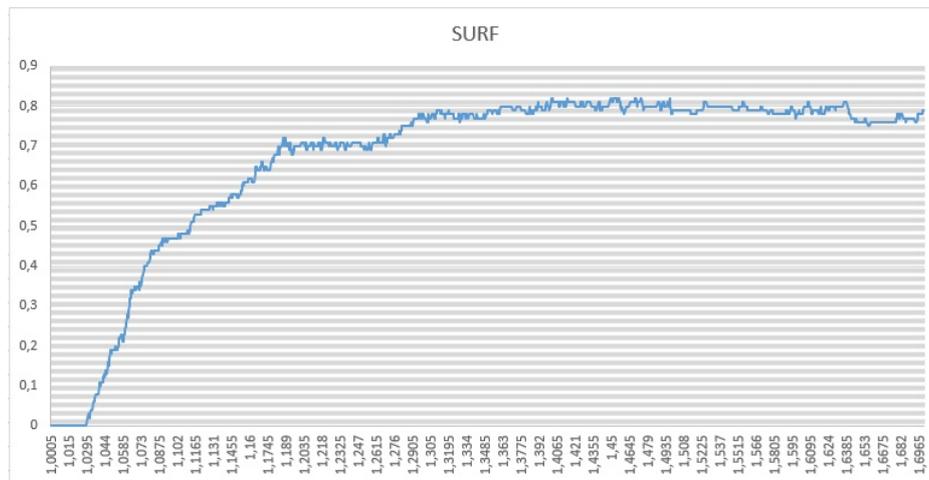


Figura 4.6: Mejor tasa de acierto obtenida con SURF(82 %) se detectó con umbral de 1.4010

4.5.1. Análisis de descriptores

Para obtener información más detallada de los descriptores, se generó un algoritmo donde se los analizó uno a uno observando cuan representativos son, contabilizando en un vector de *hits* los aciertos. Si bien, los resultados devueltos por SIFT superan a los de SURF en la tabla 4.1 se refleja que SIFT genera muchos descriptores que no se corresponden con otros que representen a la misma especie. Para realizar este análisis, se tuvieron en cuenta los siguientes criterios:

- Descriptores sin coincidencias: Son aquellos que no suman para ninguna especie de las estudiadas, generando un vector de *hits* del estilo $[0, 0, 0, 0, 0]$.
- Descriptores con un máximo incorrecto: Son aquellos que en el vector de *hits* obtuvieron un máximo incorrecto, es decir, que la especie que debería haber ganado quedó en segundo lugar con respecto a la cantidad de aciertos. Por ejemplo, a partir de un descriptor perteneciente a la especie 1, se genera el siguiente vector $[1, 0, 2, 0, 0]$.
- Descriptores con varios máximos incorrectos: Son aquellos que en el vector de *hits* obtuvieron más de un máximo incorrecto (peor de los casos). Continuando el ejemplo anterior: $[1, 0, 3, 2, 0]$.
- Descriptores con un máximo correcto: Son aquellos que en el vector de *hits* obtuvieron un máximo. Por ejemplo: Si a partir de un descriptor que corresponde a la especie 1, se obtiene el siguiente vector $[1, 0, 0, 0, 0]$.
- Descriptores con varios máximos correctos: Son aquellos que en el vector de *hits* obtuvieron varios máximos correctos. Siguiendo el ejemplo del ítem anterior, para la especie 1 un vector con varios máximos correctos sería $[3, 0, 0, 0, 3]$.

Tabla 4.1

Algoritmo	SIFT	SURF
Descriptores sin coincidencias	55.35 %	26.28 %
Descriptores con un máximo incorrecto	6.21 %	29.17 %
Descriptores con varios máximos incorrectos	0.92 %	6.99 %
Descriptores con un máximo correcto	34.12 %	23.82 %
Descriptores con varios máximos correctos	3.40 %	13.75 %

El algoritmo 4 describe lo antes expuesto:

Algoritmo 4: Pseudocódigo del proceso para los aciertos por cada vector de cada descriptor

```

umbral ← valor obtenido del algoritmo 3
hitsPorvector ← matriz para contar aciertos por especie por vector
aciertosPorEspecies ← contendrá el total de aciertos por especie
for i = 1 to Cant. de imágenes de la BDD do
  for j = 1 to Cant. de imágenes de la BDD do
    if i <> j then
      for d = 1 to Cant. descriptores imagen i do
        if cocientes(i,j,d) > umbral then
          aciertosPorEspecies(k) ++
          siendo k la especie a la que pertenece la imagen j
          hitsPorvector[i,d] ← aciertosPorEspecie
      descriptoresConAciertos ← [nroImagen, especie, descriptoresBDD, hitsPorvector]
  Fin del algoritmo ← Retornar descriptores con aciertos

```

4.6. Validación del modelo

La validación de un modelo de datos se utiliza en entornos donde se desea conocer la precisión del mismo [Reeves, 1992]. Una de las tantas formas de estimar la precisión de un modelo consiste en tomar los datos de la muestra y hacer dos subconjuntos. Un subconjunto para los datos de entrenamiento y otro para los datos de prueba. Una vez definidos los subconjuntos, se evalúa el modelo con los datos de prueba, que al estar etiquetados, permiten medir la precisión del modelo.

Según como se divida la muestra, existen distintos tipos de validaciones:

- Validación cruzada dejando uno fuera o *Leave-one-out cross validation*. Se detalla en la

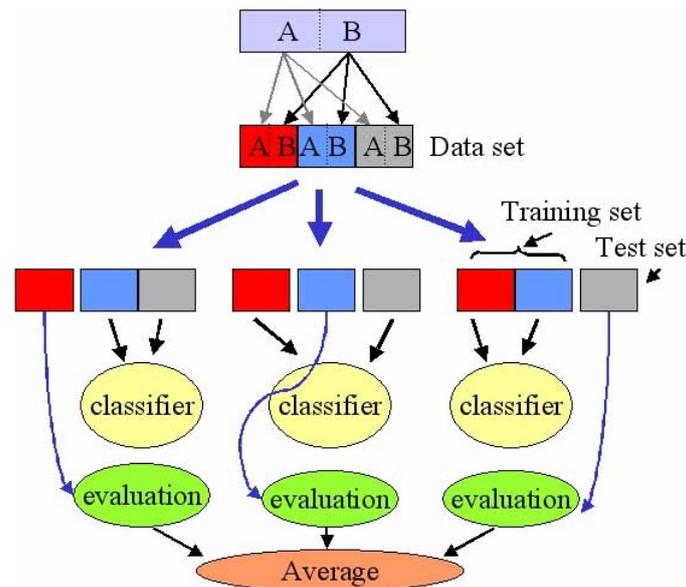


Figura 4.7: Etapas de la validación cruzada de un modelo de datos. En este caso $k = 3$ [Lang, 2010].

figura 4.8.

- Validación cruzada de K iteraciones o K -fold cross validation. Figura 4.7.

4.7. Tipos de validación cruzada

A continuación se detallan las distintas variantes que tiene el método de validación cruzada:

4.7.1. Validación cruzada dejando uno fuera

La validación cruzada dejando uno fuera [Arlot and Celisse, 2010], consiste en separar los datos de forma tal que, en cada iteración se cuente con una sola muestra para los datos de prueba, y el resto para los datos de entrenamiento. En este tipo de evaluación el costo computacional es muy elevado ya que se tienen que realizar N repeticiones, donde N es la cantidad de muestras que hay en el modelo. Para cada muestra se analizan los datos de prueba y de entrenamiento [Elkan, 2012], es decir, que las muestras no se dividen en subconjuntos, sino que, en cada repetición se omite un dato que será utilizado para realizar la prueba y el resto serán los datos de entrenamiento. Una vez terminada esta primera iteración, se almacenan los datos correspondientes a la primera corrida, y se continúa con el siguiente dato, el cual es usado como prueba y el resto, incluyendo el último dato de prueba, conforman el grupo de datos de

Iteración	Número de muestra									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	5	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	9	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	3	4	5	6	7	8	9	10

Figura 4.8: Etapas de la validación cruzada dejando uno fuera sobre 10 muestras. En cada iteración la muestra de color verde se utiliza como testeo, mientras que las restantes se usan como entrenamiento.

entrenamiento. Una vez finalizadas las evaluaciones con todos los datos, se calcula el porcentaje de aciertos [Liu et al., 2008] [Baumann, 2003]. El método utiliza la mayoría de los datos para entrenar, donde se obtiene un resultado determinista, ya que cada experimento siempre ofrecerá los mismos resultados. La figura 4.8 ilustra un ejemplo de este tipo de validación.

4.7.2. Validación cruzada de K iteraciones

La validación cruzada de K iteraciones 4.9, la cual se utiliza en esta investigación, es una generalización del método anterior. Los datos de muestra se dividen en K subconjuntos, uno se utiliza como datos de prueba y el resto ($K - 1$) como datos de entrenamiento. El proceso de validación es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos. Finalmente, se obtiene la media aritmética de los resultados de cada iteración para obtener un resultado único. En la literatura es una de las técnicas más utilizadas, ya que es considerada una medida fiable para evaluar distintos modelos de datos en un tiempo razonable [Bouckaert, 2008].

En la práctica, existen diversas formas de seleccionar el subconjunto de datos de prueba:

- Selección aleatoria: El subconjunto de datos de prueba se construye de forma arbitraria, lo que puede llevar a que haya un desbalance en el análisis.
- Selección estratificada: El subconjunto de datos de prueba se construye de forma equitativa. Por ejemplo, en las pruebas estratificadas realizadas en esta investigación, el subconjunto de prueba en cada iteración tendrá la misma cantidad de ejemplares.

La elección del número de repeticiones depende del tamaño del conjunto de datos. Para esta tesina se utilizó la validación cruzada de 10 iteraciones tanto para las pruebas aleatorias como

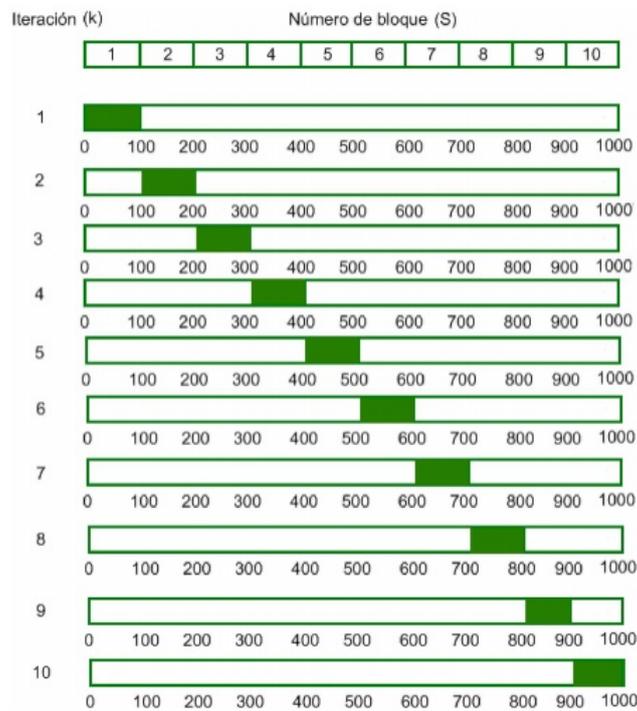


Figura 4.9: Validación cruzada de K iteraciones con $K = 10$. En cada iteración la muestra de color verde se usa como testeo. Para armar el subconjunto de prueba se utiliza el 10% de la muestra total. Con respecto a las pruebas aleatorias y estratificadas, el procedimiento es el mismo, únicamente varía la cantidad de especies de serpientes elegidas en cada muestra de prueba.

para las estratificadas.

El algoritmo 5 resume lo antes explicado.

Algoritmo 5: Pseudocódigo de validación cruzada sobre el modelo de datos

```

umbral ← valor obtenido del algoritmo 3
corridas ← 30
cantidadCortes ← 10
tipoCortes ← estratificado o aleatorio
for i = 1 to Cant. de tipoCortes do
  aciertosPorCorrida[corridas, cantidadCortes] ← matriz de 30x10
  for corrida = 1 to Cant. de corridas do
    mezcla[] ← random(cantFotosBDD) vector aleatorio de cantidad de fotos
    aciertosPorCorte[CantidadCortes, cantFotosElegidas] ← inicializarMatriz
    for corte = 1 to cantidadCortes do
      if tipoCortes = estratificado then
        elegidas ← corteEstratificado(mezcla) selección estratificada
      else
        elegidas ← corteAleatorio(mezcla) selección aleatoria
      NOElegidas ← Subconjunto de entrenamiento
      for f = 1 to Cant. elegidas do
        if cocientes(elegidas(f), NOElegidas) > umbral then
          aciertosPorEspecies(k) ++
          siendo k la especie a la que pertenece la imagen f
        if imagen(f) fue correctamente clasificada then
          aciertosPorCorte(corte, elegidas(f)) ++
        else
          aciertosPorCorte(corte, elegidas(f)) --
      aciertosPorCorrida(corrida, cantidadCortes) ← aciertorPorCorte
  Fin del algoritmo ← Retornar aciertosPorCorrida

```

4.8. Resultados obtenidos

En las tablas 4.2 y 4.3 se detallan los resultados obtenidos de las validaciones cruzadas aleatorias y estratificadas, sobre el modelo de datos construido con SIFT, mientras que las tablas 4.4 y 4.5 muestran los resultados del modelo de datos SURF.

Para obtener un análisis general, las pruebas se realizaron con distintos grupos de ejemplos formados al aplicar el criterio de selección, ya sea aleatorio o estratificado, luego de mezclarlos de manera aleatoria utilizando un generador de números al azar uniforme. Las pruebas se

repetieron 30 veces. A priori, se estima que los resultados de las pruebas estratificadas sean mejores que los de las pruebas aleatorias ya que las primeras se encuentran balanceadas en cuanto a la cantidad de especies. El análisis final que se detalla en las tablas 4.2 y 4.3 muestra que el modelo generado con SIFT es, claramente, más robusto que SURF y como se anticipaba, los subconjuntos estratificados superan claramente a los aleatorios. Con estas evidencias, se puede decir que las particiones generadas son independientes entre sí.

4.9. Conclusiones

En este capítulo se detalló la implementación de la base de datos utilizada en esta tesina. Se presentaron las dificultades que implican estos tipos de desarrollos y se explicaron y justificaron sus soluciones, teniendo en cuenta que no existe hasta el momento, una base de datos de especies de serpientes locales que sirva como punto de referencia. Por otro lado, como se mencionó en la introducción de este capítulo, los únicos modelos de datos públicos son meramente informativos y no pueden utilizarse para la resolución de este problema en particular.

En este caso de estudio, la recopilación de imágenes resulta una tarea compleja debido a los lugares de hallazgo de los distintos ejemplares. Optar por la selección de partes distintivas del cuerpo de las serpientes para realizar el armado de la base de datos y el posterior reconocimiento, favorece a la realización de estas tareas, ya que pensar en utilizar el cuerpo entero de un ejemplar resultaría una tarea casi imposible debido a problemas de oclusión.

Como se indicó en el capítulo 2, se estudiaron serpientes de las provincia de Buenos Aires, en particular las del partido de La Plata. Si bien se reduce considerablemente el número de opciones, es un comienzo para la construcción de una base de datos más robusta para el reconocimiento de especies de serpientes, implicando un desafío la adición de nuevas.

Por otro lado, se trató un tema crítico como la elección del umbral de decisión. Para llevar a cabo estas tareas, no existe un umbral genérico que garantice la clasificación correcta de un ejemplar. Se realizaron pruebas exhaustivas, a partir de un rango de valores candidatos, sobre los descriptores que conforman la base de datos hasta obtener la mejor tasa de acierto. Si bien es una tarea que tiene un costo computacional elevado, se realiza una sola vez y *offline*, por lo que no se ejecutaría al momento de realizar el reconocimiento que se explica en el próximo capítulo.

En cuanto a la elección de los descriptores, se optó por SIFT ya que mostró mejores resultados que SURF. Se detallaron la cantidad de descriptores que ambos generan y cuan significativos son. SIFT sobresale por detectar una mayor cantidad de puntos característicos que permiten una mayor precisión al momento del reconocimiento. Las pruebas de validación cruzada brindan con mayor detalle la exactitud que caracteriza al modelo de datos propuesto, remarcando la

superioridad de SIFT por encima de SURF.

Hasta el momento, se cuenta con una primera versión de una base de datos de especies originarias del partido de La Plata.

En el próximo capítulo, se explica el funcionamiento de la aplicación que utiliza el modelo de datos propuesto como fuente de consulta. Además se realizaron pruebas con imágenes nuevas que simulan una situación en la vida real. Las mismas se detallan con sus resultados y la respuesta final que devuelve el aplicativo.

Capítulo 5

Reconocimiento de Especies

5.1. Introducción

Este capítulo constituye el aporte central de la tesina y tiene por objeto explicar la manera en que se llevó adelante la tarea de reconocimiento de las distintas especies de serpientes consideradas.

Una primera tarea a realizar, que condiciona totalmente los resultados, se relaciona con la base de datos de las especies a reconocer. En el capítulo 4 se explicó la construcción de la misma, donde se detalló las partes que la componen y pruebas que evidencian su robustez. Como se mencionó en el capítulo 2, en esta primera etapa sólo estará formada por las cinco especies más características del partido La Plata. Esta base de datos será utilizada para medir el desempeño de la solución propuesta.

Si bien en la literatura existen soluciones a este problema, su funcionamiento no se basa en una representación vectorial de las serpientes sino que utiliza características específicas de los ejemplares con el fin de aplicar técnicas de Aprendizaje de Máquina tales como árboles de decisión o redes neuronales [Amir et al., 2017]. Estas características son difíciles de medir para quienes no sean especialistas en el tema. La propuesta de esta tesina es simplificar el proceso a través de una representación automática.

5.2. Método Propuesto

Para llevar a cabo el reconocimiento de serpientes, el método propuesto utiliza una base de datos construida a partir de descriptores SIFT de partes de imágenes de distintos ejemplares.

La selección de segmentos de las imágenes y sus descriptores apuntan a reducir el tiempo de comparaciones, así como la detección de falsos positivos.

El reconocimiento de un ejemplar consta de los siguientes pasos:

- Cargar una imagen de entrada de la cual el usuario podrá elegir entre recortar 1 o 2 partes que considere significativas.
- Calcular los vectores SIFT correspondientes al fragmento obtenido en el paso previo.
- Comparar cada vector obtenido en el punto anterior con los vectores de la BDD, acumulando las coincidencias por ejemplar.
- La respuesta final se corresponderá con aquel ejemplar que haya acumulado la mayor cantidad de coincidencias.

Cabe destacar que la comparación entre el conjunto de vectores SIFT correspondientes a la imagen de entrada y los almacenados en la base de datos para cada parte de cada ejemplar es una tarea que puede ser paralelizada muy fácilmente. Esto se puede conseguir almacenando en cada procesador los vectores correspondientes a cada especie de las estudiadas, y una vez finalizada la comparación con el conjunto de vectores de la imagen de entrada, sólo debe retornarse el total de coincidencias encontradas. Esto minimiza considerablemente el tiempo de comunicación entre procesos.

5.3. Resultados Obtenidos

Las mediciones realizadas se basan en la clasificación de dos ejemplares de cada una de las cinco especies más populares en el partido de La Plata y Gran La Plata. Como se mencionó en el capítulo 2, las especies analizadas son las siguientes: *Bothrops Alternatus*, *Paraphimophis Rusticus*, *Philodryas Patagoniensis*, *Tomodon Ocellatus* y *Xenodon Dorbignyi* que se detallan en la figura 2.1.

Cada imagen a clasificar debe ser transformada en primer lugar a sus descriptores SIFT correspondientes. Este proceso fue explicado con detalle en el capítulo 3. A continuación, se calcula la distancia euclídea del descriptor de la imagen de entrada contra cada descriptor de la base. Aquellos valores que estén por encima del umbral que se determinó en el capítulo 4 sumarán puntos al ejemplar correspondiente, mientras que el resto serán descartados.

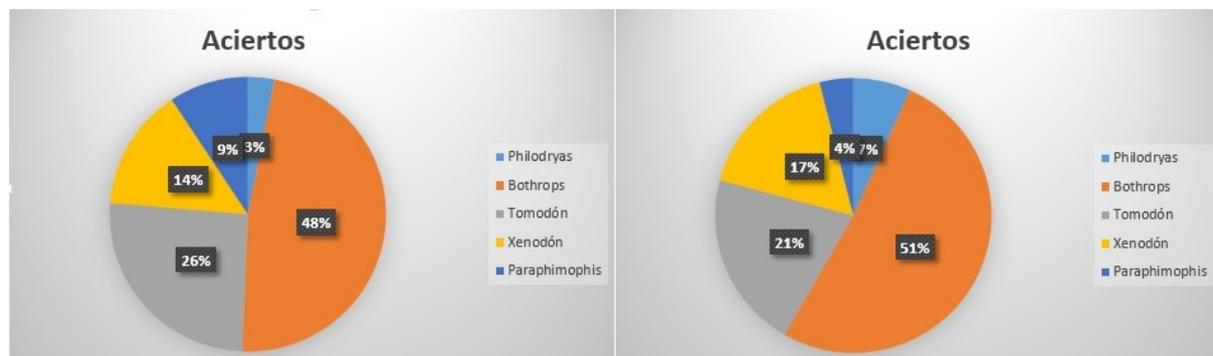


Figura 5.1: Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie *Bothrops Alternatus*.

El algoritmo 6 resume lo explicado

Algoritmo 6: Pseudocódigo del proceso de clasificación

descIMGEntrada ← calcular descriptores imagen entrada

U ← umbral determinado en el capítulo 4

for *i* = 1 to Cant. de imágenes de la BDD **do**

especieActual ← nro. de especie a la que pertenece la imagen *i*

IE = imagen de entrada

for *j* = 1 to Cant. de descriptores de la *IE* **do**

 Calcular distancias entre descriptor *j*-ésimo de la *IE* y cada descriptor de la imagen *i*

 Cociente = (d_2/d_1) siendo d_1 y d_2 las distancias mínimas tales que $d_1 < d_2$

 Si (Cociente > *U*)

 aciertos(*especieActual*) ++

Fin del for ← Retornar aciertos

A continuación se detallan resultados de pruebas realizadas con imágenes que no forman parte de la BDD. Dependiendo del contexto en que las fotos fueron tomadas, se optó por realizar un corte donde la imagen presenta una buena visibilidad del ejemplar, y dos cortes donde se consideró que un único corte no era lo suficientemente representativo. Para las pruebas se utilizó un umbral de 1,4080.

Los resultados confirman la robustez del modelo. Las imágenes 5.1, 5.2, 5.3, 5.4 y 5.5 muestran una clara diferencia de la especie ganadora con respecto al resto. Cabe destacar que las especies *Bothrops Alternatus*, *Tomodon Ocellatus* y *Xenodon Dorbignyi*, presentan dibujos en el cuerpo, que dependiendo del punto de vista pueden presentar similitudes. Las tablas 5.1 y 5.2 resumen los resultados detallados.

Es importante tener en cuenta la precisión de la respuesta obtenida. Esto se relaciona con la similitud que cada descriptor SIFT de la imagen a clasificar tiene con los descriptores almacenados en la base. Es conveniente que exista una marcada diferencia entre la imagen

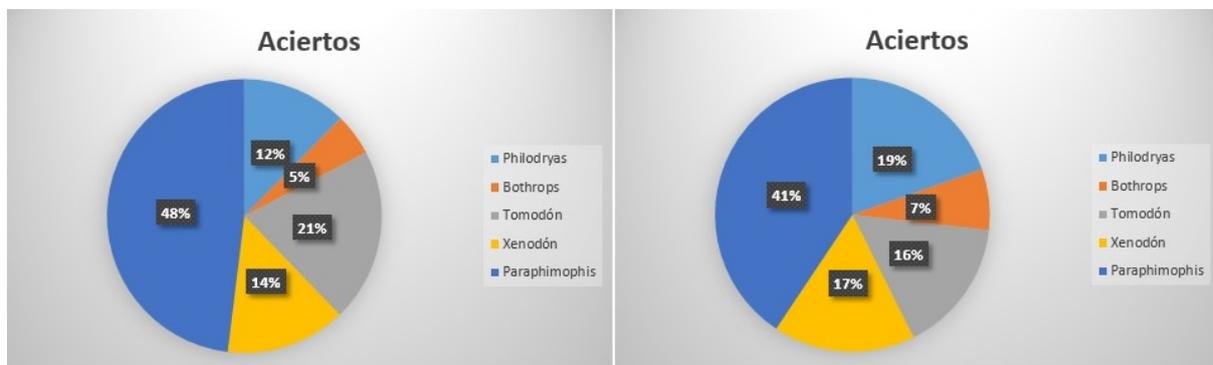


Figura 5.2: Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie *Paraphimophis Rusticus*.

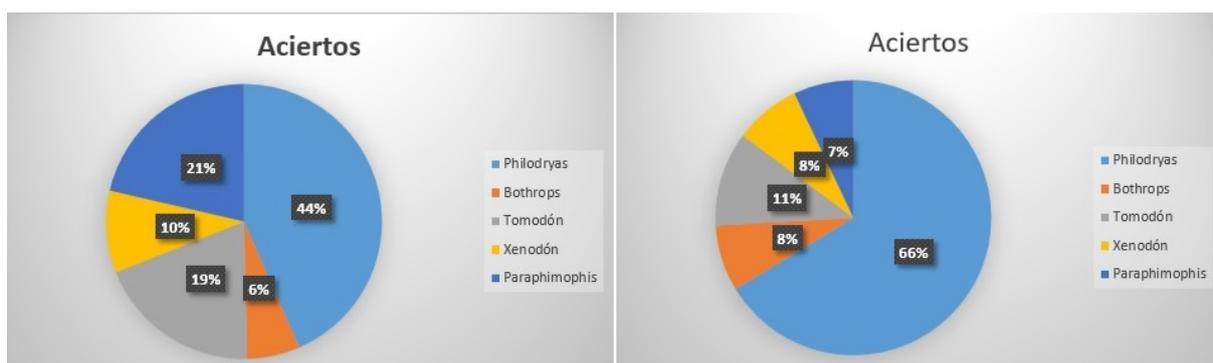


Figura 5.3: Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie *Philodryas Patagoniensis*.

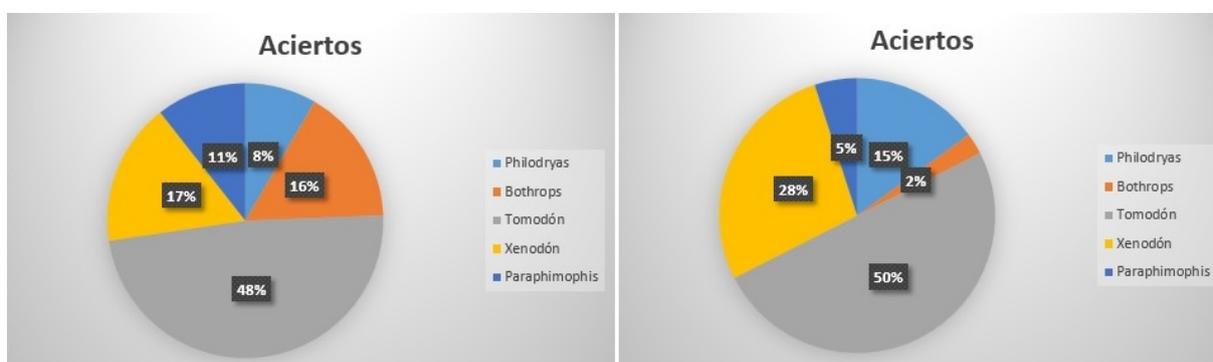


Figura 5.4: Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie *Tomodon Ocellatus*.

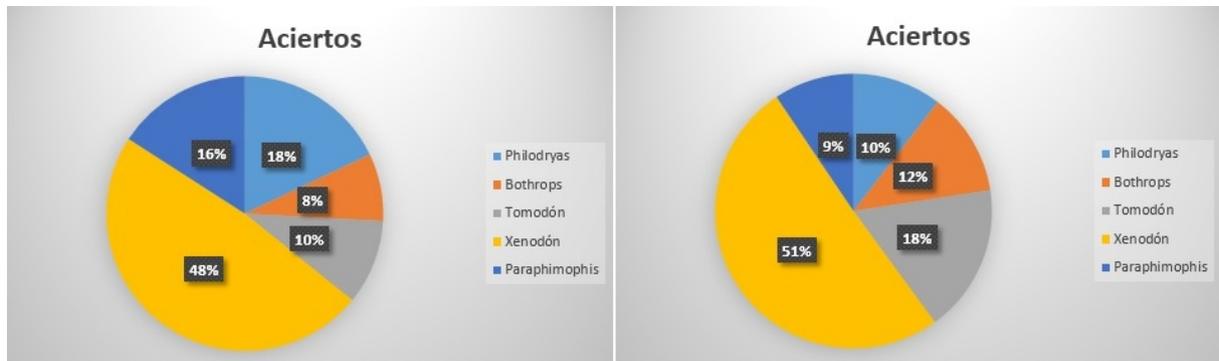


Figura 5.5: Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie *Xenodon Dorbignyi*.

Tabla 5.1: Resumen de aciertos de la primera prueba realizada sobre el modelo con imágenes nuevas.

ESPECIE	Bothrops Alternatus	Paraphimophis Rusticus	Philodryas Patagoniensis	Tomodon Ocellatus	Xenodon Dorbignyi
Bothrops A.	48 %	3 %	9 %	14 %	26 %
Paraphimophis R.	5 %	48 %	12 %	21 %	14 %
Philodryas P.	6 %	21 %	44 %	19 %	10 %
Tomodon O.	16 %	11 %	8 %	48 %	17 %
Xenodon D.	8 %	16 %	18 %	10 %	48 %

Tabla 5.2: Resumen de aciertos de la segunda prueba realizada sobre el modelo con imágenes nuevas.

ESPECIE	Bothrops Alternatus	Paraphimophis Rusticus	Philodryas Patagoniensis	Tomodon Ocellatus	Xenodon Dorbignyi
Bothrops A.	51 %	4 %	7 %	21 %	17 %
Paraphimophis R.	7 %	41 %	19 %	16 %	17 %
Philodryas P.	8 %	7 %	66 %	11 %	8 %
Tomodon O.	2 %	5 %	15 %	50 %	28 %
Xenodon D.	12 %	9 %	10 %	18 %	51 %

candidata con respecto al resto para poder afirmar con certeza que corresponde a una especie determinada. Se puede observar que en la mayoría de los casos, la especie ganadora obtiene al menos el doble de aciertos que la segunda.

5.4. Conclusiones

Este capítulo describió con detalle la tarea realizada para lograr el reconocimiento esperado. Su desarrollo implicó, en primer lugar, el estudio y análisis del método SIFT como herramienta para representar los puntos de interés presentes en la imagen para la construcción de la base de datos y el posterior reconocimiento realizado en este capítulo.

Los resultados muestran que es factible utilizar esta representación para alcanzar el objetivo planteado.

Las mediciones pueden completarse si se amplía la base de datos. Esto es un tema que debe analizarse con cuidado ya que un excesivo número de ejemplares reduce las diferencias entre ellos lo cual redundaría en una menor tasa de acierto a la hora de efectuar el reconocimiento.

Capítulo 6

Conclusiones Generales

6.1. Conclusiones

Son numerosos los contextos que pueden verse beneficiados con el reconocimiento de objetos en imágenes. En esta tesina el foco estuvo puesto en el reconocimiento de las distintas especies de serpientes que se encuentran habitualmente en el partido de La Plata.

En el capítulo 1 se realizó una introducción general al procesamiento de imágenes donde se detalló sus aplicaciones en diversas ciencias y los procesos que lo componen. Finalmente, se explicó la estructura de esta tesina y como está distribuida.

En el capítulo 2 se hizo una descripción de la distribución de los distintos ejemplares llegando a la conclusión de que la diversidad de especies en nuestra ciudad es muy inferior a la que suele verse en el norte de la provincia. Esto facilita el desarrollo de la aplicación a través de una drástica reducción del espacio de búsqueda.

En el capítulo 3, se explicó con detalle la forma de calcular descriptores de características que posteriormente, conformarán la base de datos. Para la obtención de los vectores, se utilizaron los métodos de generación de descriptores SIFT y SURF. Las pruebas realizadas en el capítulo 4, ponen en evidencia que, para este caso en particular, SIFT se comporta mejor que SURF. Si bien ambos son muy utilizados en la literatura, no existe un algoritmo genérico que se adapte a un problema en particular.

Las imágenes, si bien fueron captadas por distintas personas, tienen en común la presencia de un único ejemplar en su interior. Pese a esto, reconocer la especie en cuestión no es una tarea sencilla. La parte de la imagen habitualmente denominada “fondo”, es decir, lo que rodea al animal, es sumamente variada y difícil de estandarizar requiriendo un análisis pormenorizado y específico.

Desde el punto de vista del procesamiento de la imagen, son varios los aspectos a tener en

cuenta a la hora de detectar la parte “útil”. Por este motivo, para el armado de la base de datos se optaron por utilizar fracciones del cuerpo distintivas de cada especie. En el capítulo 4, se detallan y justifican la selección de las partes que conforman el modelo de datos y la manera en que pueden compararse las imágenes correspondientes a dos objetos distintos. La elección de cuál de ellas utilizar dependerá de las características de la información de entrada. Inicialmente se habían tomado ejemplares completos pero luego de realizar varias pruebas se concluyó que era mejor almacenar únicamente las partes más representativas. Es poco probable que todo el cuerpo del ejemplar se encuentre expuesto en la imagen; en su lugar, las serpientes suelen arrollarse provocando que en muchas ocasiones la cabeza y/o cola no se visualicen.

Como resultado de esta etapa se dispondrá de la parte correspondiente al objeto de interés y por lo tanto, puede iniciarse el reconocimiento. Esta tarea consistió en un cambio de la representación pasando de los píxeles a vectores numéricos de longitud fija correspondientes a los puntos de interés. Dichos puntos se encuentran ubicados en zonas de la imagen con textura característica. Cambios en la iluminación y/o bordes son fácilmente detectados con este tipo de representación.

Una vez obtenido el conjunto de vectores correspondiente al ejemplar a reconocer debe efectuarse un análisis comparativo con las distintas muestras presentes en la base de datos. Aquí cobra importancia el armado de la base ya que los ejemplares que contiene deben ser lo suficientemente representativos del problema. En el capítulo 5, los objetos a comparar fueron nuevos ejemplares de serpiente y las distintas mediciones obtenidas mostraron la efectividad de la solución propuesta.

Es importante destacar que estos resultados guardan una estrecha relación con el armado de la base. En el capítulo 4 se analizó el tamaño adecuado de muestra a incluir con el objeto de alcanzar una tasa de acierto adecuada.

Por todo lo antes dicho, se considera que los objetivos originalmente planteados han sido alcanzados. Sin embargo, el problema aún presenta desafíos no resueltos que se explican a continuación.

6.2. Trabajos Futuros

Existen aspectos que caracterizan a algunas especies y que no fueron contemplados en este análisis. Uno de estos aspectos es el color. Algunas especies presentan colores llamativos en su vientre o en su lomo. Por ejemplo, la especie *Lygophis (Liophis) Anomalus* posee una línea roja que atraviesa todo el costado de su cuerpo. Esta particularidad que en un principio parece sumamente útil no siempre puede ser utilizada porque depende del ángulo desde el cual haya sido captada la imagen. Sin embargo, es un aspecto a tener en cuenta y seguramente en varias

situaciones puede resultar de interés. Algunas alternativas a considerar pueden analizarse a partir del siguiente trabajo [James, 2017].

Otro inconveniente detectado tiene que ver con la posibilidad de que la imagen ingresada no corresponda a ninguno de los ejemplares almacenados en la base de datos. La solución propuesta no dispone de una clase de rechazo. Esto implica que ante cualquier imagen de entrada indicará el ejemplar más parecidos de los registrados en la base. Sólo el uso de un umbral puede impedir que se brinde una respuesta errónea. El inconveniente es precisamente determinar el valor de este umbral ya que si es demasiado bajo no impedirá que se produzcan errores y si es sumamente alto no dejará que se informe el reconocimiento. Este tema aún está en estudio.

Finalmente, es preciso completar la base de datos a fin de poder cubrir todas las especies de la provincia. Esto no sólo es una tarea ardua por la cantidad de información que se requiere también reduce la distancia entre las opciones. Esto último seguramente reducirá la tasa de acierto e incrementará la necesidad de definir adecuadamente una tasa de acierto asociada a cada respuesta dada.

Índice de figuras

1.1.	Detalle de los procesos que componen el procesamiento de imágenes	6
1.2.	(A) Imagen original (B) Detección de bordes (C) Segmentación semántica (D) Detección del objeto de interés	7
2.1.	(A) <i>Bothrops Alternatus</i> (B) <i>Paraphimophis Rusticus</i> (C) <i>Philodryas Patagioniensis</i> (D) <i>Tomodon Ocellatus</i> (E) <i>Xenodon Dorbignyi</i>	13
2.2.	Vértebras de una serpiente. ([Giambelluca, 2015], página 12 - reimpresso con permiso)	14
2.3.	Escamas ventrales de una serpiente. ([Giambelluca, 2015], página 12 - reimpresso con permiso)	14
2.4.	Escamas ventrales de una serpiente. ([Giambelluca, 2015], página 13 - reimpresso con permiso)	15
2.5.	Textura apergaminada de un huevo de un ofidio.[Ark, 2016]	16
2.6.	Huevos apergaminados. [Vig, 2013]	16
2.7.	Víbora ratonera (<i>Pantherophis obsoletus</i>) a punto de mudar de piel, momento en que la escama ocular se ve opaca. [War, 2016]	17
2.8.	Momento exacto en que el ofidio cambia su piel. Se puede apreciar el brillo en su color. [Vel, 2008]	18
2.9.	Morfología de los dientes de distintas especies de serpientes. ([Giambelluca, 2015], página 13 - reimpresso con permiso)	19
3.1.	Resultado de la convolución de una imagen con una función gaussiana para distintos valores de su desviación σ . (A) Imagen original. (B) $\sigma = 1$. (C) $\sigma = 2$. (D) $\sigma = 4$. (E) $\sigma = 8$. (F) $\sigma = 16$	25
3.2.	A izquierda, se representa el espacio escala de dos octavas. Cada una de ellas tiene 5 escalas. A la derecha, se obtiene el espacio escala de las diferencias de gaussianas.	26

3.3.	Los círculos verdes representan los píxeles vecinos de un punto de interés, que son analizados para determinar si el pixel central es un máximo o mínimo local.	27
3.4.	(a) Gradientes de la imagen. (b) Descriptor del punto clave	29
4.1.	A la izquierda, la cabeza de una Yará. Se aprecia como el dibujo de su cuerpo permanece continuo. A la derecha, la cabeza de una Culebra con sus placas distintivas.	36
4.2.	Árbol de decisión donde se detallan características de cada especie	37
4.3.	En las figuras A y C, se puede observar la piel de los ofidios con colores más vívidos debido a su contacto con el agua. En las figuras B y D, se perciben los colores más apagados debido a su contacto con la tierra y el polvo proveniente de las rocas, además se encuentran enrollados en ambos casos.	39
4.4.	Imágenes de distintos recortes de cuerpos de serpientes. En cada imagen, se aprecian los patrones que caracterizan a cada ofidio	40
4.5.	Mejor tasa de acierto obtenida con SIFT(91 %) se detectó con un umbral de 1.4080	43
4.6.	Mejor tasa de acierto obtenida con SURF(82 %) se detectó con umbral de 1.4010	44
4.7.	Etapas de la validación cruzada de un modelo de datos. En este caso $k = 3$ [Lang, 2010].	46
4.8.	Etapas de la validación cruzada dejando uno fuera sobre 10 muestras. En cada iteración la muestra de color verde se utiliza como testeo, mientras que las restantes se usan como entrenamiento.	47
4.9.	Validación cruzada de K iteraciones con $K = 10$. En cada iteración la muestra de color verde se usa como testeo. Para armar el subconjunto de prueba se utiliza el 10 % de la muestra total. Con respecto a las pruebas aleatorias y estratificadas, el procedimiento es el mismo, únicamente varía la cantidad de especies de serpientes elegidas en cada muestra de prueba.	48
5.1.	Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie <i>Bothrops Alternatus</i>	58
5.2.	Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie <i>Paraphimophis Rusticus</i>	59
5.3.	Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie <i>Philodryas Patagoniensis</i>	59
5.4.	Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie <i>Tomodon Ocellatus</i>	59

5.5. Aciertos obtenidos al realizar pruebas con dos imágenes distintas de la especie
Xenodon Dorbignyi. 60

Lista de Algoritmos

1.	Pseudocódigo del método para seleccionar los vectores SIFT que conforman la base de datos	41
2.	Pseudocódigo del cálculo de las distancias euclídeas y generación de cocientes. .	42
3.	Pseudocódigo del calculo del mejor umbral junto con su tasa de acierto	43
4.	Pseudocódigo del proceso para los aciertos por cada vector de cada descriptor . .	45
5.	Pseudocódigo de validación cruzada sobre el modelo de datos	49
6.	Pseudocódigo del proceso de clasificación	58

Indice de tablas

1.1. Entradas y salidas de los distintos niveles en los que pueden agruparse los métodos del procesamiento digital de imágenes	8
4.1.	45
4.2. Aciertos obtenidos realizando una validación cruzada aleatoria sobre el modelo SIFT de datos propuesto	50
4.3. Aciertos obtenidos realizando una validación cruzada estratificada sobre el modelo SIFT de datos propuesto	51
4.4. Aciertos obtenidos realizando una validación cruzada aleatoria sobre el modelo SURF	52
4.5. Aciertos obtenidos realizando una validación cruzada estratificada sobre el modelo SURF	53
5.1. Resumen de aciertos de la primer prueba realizada sobre el modelo con imágenes nuevas.	60
5.2. Resumen de aciertos de la segunda prueba realizada sobre el modelo con imágenes nuevas.	60

Bibliografía

- [Vel, 2008] (2008). <https://www.mascotaplanet.com/blog/por-que-cambian-la-piel-los-reptiles-n167>.
- [Vig, 2013] (2013). http://www.moroccoherps.com/ficha/Daboia_mauritanica/.
- [Ark, 2016] (2016). <https://www.scoopnest.com/es/s/Arkaprava%20Ghosh/>.
- [War, 2016] (2016). <https://allyouneedisbiology.wordpress.com/2016/03/06/snake-senses/>.
- [Abdel-Hakim and Farag, 2006] Abdel-Hakim, A. E. and Farag, A. A. (2006). Csift: A sift descriptor with color invariant characteristics. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1978–1983.
- [Ahonen et al., 2006] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041.
- [Amir et al., 2017] Amir, A., Zahri, N. A. H., Yaakob, N., and Ahmad, R. B. (2017). *Image Classification for Snake Species Using Machine Learning Techniques*, pages 52–59. Springer International Publishing, Cham.
- [Arlot and Celisse, 2010] Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4:40–79.
- [Baumann, 2003] Baumann, K. (2003). Cross-validation as the objective function for variable-selection techniques. *TrAC Trends in Analytical Chemistry*, 22(6):395 – 406.
- [Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346 – 359. Similarity Matching in Computer Vision and Multimedia.
- [Bay et al., 2006] Bay, H., Fasel, B., and Gool, L. V. (2006). Gool. interactive museum guide: Fast and robust recognition of museum objects. In *In Proc. Int. Workshop on Mobile Vision*.

- [Bouckaert, 2008] Bouckaert, R. R. (2008). Practical bias variance decomposition. In Wobcke, W. and Zhang, M., editors, *AI 2008: Advances in Artificial Intelligence*, pages 247–257, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Brown and Lowe, 2002] Brown, M. and Lowe, D. (2002). Invariant features from interest point groups. In *In British Machine Vision Conference*, pages 656–665.
- [Cohen and Medioni, 1998] Cohen, I. and Medioni, G. (1998). Detecting and tracking moving objects in video from an airborne observer. In *In 1998 Image Understanding Workshop*, pages 217–222.
- [Delgado, 2015] Delgado, D. (2015). Automatic coin and bill detection.
- [Elkan, 2012] Elkan, C. (2012). Evaluating classifiers.
- [Forsyth and Ponce, 2012] Forsyth, D. A. and Ponce, J. (2012). *Computer Vision A Modern Approach*. Pearson, Upper Saddle River, NJ, USA.
- [Giambelluca, 2015] Giambelluca, L. (2015). *Serpientes bonaerenses*. EDULP, La Plata. Bs.As. Argentina.
- [Gill and Sharma, 2017] Gill, N. K. and Sharma, A. (2017). Vehicle detection from satellite images in digital image processing. *International Journal of Computational Intelligence Research*, 13:697–705.
- [Gonzalez and Woods, 2006] Gonzalez, R. C. and Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- [Harris and Stephens, 1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *In Proc. of Fourth Alvey Vision Conference*, pages 147–151.
- [Hu, 1962] Hu, M.-K. (1962). Visual pattern recognition by moment invariants, computer methods in image analysis. *IRE Transactions on Information Theory*, 8.
- [James, 2017] James, A. (2017). Snake classification from images. PeerJ Preprints 5:e2867v1. <https://doi.org/10.7287/peerj.preprints.2867v1>.
- [Janwale, 2015] Janwale, A. (2015). Digital image processing applications in agriculture: A survey. 5:622.
- [Ke and Sukthankar, 2004] Ke, Y. and Sukthankar, R. (2004). Pca-sift: a more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–506–II–513 Vol.2.

- [Koenderink, 1984] Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, 50:363–370.
- [K.V.Bhuvaneshwari, 2017] K.V.Bhuvaneshwari, A.Abirami, N. (2017). Face recognition using pca. *International Journal Of Engineering And Computer Science*, 6.
- [Lang, 2010] Lang, J.-P. (2010). Predictors tutorial.
- [Lanzarini et al., 2013] Lanzarini, L., Ronchetti, F., Estrebou, C., Lens, L., and Bariviera, A. F. (2013). Face recognition based on fuzzy probabilistic som. In *2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pages 310–314.
- [Lanzarini et al., 2016] Lanzarini, L., Ronchetti, F., Estrebou, C., and Quiroga, F. (2016). Handshape recognition for argentinian sign language using probsom. *Journal of Computer Science and Tecnology*, 16:1–5.
- [Lazebnik et al., 2005] Lazebnik, S., Schmid, C., and Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278.
- [Lindeberg, 1994] Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA.
- [Liu et al., 2008] Liu, D., w. Sun, J., Wei, G., and Liu, X. (2008). Rbf neural networks and cross validation-based signal reconstruction for nonlinear multi-functional sensor. In *2008 9th International Conference on Signal Processing*, pages 1512–1515.
- [Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2, ICCV '99*, pages 1150–, Washington, DC, USA. IEEE Computer Society.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110.
- [Matas et al., 2002] Matas, J., Chum, O., Urban, M., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. *British Machine Vision Conference*, pages 384–396.
- [Mikolajczyk and Schmid, 2004] Mikolajczyk, K. and Schmid, C. (2004). Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86.
- [Mislav Grgic, 2007] Mislav Grgic, K. D. (2007). Face recognition homepage. <http://www.face-rec.org/databases/>.

- [Mundy, 2006] Mundy, J. L. (2006). *Object Recognition in the Geometric Era: A Retrospective*, pages 3–28. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Patel et al., 2013] Patel, C., Shah, D., and Patel, A. (2013). Automatic number plate recognition system (anpr): A survey. *International Journal of Computer Applications*, 69(9):21–33. Full text available.
- [Reeves, 1992] Reeves, S. J. (1992). A cross-validation framework for solving image restoration problems. *Journal of Visual Communication and Image Representation*, 3(4):433 – 445.
- [Roly et al., 2015] Roly, Z. Y., Hakim, M. A., Shahriar Zahan, A., Monzur Hossain, M., and Reza, M. (2015). Isob: A database of indigenous snake species of bangladesh with respective known venom composition. Roly et al. *Bioinformatics* 11 (2): 107-114 (2015):107 – 114.
- [Santillán, 2008] Santillán, I. D. G. (2008). *Visión Artificial y Procesamiento Digital de Imágenes Usando MatLab*. Pontificia Unviersidad Católica del Ecu.
- [Simard et al., 1999] Simard, P. Y., Bottou, L., Haffner, P., and LeCun, Y. (1999). Boxlets: A fast convolution algorithm for signal processing and neural networks.
- [Uetz and Etzold, 1996] Uetz, P. and Etzold, T. (1996). The embl/ebi reptile database. 27:174–175.
- [Vedaldi and Fulkerson, 2008] Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>.
- [Viola and Jones, 2001] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. 1.
- [Witkin, 1983] Witkin, A. P. (1983). Scale-space filtering. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'83*, pages 1019–1022, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- [Zhang et al., 2013] Zhang, R., Zhou, W., Li, Y., Yu, N., and Xie, Y. (2013). Nonrigid registration of lung ct images based on tissue features. 2013:834192.