Instituto Tecnologico de Costa Rica

Universidad Nacional

Universidad Estatal a Distancia

Doctorado en Ciencias Naturales para el Desarrollo

Modelo evolutivo de forma para el ajuste de organismos vermiformes en imagenes digitales

Tesis sometida a consideracion del tribunal evaluador como requisito para optar el grado de Doctor en Ciencias Naturales para el Desarrollo, con enfasis en Tecnolog as Electronicas Aplicadas

Jorge Arroyo Hernandez

Costa Rica, 12 de abril de 2021

Declaro que el presente documento de tesis ha sido realizado enteramente por mi perso-na, utilizando y aplicando literatura referente al tema e introduciendo conocimientos y resultados experimentales propios.

En los casos en que he utilizado bibliograf a he procedido a indicar las fuentes mediante las respectivas citas bibliogra cas. En consecuencia, asumo la responsabilidad total por el trabajo de tesis realizado y por el contenido del presente documento.

Jorge Arroyo Hernandez

Heredia, 12 de abril de 2021

Ced: 1-0986-0691

DOCINADE

Tesis de Doctorado

Tribunal Evaluador

Trabajo nal de graduacion presentado como requisito parcial para optar por el grado de Doctor en Ciencias Naturales para el Desarrollo, con enfasis en Tecnolog as Electronicas Aplicadas.

Miembros del Tribunal

Dr. Carlos Travieso Gonzalez
Profesor Lector

Dr. Juan Luis Crespo Mari~no
Profesor Lector

Dr.-Ing. Pablo Alvarado Moya
Profesor Asesor

Los miembros de este Tribunal dan fe de que la presente tesis de doctorado ha sido aprobada y cumple con las normas establecidas por Doctorado en Ciencias Naturales para el Desarrollo, con enfasis en Tecnolog as Electronicas Aplicadas.

Costa Rica, 12 de abril de 2021

Instituto Tecnológico de Costa Rica

Universidad Nacional de Costa Rica

Universidad Estatal a Distancia

Modelo evolutivo de forma para el ajuste de organismos vermiformes en imágenes digitales

Trabajo sometido a consideración del Tribunal Evaluador como requisito para optar por el grado de Doctorado en Ciencias Naturales para el Desarrollo con énfasis en Tecnologías Electrónicas Aplicadas

Jorge Arroyo Hernández
Sustentable

Aprobado por los miembros del tribunal examinador:

Tomás Guzmán Hernández, Ph.D.
Representante del Sistema de la Unidad de Posgrado

TOMAS DE JESUS GUZMAN HERNANDEZ (FIRMA)
Firmado digitalmente por TOMAS DE JESUS GUZMAN HERNANDEZ (FIRMA)
Fecha: 2021.12.07 07:51:08 -06'00'

Giovanni Sáez Arce, Ph.D.
Coordinador General del DOCINADE

Giovanni Sáenz-Arce
2021.12.07
17:07:59 Z

Pablo Alvarado Moya, Ph.D.-Ing.
Director de Tesis

JOSE PABLO ALVARADO MOYA (FIRMA)
PERSONA FISICA, CPF-01-0753-0713.
Fecha declarada: 07/12/2021 02:32:48 PM
Razón: Acta Jorge Arroyo Hdz
Lugar: Cartago

Juan Luis Crespo Mariño, Ph.D.-Ing.
Asesor de Tesis

JUAN LUIS CRESPO MARIÑO (FIRMA)
Firmado digitalmente por JUAN LUIS CRESPO MARIÑO (FIRMA)
Fecha: 2021.12.08 08:34:29 -06'00'

Carlos Travieso González, Ph.D.-Ing.
Asesor de Tesis

TRAVIESO GONZALEZ CARLOS MANUEL - 43761645M
Firmado digitalmente por TRAVIESO GONZALEZ CARLOS MANUEL - 43761645M
Fecha: 2021.12.09 08:59:27 Z

Resumen

Los nematodos han ganado interes por su funcion de agentes de control biologico, as como por el efecto directo en los agro-ecosistemas tales como sembrad os y sistemas de produccion animal, y en consecuencia, el impacto economico a nivel pa s.

Debido a su importancia, en el presente trabajo se desarrolla un modelo evolutivo de forma basado en subdominios permitidos de siluetas vermes expresados en hitos de frontera que ajustan la silueta de nematodo presente en una imagen digital.

El modelo desarrollado trabaja sobre subdominios de forma permitidos en conjuncion con el metodo de optimizacion enjambre de part culas multiobjetivo. A traves de este, se evalua iterativamente la bordicidad y el tama~no del nematodo hasta que se logre hallar el mejor ajuste de la forma verme en la imagen digital.

Asimismo, el modelo se apoya en dos procesos que asisten a la adaptacion de la forma en el espacio de busqueda: caminatas aleatorias entre subdomnios de formas vermes validas y deformaciones a traves de los metodos serpenteo y forrajeo. Con el algoritmo k medias se construyen subdomnios de forma permitidas.

El modelo matematico de analisis de formas se nutre de dos bases de datos: una de imagenes digitales de nematodos y, la otra un conjunto de formas de nematodos expresados con hitos de frontera y almacenados vectorialmente. Esta ultima, es normalizada en cuanto a la cantidad de hitos as como la distancia entre estos. Ademas, cada instancia vectorial es rotada, escalada y alineada a una en particular previamente de nida.

Palabras clave: ajuste de forma, vermiformes, optimizacion, n-s mplexes, serpenteo, forrajeo, segmentacion.

Abstract

Nematodes have gained interest because of their role as biological control agents, as well as for their direct e ect on agro-ecosystems such as crops and animal production systems, and consequently, their economic impact at the national level.

Due to its importance, in the present work, an evolutionary shape model is developed based on allowed subdomains of verm silhouettes expressed in boundary landmarks that adjust the nematode silhouette present in a digital image.

The developed model works on shape-allowed subdomains in conjunction with the multi-objective particle swarm optimization method. Through this, the bordicity and size of the nematode is iteratively evaluated until the best t of the shape seen in the digital image is found.

The model also relies on two processes that assist in the adaptation of the shape in the search space. of the shape in the search space: random walks between subdomains of valid verme shapes and deformations through the snaking and foraging methods. With the k means algorithm, allowable shape subdomains are constructed.

The mathematical shape analysis model is fed by two databases: one of digital images of nematodes and, the other a set of nematode shapes expressed with boundary landmarks and vectorially stored. The latter is normalized in terms of the number of landmarks as well as the distance between them. Also, each vector instance is rotated, scaled, and aligned to a particular one previously de ned.

Keywords: shape t, vermiform, optimization, n-simplex, snaking, meandering, segmentation.

A mi padre, madre y esposa. Gracias eternas ...

Agradecimientos

En primera instancia, darle gracias a Dios y a la Virgencita de los Angeles, quienes han sido acompa~nantes en este largo proceso de crecimiento personal y profesional.

Al Dr.-Ing. Pablo Alvarado-Moya, de quien estar eternamente agradecido por la extraordinaria asesor a brindada durante todo el proceso de investigacion, y sobre todo, por su enorme disposicion en cada momento que le necesit <Gracias Pablo!

A los profesores Dr. Juan Luis Crespo-Mari~no y Dr. Carlos Manuel Travieso-Gonzalez por toda la colaboracion y consejos brindados.

Finalmente a mis padres Jorge Luis y Mar a Eugenia, y a mi esposa Mar a Fernanda por ser los bastiones en mi vida.

Jorge Arroyo Hernandez

Costa Rica, 12 de abril de 2021

Indice general

Indice de guras

Indice de tablas

Cap tulo 1

Introduccion

En la actualidad, el mundo avanza hacia una econom a global y la agroeconom a es uno de los ejes centrales en pa ses como Costa Rica que depende, en gran parte, de este sector tal y como indica Mora (2019). Los indicadores macroeconomicos mostrados en este estudio, indican que solo las ventas al exterior, de los bienes de cobertura agropecuaria en el a~no 2019, alcanzaron un monto de 3665.4 millones de dolares, cifra que representa 42.6 % del total exportado por el pa s.

La supervivencia de esta actividad depende de la mejora continua y aplicacion de novedo-sas estrategias de produccion, que permitan alcanzar e innovar los estandares de calidad y cantidad demandados por los mercados locales e internacionales (Pratt y Rivera, 2003).

Debido a lo mencionado anteriormente, las nuevas estrategias deben ser decididas sobre una base de informacion de los agroecosistemas provenientes de un sumario actualizado de resultados y evidencias cient cas durante el quehacer de sus actividades.

Su n debe centrarse en el analisis y caracterizacion de los factores que intervienen en los procesos de produccion, y con ello, la mejor a de las habilidades sobre las acciones de intervencion necesarias a favor de la mitigacion y prevencion de posibles efectos ne-gativos a corto, mediano y largo plazo que puedan inuir en la calidad y cantidad de la productividad agr cola.

Por tanto, se hace menester la incorporacion de nuevas estrategias y tomas de decisiones respaldadas a partir de evidencia cient ca, proveniente de la aplicacion de modelos o pa-

trones matematicos y estad sticos sobre los datos extra dos, que permitan la optimizacion de los recursos y tiempos en la cadena productiva, en conjuncion a los requerimientos y exigencias en competitividad a nivel mundial (El as, 2014; Suprem et al., 2013).

Actualmente, la inclusion de los metodos de aprendizaje de maquina transforman tecnicas manuales de laboratorio como la deteccion, conteo y clasi cacion a procesos automatizados para agilizar labores propias de los especialistas, ademas, ofrece una opcion de mejoras del rendimiento usando modelos predictivos, de clasi cacion y toma de decisiones (Geron, 2019).

Esto permite abrir nuevas posibilidades de innovacion en los agroprocesos, conservando un balance positivo de crecimiento de la cobertura agropecuaria segun Mora (2019). Ejemplos

de tareas, en las cuales hay una aplicacion de metodos de aprendizaje de maquina y reconocimiento en los agroecosistemas para la extraccion de informacion relevante se describen en la tabla 1.1.

Tabla 1.1: Ejemplos de aplicaciones de aprendizaje de maquina en la agricultura segun Kamilaris y Prenafeta-Boldu (2018).

Aplicacion	Descripcion
Clasi cacion	Enfermedades de hojas de los cultivos Tipos de cobertura terrestre Tipos de cultivo Usos de suelo
Reconocimiento	Fenolog a de los estados de las plantas Especies de cultivos Tipos de semillas
Deteccion	Enfermedades de hojas de los cultivos Enfermedades de los cultivos
Segmentacion	Ra ces de los cultivos respecto al suelo
Estimaciones	Tama~no de los campos de cultivo Mapeo de la cobertura de vegetacion en invierno
Conteo	Unidades de cultivos
Prediccion	Enfermedades de hojas de los cultivos Humedad de los suelos Crecimiento de animales y cultivos Condiciones atmosfericas

Es por lo tanto necesario potenciar la investigacion y el desarrollo de nuevas herramientas

tecnologicas que apoyen la actividad agropecuaria en cada una de las actividades que intervienen en el proceso productivo, que complementen procesos de mejora en la calidad y produccion en los cultivos, as como la deteccion temprana de enfermedades y pestes, disminucion del impacto ambiental y sus respectivas intervenciones (Liakos et al., 2018).

En los procesos de innovacion basada en la actividad cient ca, en particular la referen-te a la agropecuaria, es trascendental la calidad y la consistencia de los datos con los que se va a producir conocimiento. En este sentido, una de las fuentes de extraccion de informacion es la microora y microfauna de los suelos (Bongers y Ferris, 1999). Estos los son bioindicadores capaces de proporcionar datos acerca de la perturbacion de los suelos y sedimentos acuaticos, su ciclo de descomposicion y regeneracion, as como de la calidad de los nutrientes, fertilidad, acidez y los efectos provocados en ellos por aplicacion de plaguicidas y otros contaminantes (Esquivel, 2011).

En este particular, se encuentran los nematodos. Estos son organismos vermiformes que pertenecen al lo Nematoda, no segmentados, pseudocelomados, tripoblastos que pueden
ser hialinos (de vida libre o toparasitos) o no hialinos (parasitos vertebrados) que viven
en la mayor a de los habitats ecologicos (Association et al., 2004; Coto, 2007).

Estos microorganismos se alimentan de una amplia variedad de especies, que incluyen plantas, animales y bacterias. Asimismo, se hallan en un orden de millones de individuos de distintas especies por metro cuadrado de suelo o sedimento acuatico. Su rol es proveer la regulacion de los ecosistemas en los suelos, r os y mares, y ademas se caracterizan por ser recicladores de nutrientes (Wilson y Kakouli-Duarte, 2009).

Algunos de estos organismos vivientes debido a su tama~no, no pueden verse a simple vista,
por lo que su estudio es posible unicamente por medio de microscopios opticos (gura 1.1).

Figura 1.1: Imagen de dos nematodos capturados en un microscopio optico.

Ademas en el caso de los nematodos toparasitos, cuya dieta se compone parcial o to-talmente por las ra ces de las plantas, Wilson y Kakouli-Duarte (2009) se~nalan que son causantes de da~nos como:

- A las celulas por accion mecanica o enzimatica.
- Adaptativos por modi cacion de las celulas del hospedero.
- Neoplasicos que inducen a crecimiento de nuevo tejido, provocando un da~no pau-latino que se mani esta en el pobre crecimiento, marchitez generalizada, baja pro-duccion, clorosis, defoliacion, ca da prematura de ores y frutos.

Por estas razones, la sobrepoblacion descontrolada de estos microorganismos en diferen-tes agroecosistemas puede llegar a ser devastador. Los da~nos y perjuicios economicos

atribuibles a los nematelmintos toparasitos supero en el a~no 2018 miles de toneladas de productos equivalentes a unos $118 billones de perdidas anuales en el mundo (Bernard et al., 2017; Duran-Mora, 2018).

Segun Bongers y Ferris (1999), las perdidas por causa de la accion de nematodos van desde los derivados de origen agr cola hasta los de procedencia animal, ademas del alto costo ambiental por la aplicacion de nematicidas; siendo esto ultimo un problema global que no se ha logrado evitar (Kenney y Eleftherianos, 2016).

El estudio de estos microorganismos proporciona criterios de evaluacion para la toma de decisiones en los procesos de sostenibilidad de los ecosistemas respecto a la conservacion y remediacion de los suelos y mares (Bongers y Ferris, 1999). Aspectos como su abundancia y simple manipulacion, facil muestreo, representatividad segun su habitat y por su respuesta bien de nida segun cambios ambientales, transforman a los nematodos en fuente primaria de informacion (Wilson y Kakouli-Duarte, 2009).

Asimismo, como lo apuntan Esquivel y Peraza (2009), su estudio inicia con su extraccion por tecnicas de muestreo en los campos de cultivo para ser analizados. Posteriormente, en laboratorios especializados, son in ltrados y transferidos a portaobjetos de vidrio (o laminillas de Cobbs) y con ayuda de un microscopio, platos de conteo y maquinas de contador multiple, son cuanti cados y analizados. Aunque las tecnicas de conteo y analisis son procesos efectivos; se tornan lentos por la cantidad de muestras y el tiempo requerido para procesar cada una (Van Bezooijen, 2006).

Los metodos de vision por computador y aprendizaje de maquina dan la posibilidad de apoyar y mejorar las capacidades de trabajo en los laboratorios de investigacion con el proposito de transformar la informacion almacenada en nubes de datos crudos a un conjunto de reglas y conocimientos aplicables a distintas areas del saber (Friedman et al., 2001). En este particular, en la tabla 1.2 se resumen tareas de sistemas expertos de vision por computadora aplicados a objetos estructurales en imagenes digitales de microscop a de nematodos.

Tabla 1.2: Tareas de vision por computador aplicables en imagenes digitales de nematodos segun Geron (2019), Gonzalez y Woods (2008), Jahne (2005).

Medio	Aplicacion
Imagenes digitales	Segmentacion
	Conteo
	Clasi cacion segun su lo
	Clasi cacion segun deformacion
	Seguimiento
	Adaptacion de forma

En este trabajo se propone un modelo de forma que se genera a partir de un conjunto de datos de entrenamiento y una forma inicial expresada en hitos de frontera, y logra evolucionar su silueta hasta adoptar y capturar la forma intr nseca del organismo usando informacion de la imagen digital.

Aunado a lo anterior, la captura de la forma intr nseca de los nematodos en imagenes digitales, le permitira a los especialistas avanzar y complementar en procesos y tareas posteriores de analisis acerca del estado actual de los nematodos respecto a su morfolog a y la interaccion con su entorno. Ejemplos de estas corresponden a:

- Analisis de su locomocion e interaccion con el medio ambiente (Bogale et al., 2020; Juhasz y Zelei, 2013).
- Extraccion de la forma y locomocion para determinar su estado nutricional (Gallag-her et al., 2013).
- Estimacion del area de la super cie y volumen a partir de datos morfometricos (Brown et al., 2016).

Ademas, con la aplicacion de este metodo, un especialista podr a encargarse unicamente de tareas propias de su quehacer profesional, tales como:

- Clasi cacion de especies dello nematoda.
- Diagnostico de da~nos ocasionados por nematodos toparasitos.
- Evaluacion de la efectividad de nematicidas y fertilizantes.

Aunado a lo anterior, este trabajo sirve de base en procesos mas generales donde la captura de la forma de objetos estructurales no vermes en imagenes digitales es preponderante. Ejemplos de estos son:

- Analisis de la variaciones o perturbaciones a partir de la formas en los elementos estructurales, en una parte o la totalidad del cuerpo (Costa et al., 2011; Sadrnia et al., 2007).
- La captura de las caracter sticas morfologicas relevantes de los elementos estructu-rales para evaluacion y diagnostico (Cury, 2015).
- Clasi cacion de objetos estructurales segun caracter sticas de forma (Nalepa y Ka-wulok, 2014; Wilder et al., 2011).
- Generacion de nuevas instancias sinteticas (Arroyo y Alvarado, 2020).

La propuesta en este trabajo se basa en un proceso de optimizacion asistido por dos metodos de deformacion intermedios y una base de datos normalizada de formas vermes expresadas como hitos de frontera. En la gura 1.2 se ilustra un diagrama de bloques general de la investigacion.

Debido a la importancia mencionada en el estudio de los nematodos y al impacto que puedan tener los sistemas automatizados de rastreo y segmentacion en imagenes digitales,

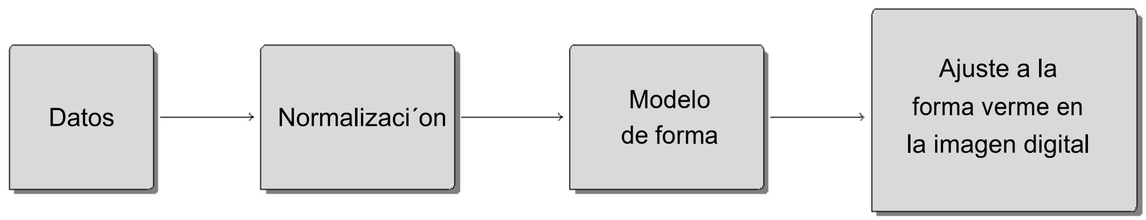


Figura 1.2: Diagrama de bloques del modelo evolutivo de forma.

este trabajo propone un modelo de forma adaptativo ajustable a la silueta de un nematodo que sirva de apoyo para las tareas expuestas en la tabla 1.2.

Asimismo, segun Berry et al. (2019), este trabajo de investigacion se enmarca en el Modelo de aprendizaje de maquina supervisado pues a partir de un conjunto de datos expresado de forma vectorial se logra modelar, a traves de un aprendizaje inductivo, la silueta de la forma verme en una imagen digital.

1.1. Hipotesis y objetivos de investigacion

En el presente trabajo se propone una investigacion en la que se analiza, desarrolla y evalua un nuevo metodo de ajuste de formas vermes en imagenes digitales. Por lo que se propone demostrar la siguiente hipotesis de investigacion:

ƒEs posible dise~nar un sistema adaptacion de formas vermes, a partir de representacio-nes de hitos de frontera mediante un proceso de optimizacion multiobjetivo asistido por deformaciones vermes simuladas,.

En relacion con la hipotesis anterior, se propone como objetivo principal de este trabajo dise~nar un modelo matematico novedoso no lineal basado en instancias de representacion de hitos de frontera en dominios de forma verme permitidos, ajustable a cualquier silueta verme en imagenes digitales que sea robusto ante distintas deformaciones y ante el ruido presente en las imagenes digitales.

Asimismo, esta investigacion se cimienta en tres objetivos espec cos. El primero consiste en dise~nar un nuevo modelo matematico de analisis de formas orientado a la descripcion e ciente de una silueta vermiforme en imagenes digitales.

El segundo objetivo espec co estriba en desarrollar un algoritmo proveniente del modelo propuesto que ajuste la silueta al organismo verme presente en la imagen digital donde el el algoritmo debe ser capaz de recopilar la informacion en la imagen digital, de modo que, a partir de esta le sea posible extraer los parametros de forma que describen su silueta.

El ultimo, es crear una estrategia de evaluacion para modelos y algoritmos de analisis de forma en imagenes digitales, que permita integrar en el proceso de evaluacion con el modelo de forma propuesto.

1.2. Contribuciones del trabajo propuesto

Las contribuciones de esta investigacion son:

- Un metodo adaptativo a los movimientos corporales de organismos vermes en image-nes digitales por medio de un proceso de optimizacion multiobjetivo usando las fun-ciones de posicionamiento y distancia, asistida por dos metodos de deformaciones vermes.
- Un metodo capaz de simular la locomocion de organismos vermes a traves de un numero limitado de parametros usando caminatas aleatorias entre n-s mplexes en-gendrados por medio de caminatas aleatorias.
- Generacion de conjuntos de datos de entrenamiento de formas vermes, expresadas en hitos de frontera para alimentar sistemas de inteligencia arti cial.
- Un metodo de truncado de siluetas no vermes a formas validas en subdominios, permitidos a traves de n-s mplexes.

Por otro lado, este trabajo se ha cimentado en contribuciones propias que son derivadas de esta investigacion y que coadyuvaron a los resultados obtenidos. Primeramente se hizo un analisis comparativo de los metodos ACP y ACPK divulgado en (Arroyo, 2016), la cual fue complemento para el desarrollo del metodo para el calculo de la preimagen en

ACPK. Este fue publicado en (Arroyo y Alvarado, 2014). Asimismo, otro aporte es el desarrollo del metodo de simulacion de movimientos vermiformes mediante caminatas aleatorias entre n-s mplex vecinos expuesto en (Arroyo y Alvarado, 2020).

1.3. Metodolog a

La investigacion se enmarca bajo el paradigma exploratorio y el enfoque inductivo usual del dise~no en ingenier a. Consiste en el establecimiento de enunciados que se desarro-llan a partir de experiencias previas con el problema y soluciones particulares realizadas (Hernandez y Torres, 2018).

La estrategia metodologica del modelo de forma recorre las siguientes fases: En la primera, se seleccionan las bases de datos de imagenes digitales y a partir de esta, se prepara la de formas vermes vectoriales normalizadas. En esta se selecciona para las pruebas a nivel de imagen digital las que cumplen las restricciones descritas en la seccion 3.1. Ademas, se establece una metodolog a que normaliza a la misma escala, rotacion, traslacion y numero de hitos toda la base de datos de formas vermes vectoriales.

En la segunda y tercera fase, de forma paralela, se desarrolla e implementa el modelo de forma. El proceso conlleva etapas iterativas de observacion, analisis y dise~no del modelo, as como la implementacion.

Estas etapas conllevan dos subprocesos bien de nidos:

- A nivel de modelo de forma.
- A nivel de modelo de forma con informacion de imagen digital.

A nivel de modelo de forma, se desarrolla una estrategia que permita la adaptabilidad a distintas siluetas, garantizando la representatividad de cualquier forma verme.

A partir de la evidencia del subproceso anterior, al modelo se le incorpora otras estrategias que permitan la captura de la silueta a nivel de imagen digital.

Finalmente, en la cuarta etapa se realiza las pruebas de validacion y conclusiones que evidencien los resultados nales con los cuales se logra demostrar la hipotesis de investi-gacion. La metodolog a de evaluacion es a nivel de modelo de forma y, a nivel de modelo de forma con informacion de la imagen digital.

En el caso de la evaluacion del modelo de forma, se evalua su capacidad de adaptabilidad a diferentes formas de silueta, y un comparativo de metodos que sigan la linea establecida.

En el caso de de la evaluacion del modelo de forma a nivel de imagen digital, se evalua la cantidad de hitos necesarios de representacion, la adaptabilidad a las formas de nematodo respecto al modelo desarrollado.

Para la evaluacion de los resultados nales, se trabaja con el metodo de validacion cruzada de 10 iteraciones y el ndice de Jaccard.

En la gura 1.3 se muestra un diagrama de ujo con estrategia metodologica propuesta con las fases propias de investigacion, en el orden de aplicacion del modelo.

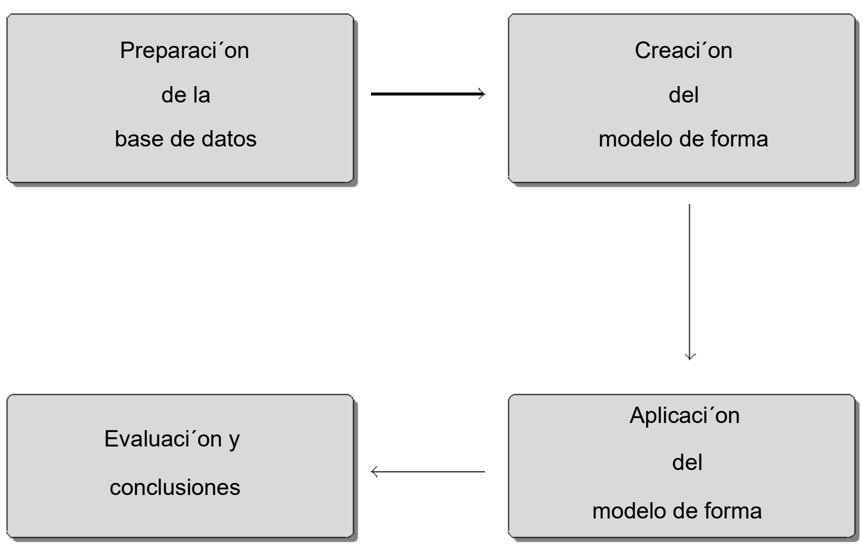


Figura 1.3: Diagrama de la estrategia metodologica de solucion al problema propuesto.

1.4. Delimitaciones al alcance de la investigacion

Este trabajo se delimita bajo las siguientes restricciones:

- El alcance de este trabajo es a nivel de propuesta metodologica. No se evaluan los tiempos de respuesta en las pruebas de convergencia y tampoco se realiza un analisis de complejidad algor tmica.
- Los algoritmos se implementan usando el lenguaje C++, junto con la biblioteca

  FLANN (Muja y Lowe, 2009) para el calculo e ciente de los centroides y de busque-da de vecinos cercanos, as como la biblioteca LTI-Lib-2 (Dorer y Alvarado, 2006)

  como base para el procesamiento digital de imagenes. Asimismo, todos los expe-rimentos se realizaron en una computadora de escritorio con procesador Intel(R) Core(TM) i5-2500 CPU @ 3.30GHz con 8GB de RAM.

1.5. Estructura del documento

Este documento se constituye con los siguientes apartados: en el cap tulo 2 se describen en detalle los contenidos y conceptos en los que esta investigacion se fundamenta. En el cap tulo 3 se detalla el modelo desarrollado y los algoritmos implementados para la solucion al problema propuesto. En el cap tulo 4 se exponen los resultados y un analisis de los mismos. Finalmente, en el cap tulo 5 se presentan las conclusiones nales y un listado de recomendaciones sobre posibles l neas de trabajo a futuro.

Cap tulo 2

Marco teorico

En este cap tulo se presenta un listado de elementos teoricos resultantes de una busqueda de estrategias en otras investigaciones a nes y de hallazgos, los cuales se han adaptado y modi cado para la generacion de la nueva propuesta y que son necesarios para la demostracion de la hipotesis de investigacion.

El orden del desarrollo teorico se ha estructurado de acuerdo con las secciones corres-pondientes donde fueron utilizados. Se inicia haciendo una referencia a los trabajos a nes a este. Luego, se describen conceptos relativos a imagenes digitales y esqueletizacion. Seguido a temas como los s mplexes geometricos, modelos deformables, interpolacion pa-rametrica y optimizacion multiobjetivo. Para nalizar, se detallan metodos de evaluacion estad stica aplicados en esta investigacion.

2.1. Trabajos relacionados

En investigaciones a nes se han desarrollado modelos para la deteccion de estructuras vermiformes en imagenes digitales basadas en el caso particular de los nematodos. Estos se fundamentan en dos areas: la primera abarca los modelos basados en informacion de la imagen digital a traves de un preprocesado de estas, por medio de operaciones morfologicas y aplicacion de ltros (Jahne, 2005); y la segunda corresponde al conjunto de tecnicas de aprendizaje de maquina supervisado y no supervisado (Berry et al., 2019). Cabe destacar que ambas areas no son excluyentes, y que por lo general son utilizadas conjuntamente.

En el area de procesamiento digital de imagenes de nematolog a existen diversas l neas de trabajo. Por ejemplo, Gomez (2009) trabajo en dos niveles de abstraccion para segmentar la imagen digital separando las partes pertenecientes a nematodos y al fondo de la imagen. Aunque en la investigacion se aplican dos metodos de segmentacion para la deteccion de las formas vermes, se limita a procesar informacion a nivel de imagen digital y no se le incorpora la de forma al metodo. Asimismo, en las conclusiones de ese trabajo se resalta que imagenes digitales ruidosas pueden ser limitantes en la aplicacion de los algoritmos que involucren operaciones morfologicas y de ltrado.

En la misma l nea de trabajo, Geng et al. (2004) propusieron un sistema de seguimiento automatico y extraccion de caracter sticas del nematodo C. elegans. Dicho sistema se desarroll en varios subprocesos con imagenes digitales binarizadas y la aplicacion de un algoritmo de esqueletizacion para la deteccion de las posiciones de los nematodos. En esta investigacion se presentan imagenes digitales con escenas no ruidosas, aspecto por el cual deja la posibilidad que en otros tipos de ambientes con imagenes ruidosas, los resultados no sean los mismos.

Similarmente, una tecnica de aprendizaje basada en contorsiones de la cabeza del C. elegans se expone en (K.-M. Huang et al., 2008). En dicho metodo se propuso medir los angulos de contorsion de la cabeza y cola en sus movimientos exploratorios de locomo-cion. La informacion espacio-temporal es aprovechada por Restif y Metaxas (2008) para predecir la posicion y ajustar la forma del organismo en procesos de segmentacion y ras-treo, utilizando para ello la evolucion de los angulos y vertices del esqueleto morfologico.

Por su parte Rizvandi, Pizurica, Philips y Ochoa (2008) presentan una propuesta para la deteccion y separacion de nematodos C. elegans con presencia de traslapes. El algoritmo pasa por las fases de pre-procesamiento de escala de grises y operaciones morfologicas. Tomando como base las ideas de la investigacion anterior, en (Rizvandi, Pizurica y Philips, 2008) se propone un algoritmo que permite la division y reconstruccion de los cuerpos de los nematodos en el punto de traslape para su identi cacion por medio de la tecnica de esqueletizacion. Para esto se utiliza el promedio del angulo que forman los vectores dados en el proceso de esqueletizacion del especimen por medio del etiquetado de p xeles de union, conexion, de cuerpo y extremo del nematodo.

En conjunto, las referencias anteriores se utilizan como base para el desarrollo de los metodos forrajeo y serpenteo en esta investigacion (secciones 3.3 y 3.4 respectivamen-te) permitiendo la simulacion de secuencias movimientos laterales de cola y cabeza del nematodo en movimientos continuos.

En la misma l nea de investigacion, K.-M. Huang et al. (2008) proponen un algoritmo de descripcion cuantitativa de los movimientos de alimentacion del nematodo C. elegans por medio de un sistema de posicionamiento automatico de v deo que detecta los distintos momentos de alimentacion y su frecuencia. En la etapa de pre-procesamiento, este algorit-mo transforma la imagen digital a escala de grises y utiliza la diferencia de una secuencia nita de marcos de un v deo para detectar la posicion del nematodo. Se umbraliza y utili-za el algoritmo de esqueletizacion morfologico, y nalmente con operaciones morfologicas se extrae el contorno del cuerpo del gusano. En la etapa de deteccion, el algoritmo, a partir de un ciclo de movimientos, detecta las contorsiones de la cabeza y la cola con un sistema de vectores que miden el angulo de la curvatura de los movimientos del cuerpo con respecto a la cabeza. Para reconocer los momentos de alimentacion se tomaron en cuenta parametros como la amplitud del angulo, el intervalo del tiempo entre dos eventos de alimentacion y la frecuencia de los eventos de alimentacion. La amplitud de los angulos de las contorsiones fueron considerados a la presente propuesta.

Por otro lado, en cuanto a tecnicas de aprendizaje automatico con informacion de forma verme, Cootes et al. (1995) desarrollaron los modelos activos de forma usando la tecnica de distribucion de puntos y analisis de componentes principales para la reduc-cion de dimensionalidad del conglomerado de los datos. A partir de este, se sientan las bases de los metodos capaces de expresar expl citamente formas validas por medio de modelos estad sticos de distribucion de puntos y la reduccion de la dimension de vectores representativos de la formas.

En la misma l nea, Mar n (2009) utilizo modelos activos de forma para detectar la ubi-cacion de los nematodos en imagenes digitales apoyado del modelo de Cootes y Taylor (1993) que describe las siluetas de las formas a traves de estructuras vectoriales de hitos de frontera. Este trabajo aborda siluetas de nematodos adaptandolos a la forma de un objeto seleccionado previamente en la imagen digital.

En este conjunto de investigaciones, se asume la premisa que la distribucion de los datos es normal gaussiana. En caso de no ser verdadero, dicha situacion repercute en la no recuperacion o no construccion de nuevas formas vermes a partir de los datos existentes, y como conclusion, estos modelos son incapaces de restringir las formas representadas a vermes validas y, cabe la posibilidad de adaptacion a siluetas anatomicamente no plau-sibles. A pesar de este inconveniente, la tecnica de expresar vectorialmente formas de nematodos a traves de hitos de frontera como metodo de esbozar su silueta es valida, robusta y representativa de las formas.

Lindenbaum et al. (2018) exponen otro concepto basado en un modelo generativo que aprende la geometr a de una variedad topologica a partir de los datos. En esta se procura ignorar su densidad, utilizando mapas de difusion para establecer la estructura geometrica de la variedad. Sin embargo, un problema asociado a esta es que el etiquetado de los datos es un proceso costoso debido a la cantidad necesaria para que en conjunto se logre cubrir el total de deformaciones.

Por tanto, dicha estrategia no es viable en esta investigacion, pues su aplicabilidad depende de una elevada dimension del espacio muestral. Esto es: el modelo requiere alimentarse de un espacio muestral lo su cientemente representativo para que los mapas de difusion sean robustos y puedan estimar nuevas formas vermes. Esto mismo aplica en los modelos de aprendizaje automatico en los que es necesario abarcar un volumen de datos cuantioso para entrenar estos sistemas de aprendizaje. En ambientes de laboratorio de nematolog a, la obtencion de dichas muestras son procesos costosos en tiempo y recursos por lo que metodos automaticos de este tipo son descartados en esta investigacion (Esquivel, 2011; Van Bezooijen, 2006).

Como alternativa para dar mas peso a la geometr a de la variedad topologica que a la densidad de puntos disponibles de entrenamiento, se utiliza en la presente propuesta una tecnica similar a los k-vecinos mas cercanos expuesta por Nutanong et al. (2010) para un contexto de consultas en bases de datos espaciales. El metodo es similar al presentado en este trabajo, en el hecho de que usa tanto la informacion de la posicion actual en la caminata, como del vecindario para la decision del siguiente paso, pero la estrategia elegida para saltar a la siguiente posicion esta en el presente caso orientada a producir formas dis miles al punto de partida, pero manteniendo su validez, en vez de procurar la descripcion geometrica de la variedad. Ademas, el uso del metodo de k-vecinos mas cercanos permite una disminucion de la dimensionalidad del conglomerado de los datos tomando como su representante el elemento mas cercano en forma del centroide en cada agrupamiento.

En otra direccion, pero siempre con tecnicas de aprendizaje, se encuentra la simulacion completa del organismo de un nematodo de la especie C. elegans en el proyecto OpenWorm (Currie et al., 2014). En ese proyecto se logr generar el movimiento locomotivo de un especimen por medio de la simulacion con 302 neuronas y 95 celulas musculares arti ciales. Los datos generados en dicha simulacion se acoplaron a renderizadores que generen las
siluetas o las imagenes, segun se requiera. Si bien es cierto, esta simulacion de abajo hacia
arriba (bottom-up) permite alcanzar altos niveles de realismo anatomico en los detalles, el consumo de recursos computacionales no lo hace viable en contextos donde la generacion de formas sea parte de otros procesos, como la segmentacion de nematodos o su rastreo en imagenes.

En este grupo de metodos tambien aparecen otras alternativas para representar el espacio
de formas validas. Por ejemplo Gruner (2015) logra describir una variedad basandose en tecnicas de aprendizaje de diccionarios. En este metodo no se lleva a nivel de imagen digital, sin embargo, se extrae la idea de generacion de otras formas de gusano a partir de coordenadas baricentricas vermes en n-s mplexes.

Por otra parte, Jimenez (2019) utiliza un modelo de aprendizaje profundo como medio de generacion y segmentacion de nuevas formas vermes. Aunque el metodo es robusto, se necesita de una gran cantidad de datos de entrenamiento y prueba que permita recuperar toda la variedad topologica de formas de gusano. Este aspecto hace que el proceso de investigacion sea costoso en terminos de recursos para la extraccion de los datos as como tiempo de procesamiento de los mismos.

Asimismo, en el proyecto Automated Wormscan Puckering et al. (2017) muestran la utili-dad e importancia en la adquisicion rapida de datos por sistemas computacionales usando tecnicas de vision por computador para la deteccion de nematodos C. elegans para ensa-yos de toxicidad, crecimiento y fecundidad. En esta investigacion, se evidencia el aumento del rendimiento y la reduccion de los costos a gran escala, y una mejora de los resultados
obtenidos de los metodos manuales y a los automaticos por computador. Al igual que la
investigacion anterior, y debido a la manera de adquisicion de datos necesarias usando
escaneres de tres dimensiones, hace que esta investigacion no sea considerada.

En general, todas estas investigaciones utilizan un proceso de ltrado en las imagenes digitales para eliminar ruido y realzar las caracter sticas de los nematodos. Luego, de-terminan un sistema de parametros de clasi cacion o normas a partir de los indicadores observados, en algunos casos un proceso de esqueletizacion y nalmente, a traves de algun
sistema automatico se entrenan para la deteccion de nematodos en imagenes digitales.

No obstante, se presentan varios problemas sustanciales. Primeramente, el crecimiento exponencial del volumen de los datos para describir las instancias de las deformaciones de las estructuras vermiformes. La consecuencia directa es que el numero de parametros necesarios para describir la forma es casi el mismo numero que la representacion directa.

Segundo, se parte de la suposicion de que se puede eliminar la redundancia del modelo si los puntos se distribuyen de manera normal en el espacio de formas; de este modo el modelo ser a su ciente para describir la variabilidad de las formas. Sin embargo, si la suposicion no se cumple, entonces la reduccion de dimensiones conduce a la perdida de informacion y obliga a usar mas dimensiones en el espacio proyectado. Es decir, a mayor cantidad y variabilidad de muestras de entrenamiento se pierden las suposiciones de entrada, y por lo que es incapaz de describir las formas de los nematodos.

Tercero, se requiere expl citamente informacion de la forma de organismos para determinar indicadores de interes y su ciente material de entrenamiento, que es en general, costoso de preparar manualmente. La informacion en los laboratorios es limitada debido al tiempo de obtencion.

Cuarto, el uso de ltros de procesamiento de imagenes aplicados a grupos particulares de entrenamiento, tales como deteccion de bordes, reduccion de las variaciones de intensidad entre p xeles vecinos para suavizar la imagen, o el aumento de las variaciones de la inten-sidad con el n de realzar sus caracter sticas, y hasta operaciones morfologicas de imagen digital. O bien, las imagenes presentadas en estos trabajos son limpias sin detritos, y con esto simpli can procesos como la segmentacion o similares, situacion que en muestras reales de laboratorio no sucede (gura 2.1).

Figura 2.1: Imagen digital de un nematodo con presencia de multiples detritos y ruido.

El problema es que las condiciones aplicadas obedecen a particularidades del dominio espacial del conjunto de imagenes digitales de entrenamiento, y en circunstancias propias de cada experimento, y no a procesos generales aplicables en cualquier escena o situacion. Esta situacion tiene como consecuencia la perdida de generalidad de los procesos descritos para ser aplicables en cualquier proceso similar.

En las referencias cient cas citadas y otras a nes, no se encontr un modelo capaz de describir parametricamente la formas de estos organismos en todos sus estados de crecimiento, sus caracter sticas de movilidad y variabilidad morfologica por medio de un numero limitado de parametros capaces de ajustarse a la informacion existente en la ima-gen, con presencia de detritos y, que permitan distinguir entre las estructuras vermes y los restantes elementos presentes en las imagenes digitales. Esto limita la aplicacion de los metodos propuestos a tareas de mayor nivel de abstraccion como lo es el conteo y el reconocimiento automatizado.

Sin embargo, de las referencias descritas se extraen aportes base para la investigacion presente. En la tabla 2.1 se resumen un listado de dichas contribuciones.

Tabla 2.1: Aportes base de otras investigaciones a MEF.

Aportes base	Referencias
Modelo de forma por hitos de frontera	(Cootes et al., 1995), (Cootes y Taylor, 1993), (Shlens, 2009), (Mar n, 2009)
Contorsiones vermes de cabeza y cola	(K.-M. Huang et al., 2008), (Geng et al., 2004), (K.-M. Huang et al., 2008)
Proceso de esqueletizacion	(Gomez, 2009), (Rizvandi, Pizurica y Philips, 2008)
k-vecinos mas cercanos para la reduccion de dimensionalidad	(Nutanong et al., 2010)
n-s mplexes como generador de formas vermes	(Gruner, 2015)

2.2. Imagenes digitales

Gonzalez y Woods (2008) definen una imagen digital como una funcion $I_{x;y}$ de dominio discreto dada por:

$$I_{x;y} : X_1 \times X_2 \to \mathbb{N}_n^3 \quad \text{con} \quad I_{x;y} := \begin{bmatrix} r_{x;y} \\ v_{x;y} \\ a_{x;y} \end{bmatrix} \tag{2.1}$$

donde $X_1 := \{0, 1, \ldots, d_1 - 1\}$, y $X_2 := \{0, 1, \ldots, d_2 - 1\}$ con $d_1, d_2 \in \mathbb{N}$. La funcion $I_{x;y}$ devuelve la combinacion de la composicion espectral en los canales de informacion rojo $r_{x;y}$, verde $v_{x;y}$ y azul $a_{x;y}$ de un píxel en la posicion $(x;y) \in X_1 \times X_2$.

Asimismo, Sonka et al. (2014) definen la transformacion a escala de grises como una funcion que mapea los valores resultantes de la composicion espectral al intervalo de $[0 ; L]$, donde cada canal realiza un mapeo de $\mathbb{N}_n^3$ a $\mathbb{N}_n$, degradando en tonalidades de grises (denominados niveles de intensidad) de 0 hasta L, donde el 0 corresponde al negro y L al blanco.

Asimismo, el vector $\nabla I_{dx;y}$ es la aproximacion discreta del gradiente direccional en señales continuas que proporciona la direccion de la mayor variacion o cambio entre los niveles de intensidad. Viene dado por:

$$\nabla I_{dx;y} = \begin{bmatrix} \frac{\partial}{\partial x} T(I_{x;y}) \\ \frac{\partial}{\partial y} T(I_{x;y}) \end{bmatrix} \approx \begin{bmatrix} S_x \circledast T(I_{x;y}) \\ S_y \circledast T(I_{x;y}) \end{bmatrix} \tag{2.2}$$

donde $\circledast$ es el operador de convolucion dado por:

$$S_x \circledast T(I_{x;y}) = \sum_{dx=-a}^{a} \sum_{dy=-b}^{b} S_x(dx;dy)\, T(I_{x+dx;y+dy}) \tag{2.3}$$

$$S_y \circledast T(I_{x;y}) = \sum_{dx=-a}^{a} \sum_{dy=-b}^{b} S_y(dx;dy)\, T(I_{x+dx;y+dy}) \tag{2.4}$$

T es la funcion que transforma la imagen a niveles de gris y, S_x y S_y corresponden a las matrices kernel de estimacion del gradiente tales como diferencia simple, Sobel, Prewitt, Robinson, Roberts, Kirsch y Ando y derivaciones orientadas de la funcion Gaus-siana (OGD por sus siglas en ingles)(Gonzalez y Woods, 2008; Jahne, 2005). Los rangos $-a \le dx \le a$ y $-b \le dy \le b$ dependen del tamaño de las matrices kernel.

2.3. Metodos de reduccion de dimensionalidad

Los metodos de reduccion de dimensionalidad (MRD) son procedimientos que transforman el conjunto de datos X en un nuevo conjunto de datos Y de menor dimensionalidad, conservando la informacion intr nseca tanto como sea posible (Arroyo, 2016). El objetivo de esta transformacion es evidenciar propiedades no observables en el conglomerado de la informacion de forma signi cativa sin redundancia (J. A. Lee y Verleysen, 2007; Van Der Maaten et al., 2009).

En las secciones siguientes se abordan los metodos de reduccion de dimensionalidad Anali-sis de componentes principales y Analisis de componentes principales con kernel respec-tivamente.

2.3.1. Analisis de componentes principales

El analisis de componentes principales (ACP) es una tecnica lineal que se utiliza para la eliminacion de la redundancia de los datos (Shlens, 2009). El metodo de ACP hace un cambio de base a una de menor dimensionalidad a traves de la ecuacion de transformacion:

$$Y = PX \tag{2.5}$$

donde P es una matriz ortogonal denominada matriz de representacion. El objetivo es calcular una matriz P que permita proyectar los datos a un espacio de menor dimension de manera que se garantice la no correlacion entre vectores de Y . Si la correlacion entre las distintas muestras es nula, se elimina la redundancia y el subespacio de datos puede ser descrito por P .

El algoritmo para el calculo de P inicia centrando y estandarizando los datos. Luego, se calcula la matriz de covarianza de X dada por C_{cov} que es simetrica y diagonalizable, y que cuanti ca la covarianza entre las mediciones. Luego, se obtienen los vectores propios de C_{cov} ordenados de acuerdo al valor propio respectivo y que sirven de nuevas coordenadas del sistema donde se maximiza la varianza de los mismos.

Los vectores propios de C_{cov}, denominados componentes principales, describen la infor-macion del conjunto de datos de acuerdo con su coe ciente de inercia.

2.3.2. Analisis de componentes principales con kernel

El analisis de componentes principales con kernel (ACP K) es un metodo para la reduccion de dimensionalidad de datos en el que se aplica el metodo de analisis de componentes principales en un espacio F denominado caracter stico (Scholkopf, Smola et al., 1998).

Para esto, los datos de entrada E IR^n, son mapeados a F a traves de una funcion de nida por : E ! F. El calculo de la q esima componente principal no lineal de un x 2 E es dado por:

$$V^{q} \quad (x) = \overset{m}{\underset{=1}{\scriptstyle x}} \; {}_{i}^{q} h\,(x_i);\ \ (x)i \tag{2.6}$$

donde V^{q} es q-esimo vector propio de la matriz de covarianza de los datos proyectados, los $_i$ son entradas del vector normalizado determinado en el eigensistema:

$$m \quad = K \tag{2.7}$$

y K es una matriz de orden m con entradas $K := (k_{ij})$ de nidas por:

$$k_{ij} := k(x_i; x_j) = h\,(x_i);\ \ (x_j)i \tag{2.8}$$

donde k(;) recibe el nombre de kernel y evita hacer la proyeccion expl cita de los datos. Ejemplos de kernels mas utilizados segun Honeine y Richard (2011) se muestran en la tabla 2.2.

Tabla 2.2: Ejemplos de kernels para c 0; n 2 IN y > 0.

Funcion kernel	Criterio
Proyectivo polinomial	$(x_i; x_j) := (hx_i; x_j i + c)^n$
Proyectivo exponencial	$(x_i; x_j) := \exp(hx_i; x_j i =(2\ ^2))$
Proyectivo sigmoide	$(x_i; x_j) := \tanh(hx_i; x_j i = \ + c)$
Radial Gaussiana	$(x_i; x_j) := \exp(\quad x_i \quad x_j jj^2 =(2\ ^2))$
Radial Laplaciana	$(x_i; x_j) := \exp(\quad x_i \quad x_j jj =(2\ ^2))$
Radial Multicuadratico	$(x_i; x_j) :=$ p $\overline{jjx_i \quad x_j jj^2 + c}$

Calculo de la preimagen en ACP K

Las funciones descritas en la seccion 2.3.2 en general son no lineales, lo que conlleva a que la dimension del espacio F engendrado por las imagenes del conjunto de datos exceda la dimension del espacio de entrada E, y en ese particular, en el ACP K no siempre es posible el calculo directo de los vectores preimagenes de un vector en F (Scholkopf, Mika et al., 1998; Scholkopf, Smola et al., 1998).

Lo anterior se denomina el problema de la preimagen. Mas precisamente, el problema a resolver consiste en determinar la mejor aproximacion para x 2 E, suponga x 2 E, tal que para 2 F que cumpla:

$$(x\ ) \qquad (2.9)$$

para algun x 2 E y que cumpla que kx x k < En en la gura 2.2 se muestra un diagrama que ilustra el problema de la preimagen.

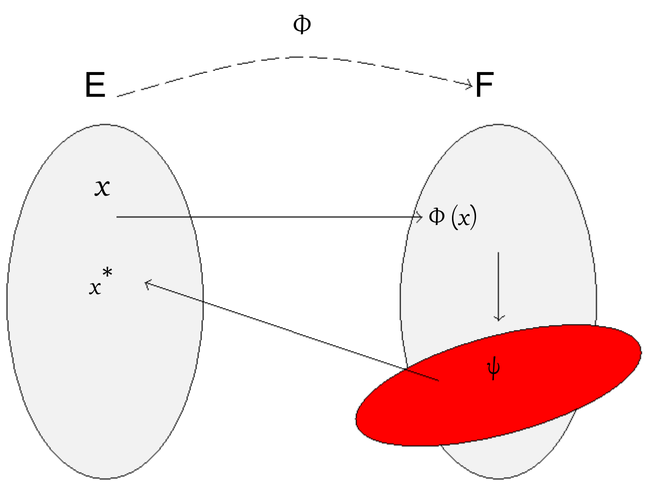


Figura 2.2: Problema de la preimagen. Diagrama de ACP K para x 2 E y su respectiva aproximacion x 2 E. En rojo se colorea el subespacio de proyeccion para .

.

Lo anterior se deriva en un problema de optimzacion de distancia m nima entre el vector en el espacio de entrada original x y el vector preimagen x del vector proyectado correspondiente a (x) (Arroyo y Alvarado, 2014). Una alternativa para la reconstruccion de las preimagenes x es dada por una variacion del metodo de Aproximacion por mapeos conformes(AMP)(Honeine y Richard, 2011).

En este metodo se parte de la representacion impl cita de a traves de la combinacion lineal de los vectores de entrenamiento proyectados (x_j) en F dada por:

$$= \sum_{k=1}^{n} {}_{k} \ (x_k) \qquad (2.10)$$

O bien, por componentes:

$$_i = \sum_{=1}^{n} {}_k^{(i)} \; (x_k) \qquad (2.11)$$

donde los coe cientes ${}_k^{(i)}$ se extraen de la solucion de (2.7). En esta se asume la existencia de $_i$ como imagen de un vector sin ruido generada por m (m < n) vectores propios:

$$= \sum_{=1}^{m} h \, (x \,); \;\; {}_i i \; {}_i \qquad (2.12)$$

Si se sustituye (2.11) en (2.12) y se aplica sesquilinealidad en el producto interno, se obtiene:

$$\begin{aligned}
&= \sum_{i=1}^{m} h \, (x \,); \;\; {}_i i \; {}_i \\
&= \sum_{i=1}^{m} \hat{} (x \,); \sum_{=1}^{n} {}_l^{(i)} (x_l) + \sum_{k=1}^{n} {}_k^{(i)} \; (x_k) \; ! \\
&= \sum_{i=1}^{m} \sum_{=1}^{n} {}_l^{(i)} h \, (x \,); \;\; (x_l) i ! \sum_{k=1}^{n} {}_k^{(i)} (x_k) ! \qquad (2.13) \\
&= \sum_{k=1}^{n} \sum_{i=1}^{m} {}_k^{(i)} \sum_{l=1}^{n} {}_l^{(i)} h \, (x \,); \;\; (x_l) i \; !! \; (x_k) \\
&= \sum_{k=1}^{n} {}_k \; (x_k)
\end{aligned}$$

de donde se in ere que:

$$_k = \sum_{i=1}^{m} {}_k^{(i)} \sum_{l=1}^{n} {}_l^{(i)} h \, (x \,); \;\; (x_l) i \; ! \qquad (2.14)$$

Finalmente, al aplicarle el metodo de m nimos cuadrados lineales a la ecuacion solucion sugerida en AMP se obtiene:

$$X^T x = (X^T X K^{\;1}) \qquad (2.15)$$

y usando (2.13) se logra dar con una solucion cerrada al calculo de la preimagen x a traves de la ecuacion:

$$x = (XX^T\ )\ ^1X(X^T\ X\ \ K\ \ ^1) \tag{2.16}$$

donde es el parametro de regularizacion, $X = x_1\ x_2 \quad x_n$ es el conjunto de datos expresados matricialmente, y es el vector dado por:

$$= \ _1 \ \ _2 \ ::: \ \ _n{}^T \tag{2.17}$$

en el cual cada $_i$ se calcula a traves de la ecuacion (2.14).

En este documento, se utiliza el metodo ACP K en conjuncion con la variante del metodo AMP propuesta para describir las formas de nematodos a traves de hitos de frontera. El objetivo es reconstruir nuevas formas vermes usando un numero menor de dimensiones al del original sin perdida de informacion en un espacio de caracter sticas para luego ser proyectado al espacio original.

2.4. Modelos deformables

Los modelos deformables, introducidos por Terzopoulos et al. (1987), son un grupo de algoritmos que tienen como objetivo modelar la variabilidad de la estructura geometrica de un objeto en curvas de dos dimensiones o super cies de tres dimensiones dado como extension de trazadores generalizados de orden mixto empleados en topicos de vision por computador. Estos comparten la caracter stica de modelar la variabilidad de objetos deformables restringidos a un conjunto de caracter sticas propias determinadas a priori (Albrecht et al., 2009).

Hidalgo et al. (2012) presentan una clasi cacion de modelos deformables para represen-taciones discretas y continuas (gura 2.3).

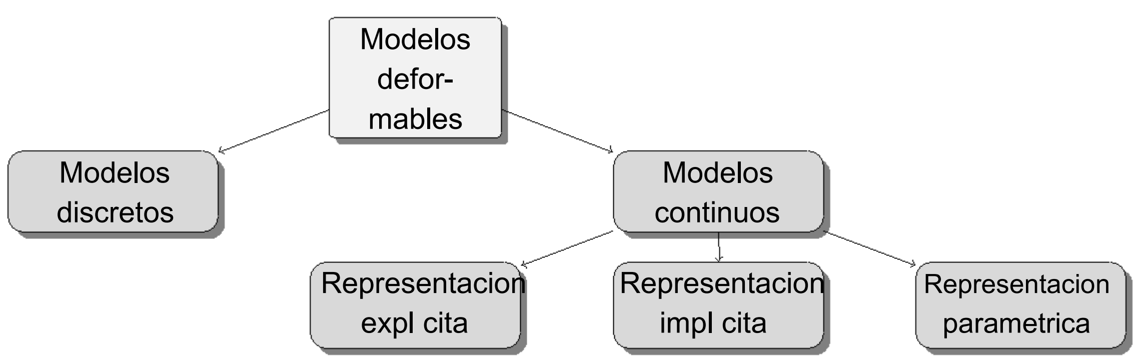


Figura 2.3: Clasi cacion modelos deformables basados en modelos discretos y continuos basado en lo que expresa Hidalgo et al. (2012)

En caso de los modelos discretos, su representacion es inferida a traves de un conjunto nito de hitos que moldean la gura del objeto. En este estadio se hallan algunos modelos como mallas poligonales (Qing et al., 2019), triangulacion de Delaunay (Chew, 1989), mallas por n-s mplexes (Delingette, 1999) y sistemas de part culas orientadas (Szeliski y Tonnesen, 1992).

En los modelos de representacion continua, se asume que su geometr a y los parametros que la describen son conocidos. Estos son trifurcados en modelos impl citos, expl citos y parametricos. Los modelos de representacion continua expl citos, en el caso de IR^2, son descritos desde un vector de parametros de forma Q = q_1 q_2 : : : q_{nq} y un espacio es:

parametrico de objetos de nidos, donde su representacion por objetos

$$S_q : IR^{n_q} ! IR^2$$

$$S_q (Q ; u) ! (x(Q;u) ; y(Q;u)) \quad (2.18)$$

Ejemplos de este son cilindros generalizados, objetos continuos y modelos de contornos activos. En el caso de los modelos impl citos, son representaciones a traves de funciones reales f que son descritos para el conjunto de valores que los anula:

$$f : IR^2 ! IR$$

$$S_f :=p \ 2 \ IR^2 : f(p) = 0 \quad (2.19)$$

Algunos ejemplos de modelos de representacion impl cita son las super cies algebraicas (Ponce et al., 1989), isosuper cies (Carr et al., 2010), conjuntos de nivel (Sethian, 1999), y representacion de objetos muestreados (Chen et al., 2007).

Los modelos de representacion parametrica de un objeto estan dadas por una aplicacion S de nida en un dominio IR^2 donde cada (u;v) 2 es mapeada a IR^2 como un punto en (x(u);y(v)) dado por:

$$S: ! IR^2$$

$$S (u;v) ! (x(u) ; y(v)) \quad (2.20)$$

El mismo puede ser tambien aplicado en super cies en IR^3 (Hidalgo et al., 2012). En esta clasi cacion se encuentran modelos tales como supercuadratico (Terzopoulos y Metaxas, 1991), la descomposicion modal (Q.-X. Huang et al., 2009), la subdivision de super cies (Loop et al., 2009), las curvas de Bezier (Farin, 1983), los B-trazadores (Atkinson, 2008) y las Nurbs (Piegl y Tiller, 1995).

Por otra parte, Mesejo et al. (2016) proveen otra clasi cacion basada en una representacion h brida entre geometrica-impl cita y parametrica-expl cita (gura 2.4).

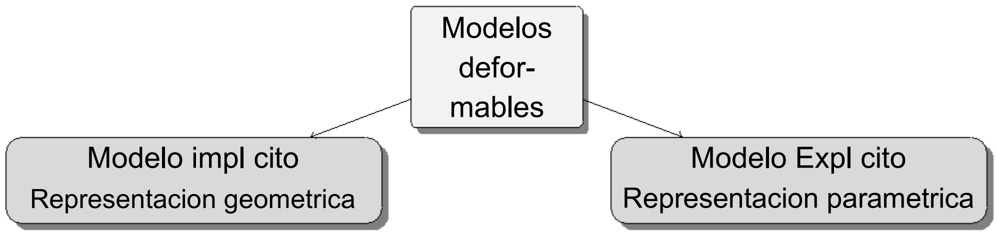


Figura 2.4: Clasi cacion modelos deformables basados en expl citos e impl citos basado en (Mesejo et al., 2016)

Segun las clasi caciones anteriores, el presente trabajo enmarca como un modelo continuo-expl cito con representacion discreta-geometrica. Ademas, el conjunto de las posibles elon-gaciones permitidas por un modelo deformable pertenece a una variedad topologica con-tenida en un espacio eucl deo IR^k.

2.4.1. Modelos de forma basados en aprendizaje profundo

Segun Goodfellow et al. (2016) y Liu et al. (2018), los modelos basados en aprendiza-je profundo son tecnicas de aprendizaje fundamentadas en conjuntos de parametros que le permiten ejecutar tareas automatizadas como segmentacion, identi cacion, posiciona-miento y clasi cacion de objetos en imagenes digitales.

Ejemplos de estos sistemas de aprendizaje de estructuras y patrones aplicables son redes neuronales arti ciales (Livingstone, 2008), redes profundas prealimentadas (feedforward networks) (Raiko et al., 2012), redes de creencia profunda (H. Lee et al., 2009) y codi - cadores automaticos apilados (Masci et al., 2011).

Entre las tareas de estos modelos, se halla el aprendizaje de movimientos naturales de organismos vivos, con el objetivo de generar otros nuevos, reconstruir partes ocultas y de adaptacion de formas. Por ejemplo, Holden et al. (2016) logran sintetizar los movimientos de una silueta de ser humano producidos a partir de una variedad topologica de movi-mientos humanos naturales. La variedad es aprendida a traves de las unidades ocultas de una red neuronal profunda prealimentada entrenada con un autocodi cador variacional (Pu et al., 2016).

Por otra parte, Litany et al. (2018) muestran a traves de un autocodi cador variacional en espacios latentes el aprendizaje de movimientos de distintas extremidades, tronco y cabeza de guras de seres humanos. A partir de estas, se ejempli ca con la reconstruccion de trozos ocultos o difusos de las siluetas en imagenes digitales obtenidas de escaneres tridimensionales.

Asimismo, Jimenez (2019) describe la estrategia para la segmentacion de nematodos en imagenes de microscopia usando una red de aprendizaje profundo para clasi cacion de p xeles del fondo con los hitos que describen la forma de nematodos.

El exito de estos metodos de aprendizaje automatico se basa en la prealimentacion con su cientes muestras datos por tipo de objeto, o su deformacion pertenecientes a una variedad topologica. En el caso de nematodos en imagenes digitales, se necesita una densa base de datos que le permita al modelo de aprendizaje automatico inferir todas las parti-cularidades de las distorsiones y transformaciones de forma de estos gusanos. Dada esta situacion, se opta por no seguir este tipo de estrategias. Sin embargo, se hace necesario lograr emular en las siluetas una variedad topologica verme pero con cantidad limitada de datos de formas de gusano.

2.4.2. Modelos de forma basados en hitos de frontera

Los modelos de forma basados en hitos de frontera son prototipos deformables capaces de emular variaciones validas de forma de algun objeto de acuerdo a un conjunto de restricciones previamente de nidas (Cootes y Taylor, 1993).

Estos modelos son generados a traves de un conjunto de entrenamiento C que capta la variabilidad de todas las posibles deformaciones de los objetos representados a traves de descriptores matematicos. El i-esimo elemento de C de ne un pol gono de n vertices o hitos denotados por h_j. Cada forma se codi ca en un vector de las coordenadas de los hitos:

$$f_i = x_{i\,0} \quad y_{i\,0} \quad x_{i\,1} \quad y_{i\,1} \quad ::: x_{i\,n\ 1} \quad y_{i\,n\ 1} \tag{2.21}$$

o alternativamente como vector de hitos:

$$H_i = h_0 \quad h_1 \quad h_2 \quad ::: \ h_{n\ 1} \tag{2.22}$$

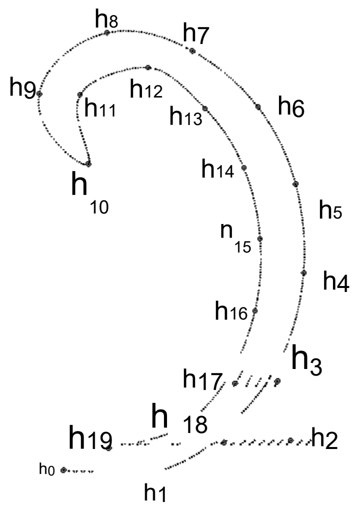
donde cada hito h_j es un vector de coordenadas cartesianas de nido por:

$$h_j = x_{ij} \quad y_{ij}\ ^T \tag{2.23}$$

Acomodados en secuencia, los hitos h_j se ubican a lo largo del contorno del objeto para representar la forma (gura 2.5).

Modelos activos de forma

Segun Cootes y Taylor (1993), los modelos activos de forma (MAF) son tecnicas de ajuste iterativa de las formas de objetos en imagenes digitales, donde los datos de objetos se representan por hitos de frontera. Esta tecnica se basa en modelos de distribucion



1

Figura 2.5: Una forma verme f_i representada por 20 hitos de frontera h_j.

de puntos que describen deformaciones de objetos a partir de variaciones inferidas del conjunto de instancias de entrenamiento y de su media geometrica.

Cada instancia de la base de datos de entrenamiento C de formas a nes es alineada a una instancia escogida a priori. Esto produce que el vector promedio de este conjunto sea nulo. As , la ecuacion de MAF es dada por:

$$x = \bar{x} + Qb \tag{2.24}$$

donde $\bar{x}$ es un vector que representa la forma media, Q es la matriz de los primeros t vectores propios de la matriz de la covarianza del conjunto de entrenamiento C de todas

las formas vermes centradas y $b = b_1 b_2 \; ::: \; b_t$ es el vector de parametros de forma.

Para el ajuste a la imagen, se utiliza un procedimiento iterativo que alinea los hitos h_j de H_i hacia otra forma de nida anteriormente. El conjunto Q, que es una base vectorial de las formas, se combina linealmente con pesos b. Su calculo es a traves del metodo de analisis de componentes principales (Cootes et al., 1995).

La forma resultante se alinea y se proyecta con Q al espacio parametrico b, en donde se puede evaluar si la forma ajustada es probable (se asume que la distribucion de los mismos es normal) y en el caso de no ser se modi ca truncando b para que quede dentro de un rango probable (delimitado en un hiperprisma rectangular en el espacio parametrico entre un 3 para cada componente principal). Luego, se regenera la forma en el espacio de la imagen y se vuelve a ajustar. El ciclo continua hasta que se satisfaga algun criterio de convergencia previamente de nido.

2.5. Alineamiento de formas vectoriales

El procedimiento de alineamiento de dos formas vermes es a traves de la ecuacion:

$$h_j = T(s;)[h_j] + t \tag{2.25}$$

donde T (;s) es la funcion de rotacion y escalamiento determinados respectivamente por y s, y $t = [\, t_x \,;\, t_y \,]^T$, donde t es vector de traslacion.

El alineamiento de dos formas utilizado en modelos basados en hitos de frontera es una tecnica de ajuste de posicion que minimiza las distancias entre los h_j correspondientes de una forma vectorial H_1 a otra ja H_2, esto es:

$$\mathrm{m\;n}_{;s;t_x;t_y} \sum_{j=0}^{n} \sqrt{(x_{1j} \quad x_{2j})^2 + (y_{1j} \quad y_{2j})^2} \tag{2.26}$$

donde H_1 y H_1 se de nen en (2.22).

Cootes y Taylor (1993) proveen un metodo de alineamiento de H_1 y H_2 respecto a los parametros , s, t_x, t_y. Este consiste en, dados los correspondientes hitos $h_j = \; x_{1j} \quad y_{1j}$ ^{T} de H_1 y los hitos $h_j = \; x_{2j} \quad y_{2j}$ ^{T} de H_2, minimizar su distancia ‘ $_2$ usando los parametros de angulo de rotacion , el factor de escala s y el vector de traslacion $t = \; t_x \quad t_y$ ^{T}

mediante la ecuacion:

$$E = PWP^T \tag{2.27}$$

donde

$$P = \begin{bmatrix} x_{1j} & y_{1j} \end{bmatrix}^T \quad M(s;) \begin{bmatrix} x_{2j} & y_{2j} \end{bmatrix}^T \quad t \tag{2.28}$$

$$M(s;) \begin{bmatrix} x_{kj} \\ y_{kj} \end{bmatrix} = \begin{bmatrix} s\cos( )x_{kj} \quad s\sin( )y_{kj} \\ s\sin( )x_{kj} + s\cos( )y_{kj} \end{bmatrix} \tag{2.29}$$

$j = 0 ; \quad ; n \;\; 1$ y W es una matriz diagonal de pesos w_i de cada punto.

Con las sustituciones $a_x = s\cos( )$ y $a_y = s\sin( )$ y por el metodo de m nimos cuadrados (diferenciando respecto a las variables a_x, a_y, t_x y t_y) se obtiene el siguiente sistema de ecuaciones lineales:

$$\begin{bmatrix} Y_2 & X_2 & 0 & W \\ X_2 & Y_2 & W & 0 \\ 0 & Z & Y_2 & X_2 \\ Z & 0 & X_2 & Y_2 \end{bmatrix} \begin{bmatrix} a_y \\ a_x \\ t_x \\ t_x \end{bmatrix} = \begin{bmatrix} Y_1 \\ X_1 \\ C_2 \\ C_1 \end{bmatrix} \tag{2.30}$$

y resuelto por el metodo de descomposicion por valores singulares, dando por resultado:

$$X_i = \sum_{k=0}^{n-1} w_k x_{ik} \qquad Y_i = \sum_{k=0}^{n-1} w_k y_{ik} \tag{2.31}$$

para i = 1; 2, y

$$\begin{aligned} Z &= \sum_{k=0}^{n-1} w_k \left(x_{2k}^2 + y_{2k}^2\right) & W &= \sum_{k=0}^{n-1} w_k \\ C_1 &= \sum_{k=0}^{n-1} w_k \left(x_{1k} x_{2k} + y_{1k} y_{2k}\right) & C_2 &= \sum_{k=0}^{n-1} w_k \left(y_{1k} x_{2k} - x_{1k} y_{2k}\right) \end{aligned} \tag{2.32}$$

El proceso anterior es conocido como analisis de Procrusto.

2.6. Curva de interpolacion polinomica parametrica

Foley y Nielson (1989) se~nalan que el metodo de interpolacion polinomica parametrica permite calcular curvas parametricas, denotadas C_{sp}, a partir de un conjunto nito de hitos conocidos. Este metodo se basa en el de interpolacion segmentaria polinomica B-trazadores (tambien denominada B-spline) (Atkinson, 2008).

La curva C_{sp} se determina a partir de un conjunto de n puntos de control c_i, polinomios trazadores de grado d, y una sucesion de numeros reales $t = (t_i)_{i=1}^{n+d+1}$. Entonces la curva que pasa por dichos hitos viene dada por:

$$C_{sp}(t) := \sum_{i=d+1}^{n} c_i B_{i;d}(t) \tag{2.33}$$

donde la base $B_{i;d}$ se de ne recursivamente por:

$$B_{i;d}(t) = \frac{t - t_i}{t_{i+d} - t_i} B_{i;d-1}(t) + \frac{t_{i+1+d} - t}{t_{i+1+d} - t_{i+1}} B_{i+1;d-1}(t) \tag{2.34}$$

con

$$B_{i;0}(t) = \begin{cases} 1 & \text{si } t_i \le t < t_{i+1} \\ 0 & \text{en otro caso} \end{cases} \tag{2.35}$$

con $t \in \mathbb{R}$.

En el caso de las curvas parametricas en IR^2 y para el conjunto de n hitos $h_i = [x_{ij} \; y_{ij}]^T$, la curva viene dada por:

$$C_{sp}(t) = \sum_{i=1}^{n} h_i B_{i;d}(t) = \left[\sum_{i=1}^{n} x_{ij} B_{i;d}(t) \quad \sum_{i=1}^{n} y_{ij} B_{i;d}(t) \right]^T \tag{2.36}$$

En la gura 2.6 se observa, dada una colocacion arbitraria de 40 hitos o puntos de control coloreados en rojo, una curva C_{sp} resultante de aplicar el metodo de interpolacion polinomica parametrica.

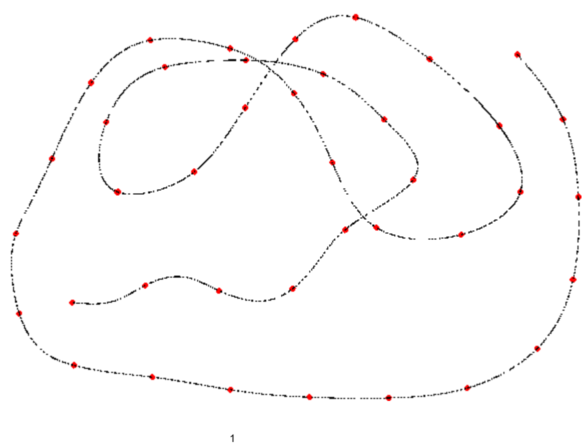


Figura 2.6: Ejemplo de una curva C_{sp} calculada por el metodo de interpolacion polinomica parametrica con 40 hitos colocados arbitrariamente coloreados en rojo.

2.7. S mplexes geometricos

Para la generacion de nuevas formas vermes a partir de las conocidas en la base de datos de entrenamiento C, se utiliza el concepto de los n-s mplexes geometricos (Boyd y Vandenberghe, 2004). Estos son envolturas convexas determinadas por n + 1 nodos linealmente independientes dados por:

$$v = [v_0 \quad v_1 \quad \dots \quad v_n] \in IR^n \tag{2.37}$$

La region interna de los n-s mplexes S_n es un conjunto in nito de puntos de m dimensiones de nidos por:

$$S_n := \left\{ s_i \in IR^m \mid s_i = \sum_{i=0}^{n-1} \lambda_i v_i ; \; \lambda_i \geq 0 ; \; 1^T \lambda = 1 \right\} \tag{2.38}$$

con $\lambda := [\lambda_0 \; \lambda_1 \dots \lambda_{n-1}]^T$. Los valores λ_i $(0 \leq i \leq n-1)$ son denominados coordenadas baricentricas de s_i. Ademas, los nodos del n-s mplex se denominan vertices.

En la gura 2.7 se ilustra un 3-s mplex y un elemento s_i que pertenece a este.

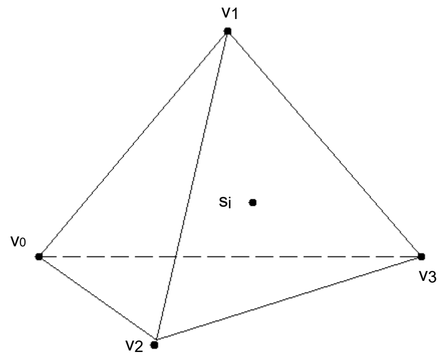


Figura 2.7: Un elemento $s_i \in S_3$.

La generacion de nuevas formas vermes en este trabajo se realiza a traves de caminatas aleatorias entre n-s mplexes vecinos. Una caminata aleatoria es un proceso aleatorio que describe un camino a traves de una sucesion de pasos aleatorios $\{P_n\}_{n \in \mathrm{IN}}$ en algun espacio matematico (Jukna, 2011; Lawler y Limic, 2010). En este trabajo, el espacio corresponde a IR^{80} y cada elemento generado por la caminata aleatoria es una forma verme valida de 80 dimensiones.

En una caminata aleatoria, cada paso es representado como un salto de un n-s mplex a otro. Ademas, el conjunto de saltos entre n-s mplexes vecinos forman una ruta y las probabilidades que salte a cada uno de sus vecinos son las mismas. De este modo, la caminata aleatoria se da sobre una variedad topologica de m dimensiones, aproximada por la estructura de n s mplexes vecinos.

2.8. Optimizacion multiobjetivo

En un problema de decision cuya solucion optima involucra dos o mas funciones es comun aplicar los metodos de optimizacion multiobjetivo. Estos determinan un conjunto de valo-res que son optimizados de forma simultanea a partir del comportamiento de m funciones de costo o aptitud. Formalmente, se de ne optimizacion multiobjetivo como:

$$\min \{F_0(\vec{x}), F_1(\vec{x}), F_2(\vec{x}), \ldots, F_m(\vec{x})\} \tag{2.39}$$

sujeto a $\vec{x} \in S$, que envuelve m funciones objetivo de costo[1] de nidas por: $F_i : \mathrm{IR}^n \to \mathrm{IR}$ las cuales se desean minimizar (o maximizar para problemas de optimizacion de aptitudes) simultaneamente (Coello et al., 2004).

[1] Sin perdida de generalidad, los conceptos desarrollados son validos para problemas de optimizacion de aptitudes.

El vector decision corresponde a:

$$\sim x = [x_0\ x_1\ \ :::\ x_n]^T \tag{2.40}$$

y pertenece una region factible S 2 IR^n, donde la region factible esta delimitada por un conjunto de funciones de restriccion previamente determinadas.

El proceso de ajuste de una forma verme $\hat{f}_i$ a un nematodo en una imagen digital se realiza a traves de un proceso de optimizacion multiobjetivo de costos utilizado en dominancia de Pareto (Marler y Arora, 2004).

2.8.1. Dominancia de Pareto

Sea F := fF_1; F_2; : : : ; F_mg un conjunto de funciones objetivo y S una region factible. Se de ne dominancia de Pareto ~x_1 sobre ~x_2, denotado ~x_1 ~x_2, si un vector ~x_1 2 S domina a otro vector ~x_2 2 S (Abbass et al., 2001). Esto ocurre si se cumplen a la vez las siguientes condiciones:

- El vector de decision ~x_1 no es peor que ~x_2 en todas las funciones objetivo. Es decir:

  $$F_i(\sim x_1)\quad F_i(\sim x_2);\ 8i;\ 1\quad i\quad m$$

- El vector decision ~x_1 es estrictamente mejor que ~x_2 en al menos una funcion objetivo. Es decir:

  $$F_i(\sim x_1) < F_i(\sim x_2)$$

  en al menos un i, 1 i m.

Por otra parte, un vector ~x_1 2 S domina debilmente a ~x_2 2 S, denotado ~x_1 ~x_2, si y solo si:

- El vector decision ~x_1 no es peor que ~x_2 en todos los objetivos. Es decir:

  $$F_i(\sim x_1)\quad F_i(\sim x_2);\ 8i;\ 1\quad i\quad m$$

  .

El conjunto optimo de Pareto, denotado CP, es el conjunto de todos los vectores ~x_1 2 S tales que no existe otro vector ~x_2 2 S que lo domina. Se de nen por:

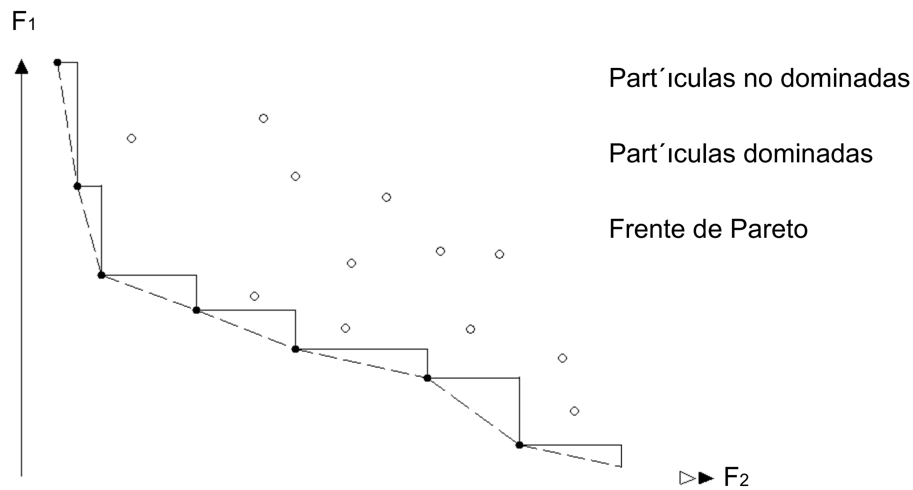
$$CP := f\sim x_1\ 2\ S\ j\ @\sim x_2\ 2\ S \text{ tal que } \sim x_1\quad \sim x_2 g \tag{2.41}$$

Los elementos de este conjunto corresponden a los vectores decision asociados a CP.

El frente optimo de Pareto, denotado FP, es el conjunto de todos los vectores que son la imagen de los vectores decision. Se de nen por:

$$FP := \{\vec{F} = [F_1(\vec{x})\ F_2(\vec{x}) : : : F_k(\vec{x})]^T \mid \vec{x} \in CP\} \qquad (2.42)$$

Los elementos del conjunto FP se denominan vectores objetivo. En la gura 2.8 se muestra el concepto de dominancia de Pareto.



1

Figura 2.8: Dominancia de Pareto para dos funciones F_1 y F_2. Los c rculos con relleno de color negro corresponden a part culas no dominadas y los que no tienen relleno corresponde a part culas dominadas. La l nea negra continua representa el frente de Pareto.

2.8.2. Optimizacion por enjambre de part culas

Optimizacion por enjambre de part culas (PSO por sus siglas en ingles) es un metodo heur stico propuesto por Eberhart y Kennedy (1995) que se basa en la simulacion del movimiento de enjambres capaces de localizar un valor optimo en un espacio de busqueda $S \subseteq IR^n$.

Cada part cula del enjambre se inicializa aleatoriamente y se mueve en S para localizar la mejor posicion de acuerdo con una funcion de ajuste F.

Asumiendo m part culas en S, la posicion de la i-esima part cula se representa por x_i y su vector velocidad actual es v_i. Los movimientos de las part culas son controlados por F hasta localizar la mejor posicion local x_{pbest} , y entre todas las posiciones locales, se localiza la mejor posicion global x_{gbest} .

Cada x_i es evolucionado hacia su nueva posicion a traves de un proceso iterativo dado por las ecuaciones:

$$v_i(t+1) := w\, v_i(t) + c_1\, r_1\, (x_{pbest} \quad x_i(t)) + c_2\, r_2\, (x_{gbest} \quad x_i(t)) \tag{2.43}$$

y

$$x_i(t+1) := x_i(t) + v_i(t+1) \tag{2.44}$$

donde t es el parametro de iteracion, w es el factor de inercia sobre $v_i(t \quad 1)$, c_1 y c_2 factores de aceleracion de las part culas local y global respectivamente y, $r_{1;t}$ y $r_{2;t}$ son numeros aleatorios en el intervalo]0;1[que dependen de la iteracion t.

La actualizacion de x_{pbest} es $x_i(t)$ si se cumple que $F(x_{pbest}) > F(x_i(t))$. La actualizacion x_{gbest} es x_{pbest} si se cumple que $F(x_{pbest}) > F(x_{gbest})$. El ciclo se sigue hasta algun criterio prede nido o un numero de iteraciones N_{it} prede nido. El algoritmo del P SO se muestra en la gura 2.9.

Entrada: $\vec{x}_i \leftarrow$ localización de la partícula *i*
$\vec{v}_i \leftarrow$ velocidad de la partícula *i*
$\vec{x}_{pbest} \leftarrow$ Mejor posición local inicial
$\vec{x}_{gbest} \leftarrow$ Mejor posición global inicial

mientras $it < N_{it}$ **hacer**
/* N_{it} número de iteraciones */
  Para cada $\vec{x}_i$ *de* S **hacer**
    Para cada $t \in \{1, 2, \ldots, n\}$ **hacer**
      $\vec{v}_i(t+1) \leftarrow w\, \vec{v}_i(t) + c_1\, r_{1,t}\, (\vec{x}_{pbest} - \vec{x}_i(t)) + c_2\, r_{2,t}\, (\vec{x}_{gbest} - \vec{x}_i(t))$
      $\vec{x}_i(t+1) \leftarrow \vec{x}_i(t) + \vec{v}_i(t+1)$
    fin
  fin
  Para cada $\vec{x}_i$ *de* S **hacer**
    si $F(\vec{x}_i) < F(\vec{x}_{pbest})$ **entonces**
      $\vec{x}_{pbest} \leftarrow \vec{x}_i$ /* Actualizar $\vec{x}_{pbest}$ */
    fin
    si $F(\vec{x}_{pbest}) < F(\vec{x}_{gbest})$ **entonces**
      $\vec{x}_{gbest} \leftarrow \vec{x}_{pbest}$ /* Actualizar $\vec{x}_{gbest}$ */
    fin
  fin
  $it \leftarrow it + 1$
fin

Figura 2.9: Algoritmo del P SO.

2.8.3. Optimizacion por enjambre de part culas multiobjetivo

Por otra parte, el metodo de Optimizacion por enjambre de part culas multiobjetivo es una extension del metodo P SO (seccion 2.8.2) con multiples funciones objetivo (Coello et al., 2004), de nido por:

$$F := \{F_0 ; F_1 ; \ldots ; F_n\} \tag{2.45}$$

para un enjambre de part culas P.

Este algoritmo denotado M OP SO (por sus siglas en ingles) inicia tomando los conjuntos de part culas P y de velocidades V en forma aleatoria y, almacena en R los elementos no dominados de P que conforman el conjunto del frente Pareto (gura 2.8).

Luego, se itera para evolucionar las posiciones de las part culas almacenadas en R por otras nuevas usando (2.44) y para las velocidades calculadas se usa (2.43). En cada iteracion, las part culas de P son truncadas en un hipercubo y se reemplazan las part culas por otras no dominadas en R. El ciclo se sigue hasta algun criterio prede nido o un numero N_{it} de iteraciones prede nido. El algoritmo de este metodo se muestra en la gura 2.11. Asimismo, un frente de Pareto generado por el M OP SO se visualiza en la gura 2.10.

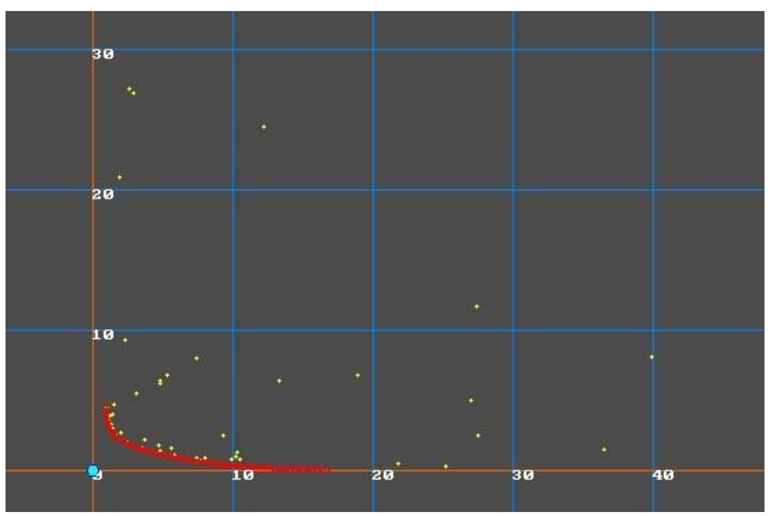


Figura 2.10: Aplicacion del algoritmo del M OP SO para dos funciones de costo. En color rojo se marcan las part culas del conjunto frente de Pareto, y en amarillo las part culas dominadas.

```
Entrada: P ← {~x0, ~x1 , . . . , ~xm}                                  /* Población */
         V ← {~v0, ~v1 , . . . , ~vm}                                  /* velocidades */
         R←{}                                                          /* Repositorio */
Para cada ~xi ∈ P hacer
    si (~xi   ~u, ∀~u ∈ R) entonces
        R ← {~u ∈ R : ~u  ~xi}                  /* Elimina los u~ ∈ R dominados por ~xi */
        R ← R ∪ {~xi}                            /* Agrega ~xgbest al repositorio */
    fin
    si ~x_i  ~x_pbest  ∨  (~xi  ~xpbest) ∧ (~xpbest  ~xi)  entonces
        pbest ← ~xi                              /* Actualización del  pbest */
        ~x                                                 ~x
    fin
fin
mientras it < Nit hacer
                                                 /* N es el número de iteraciones */
    Para cada ~xi ∈ P hacer
        Para cada t ∈ {1, 2, . . . , n} hacer
            ~vi(t + 1) ← w ~vi(t) + c1 r1,t (~xpbest − ~xi(t)) + c2 r2,t (~xgbest − ~xi(t))
            ~xi(t + 1) ← ~xi(t) + ~vi(t + 1)
        fin
        P, V ← check(P, V)                       /* Chequear fronteras */
        si (~xi   ~u, ∀~u ∈ R) entonces
            R ← {~u ∈ R : ~u  ~xi}              /* Elimina los u~ ∈ R dominados por ~xi */
            R ← R ∪ {~xi}                        /* Agrega ~xgbest al repositorio */
        fin
        pbest ←        i                                   pbest
        si ~x_i  ~x_pbest  ∨  (~xi  ~xpbest) ∧ (~xpbest  ~xi)  entonces
            ~x        ~x                         /* Actualización del ~x */
        fin
    fin
    it → it + 1
fin
```

Figura 2.11: Algoritmo del M OP SO.

2.9. Metricas de evaluacion

Las metricas de evaluacion son metodos que tienen como objetivo validar de forma es-tad stica el desempe~no y abilidad de los algoritmos. En las seccion 2.9.1 y 2.9.2 se aborda la conceptualizacion de dos metricas asociadas a las areas de procesamiento digital de imagenes y reconocimiento de patrones: validacion cruzada de k-iteraciones y el ndice de Jaccard.

2.9.1. Validacion cruzada de k-iteraciones

La tecnica de validacion cruzada de k-iteraciones es un estad stico de evaluacion de al-goritmos que comienza con una particion de la base de datos C en k subconjuntos inde-pendientes C_i (Refaeilzadeh et al., 2009). Con ellos se determinan de forma cuantitativa resultados a traves de alguna medida estad stica como la ra z del error medio cuadrati-co (E_{rr}) (Willmott y Matsuura, 2005). En la gura 2.12 se muestra un esquema de esta tecnica de validacion.

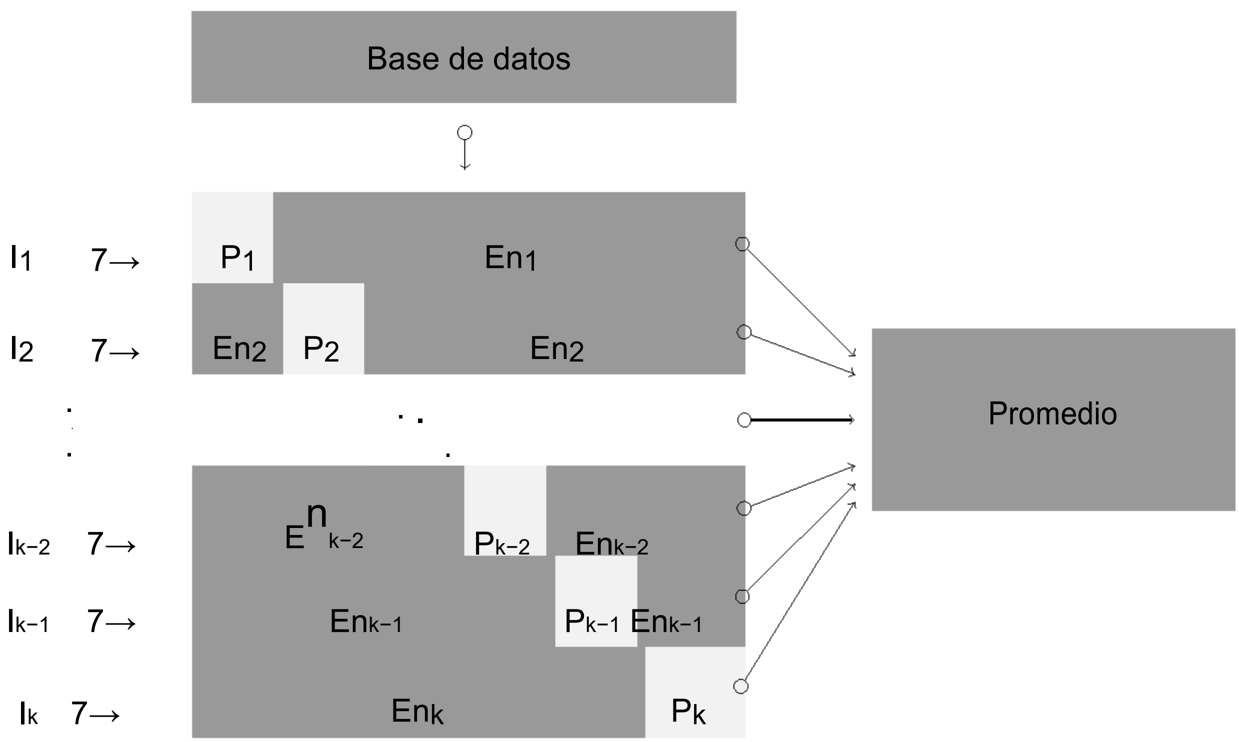


Figura 2.12: Esquema de la tecnica de validacion cruzada de k iteraciones. En_i y P_i corresponden los i esimos conjuntos de entrenamiento y prueba respectivamente, para la i-esima iteracion I_i.

Las k particiones se utilizan para determinar los i-esimos conjuntos de entrenamiento (En_i) y pruebas (P_i). Para la ejecucion de la i-esima iteracion experimental de este metodo, se escogen los conjuntos entrenamiento y prueba de la siguiente forma:

$$En_i := \{C_1; C_2; \ldots; C_k\} \setminus \{C_i\} \;\; y \;\; P_i := \{C_i\} \; (1 \leq i \leq k) \tag{2.46}$$

es decir, se utilizan k 1 conjuntos para entrenar En_i y un conjunto para pruebas P_i. Las particiones a su vez deben cumplir:

-
- La union de todas las particiones resultan la base de datos, es decir, $C = \bigcup_{i=1}^{k} C_i$.
- Las particiones C_i poseen la misma cardinalidad.

El proceso se repite k veces. Finalmente, se obtiene un promedio general de los resultados de todas las iteraciones. La tecnica de validacion cruzada de k-iteraciones se utiliza en los casos que se cuenta con un numero limitado datos experimentales.

2.9.2. Indice de Jaccard

El ndice Jaccard, denotado J (A;B), es una metrica utilizada para determinar la similitud entre muestras de conjuntos nitos A y B (Real y Vargas, 1996). Al conjunto A se le denomina region verdadera y, al conjunto B se le denomina region calculada. El ndice de Jaccard se de ne por:

$$J(A;B) := \frac{card(A \cap B)}{card(A \cup B)} = \frac{card(A \cap B)}{card(A) + card(B) - card(A \cap B)} \tag{2.47}$$

El ndice de Jaccard provee un valor en el intervalo [0;1], donde J (A;B) = 0:0 indica que no hay similitud entre los conjuntos dados y J (A;B) = 1:0 expresa que ambos conjuntos son totalmente coincidentes. A esta metrica se le denomina Interseccion sobre la union y se denota IoU (por sus siglas en ingles).

En el area de procesamiento digital de imagenes, el ndice Jaccard se utiliza para medir la e cacia de algoritmos relativos a tareas de posicionamiento, segmentacion, deteccion, seguimiento de objetos, as como evaluacion estad stica de los algoritmos a nes a estos (Berman et al., 2018; Rahman y Wang, 2016).

Cap tulo 3

Modelo evolutivo de forma

El modelo evolutivo de forma (MEF) es una tecnica de segmentacion adaptativa de objetos basada en un proceso de evolucion de forma mediante hitos de frontera, enmarcado en un sistema de representacion continua expl cita de representacion discreta-geometrica segun los sistemas de clasi cacion explicados en la seccion 2.4.

En la gura 3.1 se muestra el diagrama de bloques en el que se exhibe la organizacion del modelo evolutivo de forma. En verde se muestran los procesos que conforman el sistema. En color celeste los subprocesos que apoyan al metodo de optimizacion.

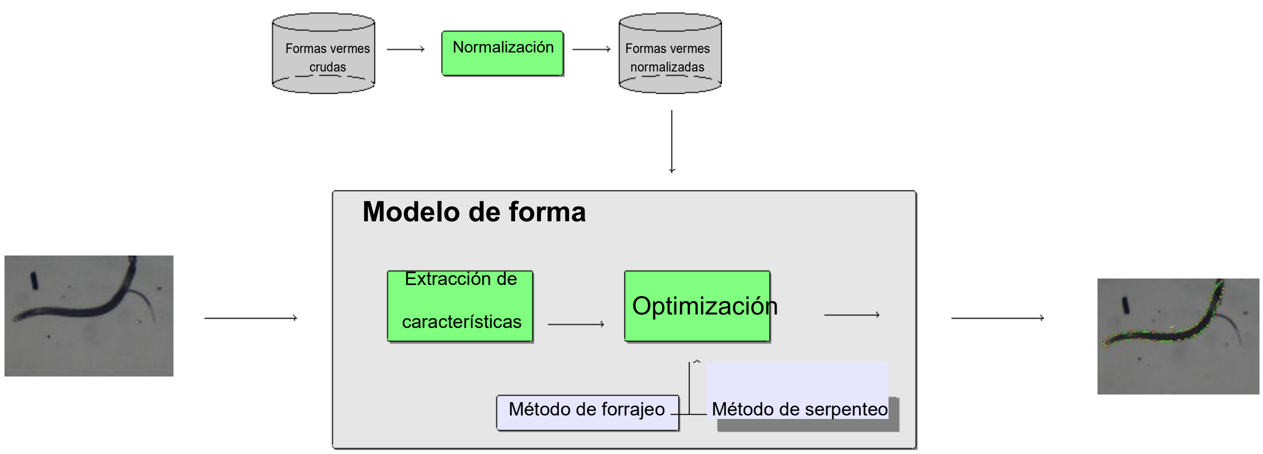


Figura 3.1: Diagrama de bloques del modelo evolutivo de forma.

En concordancia con la estrategia metodologica de la seccion 1.3, en este cap tulo se describe la propuesta de solucion del problema en cuestion fragmentado en tres fases principales. En la primera fase se explica la preparacion de las bases de datos de entrenamiento y pruebas de imagenes digitales y de las siluetas vermes de nidas por hitos de frontera (seccion 3.2).

En la segunda fase se detallan los metodos de generacion de nuevas siluetas a traves de caminatas aleatorias entre n-s mplexes. Asimismo, incluye el metodo de truncamiento de formas vermes a subdominios validos (seccion 3.3.1).

La tercera fase especi ca el metodo del serpenteo y forrajeo, as como las funciones de costo usadas en el M OP SO (secciones 3.4 y 3.6). En esta se incluye el proceso completo
de adaptacion de las forma vermes a los nematodos en imagenes digitales (seccion 3.7).
La ultima fase se abordar en el cap tulo 4.

En la gura 3.2 se muestra un diagrama de ujo con las fases propias de esta investigacion en el orden de aplicacion del modelo.

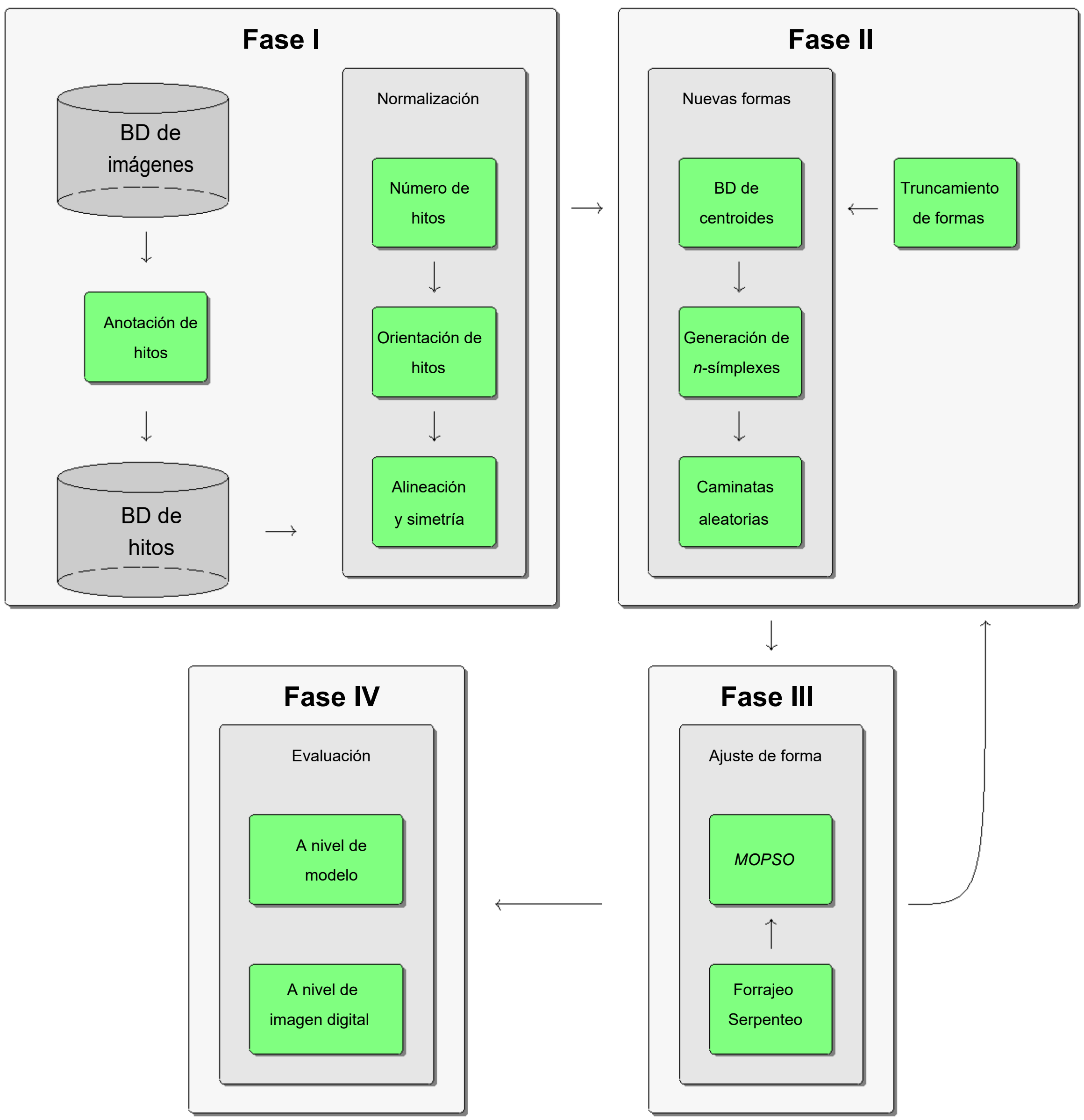


Figura 3.2: Diagrama de los componentes del modelo evolutivo de forma. En este se muestra la secuencia de la solucion planteada al ajuste de forma. En verde se representan los subprocesos de cada fase del modelo.

3.1. Base de datos de imagenes digitales de nematodos

Para las pruebas de entrenamiento y validacion, se utiliza un conjunto de 667 imagenes digitales de diferentes de muestras de nematodos. El tama~no de cada imagen digital obtenida es 421 301 p xeles. Las imagenes se capturaron con una camara Basler A630f adaptada al ocular de un microscopio Nikon Eclipse E200.

A pesar de que las imagenes digitales contienen organismos vermes con una alta varia-bilidad en sus deformaciones, factores de escala y contrastes para el entrenamiento y validacion del metodo modelo de forma propuesto, estas presentan retos reales en cuanto a mostrar escenas ruidosas producto de la presencia de detritos, residuos y sedimentos provenientes de los procesos respectivos de tamizado, y que en consecuencia, limitan el contraste del contorno y cuerpo de los organismos vermes respecto al fondo de la imagen (ver gura 3.3).

Figura 3.3: Imagen de un nematodo con presencia de multiples de detritos.

Asimismo, las imagenes digitales utilizadas en los experimentos se seleccionaron con base a la siguiente lista de restricciones:

- Los niveles de contraste y ruido deben permitir la deteccion de los bordes de cada organismo, aun cuando estos se confundan con otros elementos.
- El cuerpo de cada nematodo debe estar enfocado de modo que permita distinguir el contorno y su cuerpo. El fondo de las imagenes tendra niveles de intensidad altos cercanos a tonalidades de gris claro, mientras que el cuerpo de los nematodos seran con niveles de intensidad bajos similares a grises oscuros.
- La informacion que proporcione la imagen digital en cuanto a colorimetr a es irrelevante para el analisis de forma, debido a que estas contienen poca informacion cromatica, por lo que unicamente se trabajar con imagenes en escala de grises.
- En cuanto al borde de los nematodos en las imagenes digitales, se tomar como premisa que el ancho del contorno del nematodo medira entre 4 y 6 p xeles (gura 3.4).

Se tomar este como el rango valido de posicionamiento entre hitos con las posicio-nes calculadas por modelo de forma y anotados manualmente con la herramienta de colocacion de hitos descrita en el apendice A.

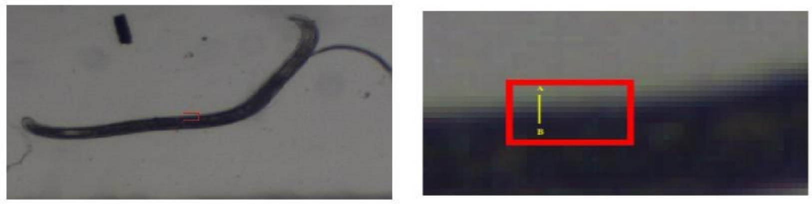


Figura 3.4: En la imagen izquierda se presentan dos nematodos y una parte del contorno enmarcado en un recuadro rojo. En la imagen de la derecha, se amplia el sector del recuadro rojo, y se muestra un segmento de un punto A al punto B, que mide aproximadamente 4.3 p xeles correspondiente al ancho del borde del nematodo.

3.2. Base de datos de formas vermes expresadas en hitos de frontera

La base de datos de entrenamiento y validacion de formas vermes expresadas en hitos de frontera, denotada C, es un conjunto de m instancias de vectores que representan formas vermes f_i, con i 2 f0; ; m 1g. Los hitos de las estancias vermes se denotan h_j, con j 2 f0; ; n 1g.

La obtencion de forma vectorial f_i se realiza a partir de la anotacion manual de hitos h_j sobre el contorno del nematodo, y que en conjunto, moldean su silueta verme de forma secuencial como se muestra en la gura 3.5. El conjunto de hitos se almacena en un vector de datos dado por (2.21).

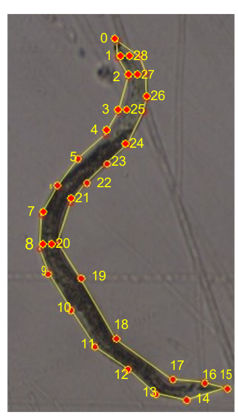


Figura 3.5: Ejemplo de anotacion manual de 29 hitos en secuenciacion sobre el contorno de nematodo en una imagen digital.

En total se demarcaron, por medio de hitos de frontera, 2744 instancias de nematodos con distintos tama~nos que oscilan entre 100 y 800 p xeles de longitud D. La longitud D de cada nematodo se aproxima a traves de la suma de las distancias eucl deas entre vertebras consecutivas de cola a cabeza. En la seccion 3.6.1 se abordar sobre la medida D. En la gura 3.6 se muestra el esqueleto de un nematodo, los hitos de frontera y las vertebras coloreadas en rojo.

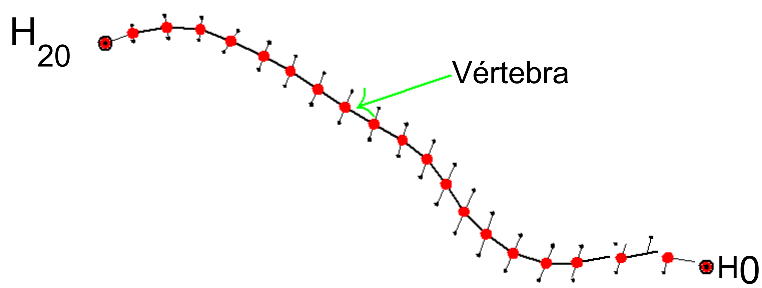


Figura 3.6: Esqueleto de una instancia de nematodo f_i con las vertebras en color rojo.

En la tabla 3.1 se muestra una clasi cacion cuantitativa de instancias vermes de acuerdo a su longitud D de su esqueleto en p xeles.

Tabla 3.1: Tabla de distribucion de instancias de representacion por hitos de nematodos segun longitud D del esqueleto.

Clases	Longitud (en p xeles)
a_1	100 D < 200
a_2	200 D < 300
a_3	300 D < 400
a_4	400 D < 500
a_5	500 D < 600
a_6	600 D < 700
a_7	700 D < 800

Respecto a la tabla 3.1, se construye un histograma (gura 3.7) que resume la distribucion de las longitudes D de las estancias de nematodos medidos en p xeles con los que se realiza los procesos de experimentacion.

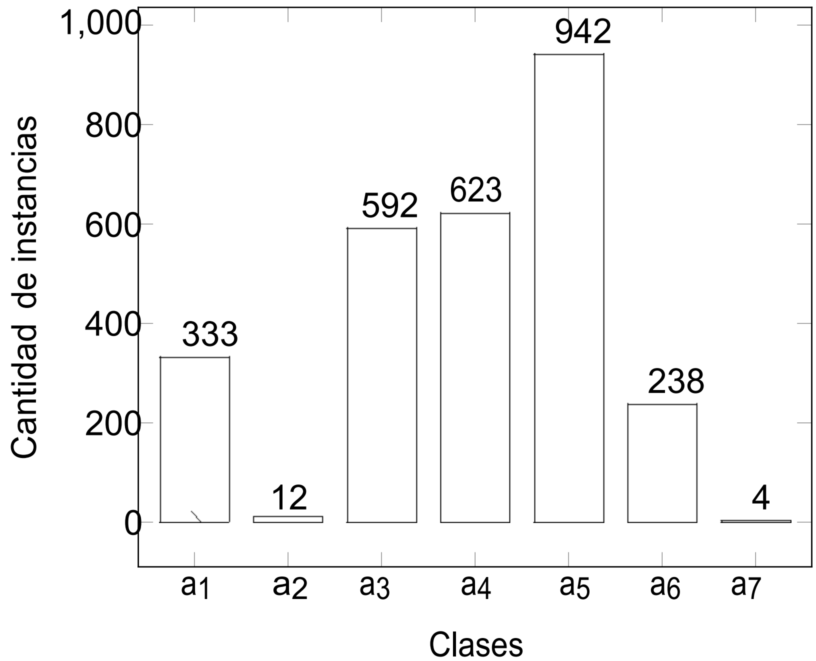


Figura 3.7: Histograma de longitudes D de los esqueletos de estancias de nematodos de la base de datos.

El proceso de la obtencion de la base de datos nal se da en tres subfases: obtencion de la coleccion de imagenes digitales, anotacion manual de hitos de frontera sobre el contorno del nematodo y nalmente la normalizacion de la base de datos (seccion 3.2.1). En la gura 3.8 se ilustra este proceso.

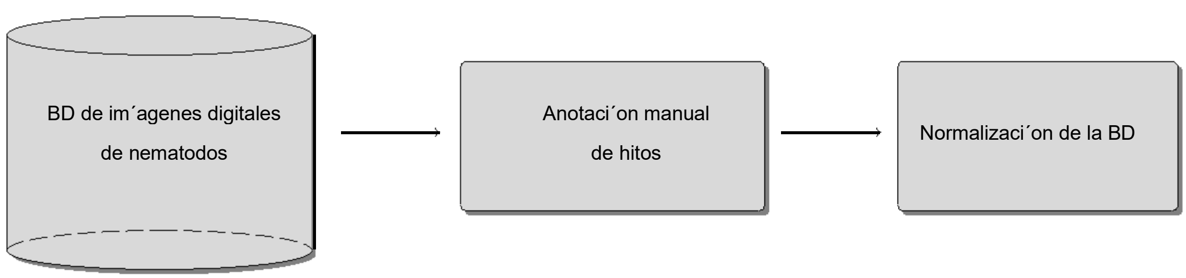


Figura 3.8: Diagrama de ujo de la obtencion de la base de datos.

3.2.1. Normalizacion de la base de datos de hitos de frontera

El proceso de normalizacion se re ere a un ajuste de las estructuras vermiformes con el n de minimizar las tendencias particulares de subconjuntos de datos dominantes sobre las caracter sticas generales de toda la base de datos C, tales como tama~no de las siluetas, numero de hitos de representacion y sentido de orientacion en la secuencia de posiciona-miento de los hitos. Con esto se logra que todos los elementos de C se proyecten a una escala comun.

La normalizacion se divide en tres etapas ordenadas segun su aplicacion: orientacion de la secuencia de hitos para representar la forma verme, numero de hitos y, alineacion y escalado de la silueta. El orden de aplicacion de cada etapa se visualiza en la gura 3.9.

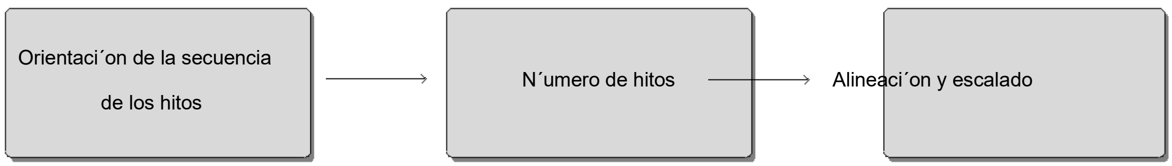


Figura 3.9: Diagrama de bloques de las etapas de normalizacion de la base de datos de entre-namiento y pruebas.

En las siguientes subsecciones se detallan las etapas de la normalizacion de la base de datos.

Etapa i: Orientacion en la secuencia de los hitos: La primera etapa tiene como tarea normalizar la orientacion de la secuencia de hitos de cada instancia vectorial. Esto es posicionar los hitos, en sentido antihorario, en una cadena.

El metodo se basa en una variacion dada en (Garza-Hume et al., 2018; Sedgewick y Wayne, 2011) sobre la interpretacion del resultado numerico del area de un pol gono irregular dada por el matematico Carl Friedrich Gauss[1], y consiste en determinar el sentido de secuenciacion a traves de la constante de orientacion (f_i).

En el caso que (f_i) > 0 se veri ca que la region interna delimitada por una curva, simple y cerrada de n hitos, se encuentra a la izquierda de la trayectoria de anotacion de estos (Levinson y Redhe er, 1975). Dicho trazo se denomina curva de Jordan.

[1] Matematico, astronomo, geobotanico y f sico, nacido en Alemania en el a~no 1777.

El calculo de (f_i) es dado por:

$$(f_i) := \sum_{i=0}^{2n-2} \det \begin{pmatrix} x_i & x_{i+1} \\ y_i & y_{i+1} \end{pmatrix} \tag{3.1}$$

En caso que (f_i) sea negativo, se reacomoda el vector con las coordenadas de los hitos en secuencia inversa (gura 3.10). El proceso de inversion en el orden de los hitos se denota $rev(f_i)$ y se aplica cuando $(f_i) < 0$, y es tal que:

$$rev(f_i) := x_{i\,n-1}\ y_{i\,n-1}\ x_{i\,n-2}\ y_{i\,n-2}\ \ldots\ x_{i0}\ y_{i0} \tag{3.2}$$

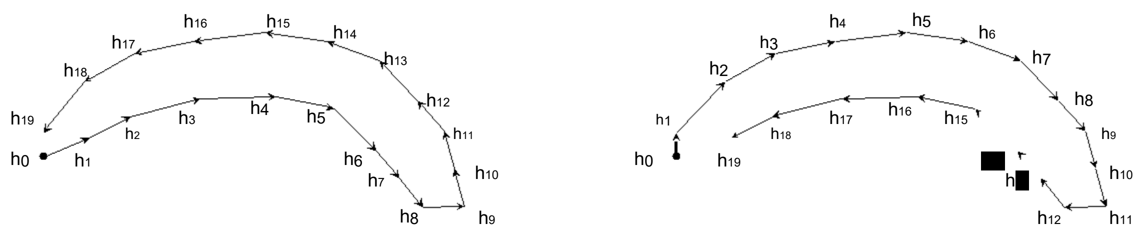


Figura 3.10: Ejemplo de la aplicacion de la etapa I de normalizacion de la secuenciacion de hitos en sentido positivo y negativo respectivamente.

Etapa ii: Numero de hitos equidistante: La segunda etapa del proceso consiste en normalizar todas las instancias vectoriales a un mismo numero m de hitos, equidistantes entre s , de manera que todas las siluetas puedan ser representadas adecuadamente.

La representacion adecuada de una forma verme es una aproximacion de la curva ideal C_r del contorno de la silueta del organismo que es muestreada por hitos de frontera. Es dada por:

y_{ij}

$$c_r := \underline{h}_0 \; \underline{h}_1 \; \underline{h}_{p-1} \quad (3.3)$$

donde $\underline{h}_j$ ($0 \; j < p$) son hitos secuenciados de posicionamiento real. La aproximacion de c_r es una curva de interpolacion c_{spl} de nida por una secuencia de hitos anotados h_j proximos a los $\underline{h}_j$ respectivamente. Cada hito anotado h_j es de nido por:

$$h_j := \underline{h}_j + {}_j \quad (3.4)$$

y, donde se asume que x_{ij} ; y_{ij} N $(0; {}^2)$. Dicho error se atribuye a la anotacion manual de los hitos, y se establece a partir de la diferencia espacial entre las posiciones de los hitos h_j y $\underline{h}_j$.

Asimismo, la curva de aproximacion de la silueta es:

$$c_{spl} := c_r + \tag{3.5}$$

con $:= \; 0 \; 1 \; p \; 1$. Para la estructura vectorial original f_i de p hitos h_j anotados y tomados como puntos de control, la curva interpolante parametrica C_{spl} dada en (2.36) que aproxima la totalidad el contorno de cada forma verme se de ne por:

$$C_{spl}(t) := \left[\sum_{j=0}^{p-1} x_j B_{j;d}(t) \quad \sum_{j=0}^{p-1} y_j B_{j;d}(t) \right]^{\#} \tag{3.6}$$

La curva C_{spl} de una forma verme y sus hitos anotados manualmente se muestran en la gura 3.11.

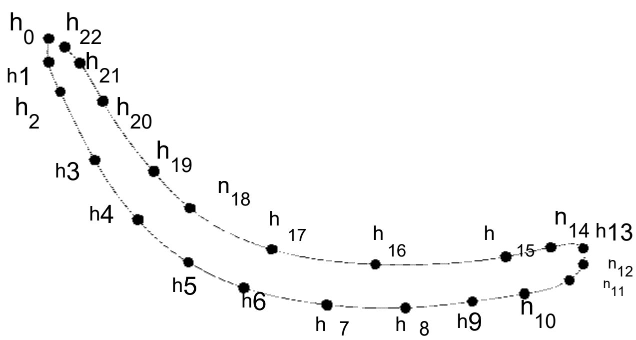


Figura 3.11: Forma de nematodo representado por f_i y su curva interpolante parametrica $C_{spl}(t)$ para 23 hitos.

Luego, el vector normalizado a m hitos, denotado part $(C_{spl}(t); m)$, corresponde a la estancia vectorial de nida en (3.6) que contiene las nuevas posiciones de los hitos consecu-tivamente equidistantes. Viene dado por:

$$\text{part}\,(C_{spl}(t); m) := \hat{x}_{i\,0} \;\; \hat{y}_{i\,0} \;\; \hat{x}_{i\,1} \;\; \hat{y}_{i\,1} \;\; \dots \;\; \hat{x}_{i\,m-1} \;\; \hat{y}_{i\,m-1} \tag{3.7}$$

Ademas, el numero m de hitos a elegir debe proveer informacion de manera que la perdida y redundancia en la informacion que se suministra del contorno del organismo sea m nima. Las particiones son calculadas sobre la curva C_{spl} en m 1 en subcurvas dadas por:

$$C_{spl_0} \; ; C_{spl_1} \; ; \dots ; C_{spl_{m-1}} \tag{3.8}$$

donde en cada subcurva C_{spl_i}, los hitos inicial y nal corresponden respectivamente a $\hat{x}_{ij} \;\; \hat{y}_{ij}$ y $\hat{x}_{ij+1} \;\; \hat{y}_{ij+1}$. La respectiva longitud de arco de cada subcurva viene dada respectivamente por:

$$\ell_{C_{spl_0}} ; \ell_{C_{spl_1}} ; \dots ; \ell_{C_{spl_{m-1}}} \tag{3.9}$$

donde ℓ_{Cspl_i} es dada por el segmento de recta de los hitos inicial y nal de cada subcurva, es decir:

$$\ell_{Cspl_i} \quad q \overline{\quad x_i^2{}_j + y_i^2{}_j \quad} \qquad b \qquad b \tag{3.10}$$

con $xb_{i\,j} := xb_{i\,j+1} \quad xb_{i\,j}$ y $yb_i := yb_{i\,j+1} \quad yb_{i\,j}$.
El criterio de particion corresponde a que la longitud de arco entre subcurvas consecutivas sea la misma. Es decir:

$$\ell_{Cspl0} = \ell_{Cspl1} = \ldots = \ell_{Csplm\ 1} \quad \frac{\sum_{j=0}^{n\ 1} q \overline{x_i^2{}_j + y_i^2{}_j}}{n\ 1} \tag{3.11}$$

En la gura 3.12 se ilustra con un ejemplo la segunda etapa. La curva interpolante respecto a sus hitos normalizados en numero y distancia.

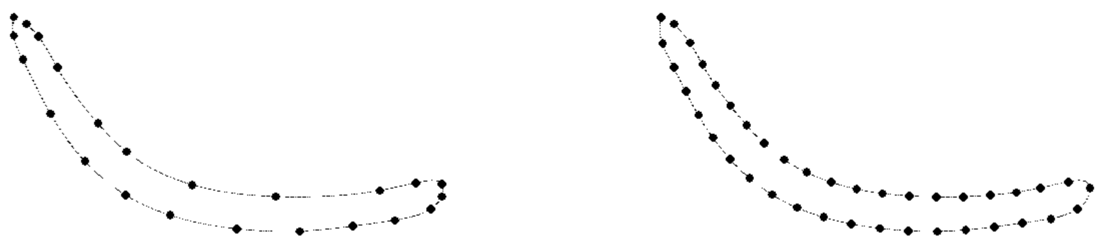


Figura 3.12: Ejemplo de la aplicacion de la etapa II. A la izquierda la curva que delinea un nematodo con hitos a diferentes distancias, y a la derecha la curva con m = 40 hitos de una forma verme f_i aproximadamente equidistantes entre s .

Etapa **III**: Alineacion y escalado: La ultima etapa es el proceso de alineacion de todos los elementos f_i del conjunto de entrenamiento respecto a una forma verme $\bar{f}$ usada como referencia.

Este es un metodo que optimiza el angulo de rotacion , el factor de escala s y el vector de traslacion $t_{xy} = \left[t_x \ t_y \right]^T$ de cada forma f_i de C de manera que la distancia entre los sea m nima (gura 3.13).

El proceso es denotado con align f_i ; $\bar{f}$ y se utiliza el metodo detallado en la seccion 2.5.

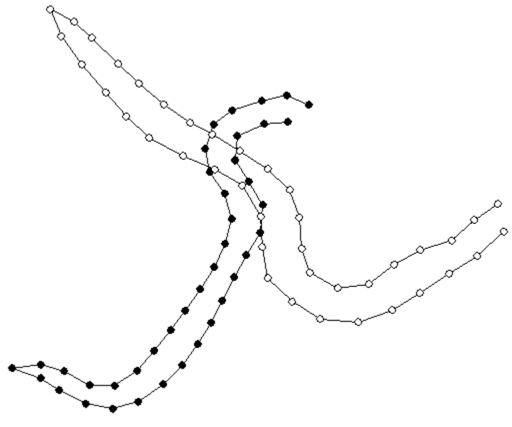


Figura 3.13: Ejemplo de la aplicacion de la fase III de alineacion y escalado. A la izquierda dos formas vermes (una ellas es la de referencia en la que se representa sus hitos con c rculos con un relleno de color negro), y a la derecha la forma verme alineada a la de referencia respecto a la posicion, escala y tama~no.

3.2.2. Aumento de cardinalidad de la base de datos por simetr a de forma

Despues de la etapa de normalizacion, el ultimo proceso consiste en reejar cada elemento del conjunto de entrenamiento C a traves de una transformacion lineal isometrica R en la que a cada forma verme f_i se le asocia otra $f_i^{(r)}$ de modo que los hitos respectivos se se ubiquen a igual distancia de un eje simetrico (gura 3.14). El proceso de reexion por simetr a de forma de f_i es denotado sim $f_i^{(r)}$. Cada instancia nueva $f_i^{(r)}$ es agregada a C.

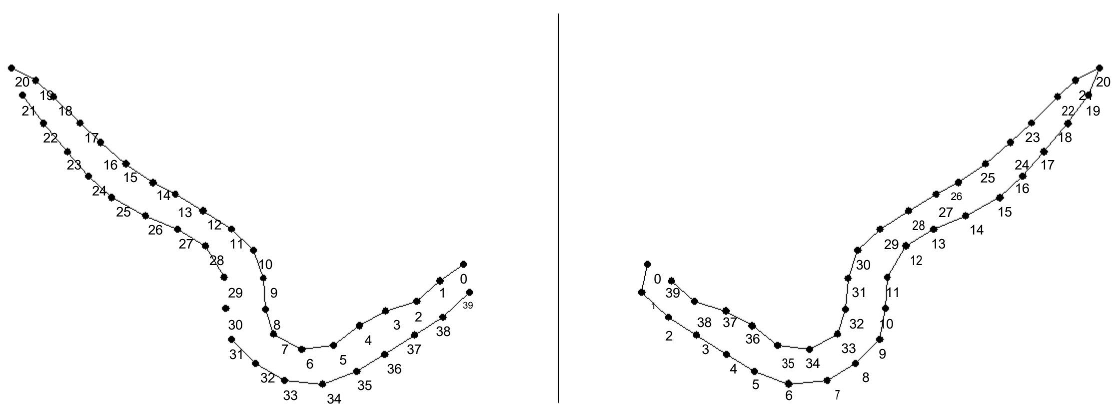


Figura 3.14: Ejemplo de la aplicacion de la normalizacion de simetr a de forma de f_i y $f_i^{(r)}$ respecto a un eje vertical.

En adelante, una forma verme f_i tratada bajo las tres etapas se denota $\hat{f}_i$, y este proceso se denomina pre-procesamiento de formas vermes.

Asimismo, despues de las fases de normalizacion y aumento de cardinalidad de la base de datos por simetr a de forma, el conjunto de datos entrenamiento con r instancias vermes y cada una de estas con m hitos, se denota por $\hat{C}$.

Matricialmente, $\hat{C}$ se representa por:

$$\hat{C} = \begin{bmatrix} \hat{f}_0 \\ \hat{f}_1 \\ \vdots \\ \hat{f}_r \end{bmatrix} = \begin{bmatrix} x_{0\,0} & y_{0\,0} & x_{0\,1} & y_{0\,1} & \cdots & x_{0\,m-1} & y_{0\,m-1} \\ x_{1\,0} & y_{1\,0} & x_{1\,1} & y_{1\,1} & \cdots & x_{1\,m-1} & y_{1\,m-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{r\,0} & y_{r\,0} & x_{r\,1} & y_{r\,1} & \cdots & x_{r\,m-1} & y_{r\,m-1} \end{bmatrix} \quad (3.12)$$

donde cada forma verme representa una la de $\hat{C}$. El algoritmo de normalizacion se muestra en la gura 3.15.

Entrada: C /* Conjunto de datos de entrenamiento */
$\bar{f}$ /* Forma promedio verme */

Salida : $\hat{C}$ /* Conjunto de datos de entrenamiento normalizado */

Para cada f_i *de* C **hacer**
  si $\rho(f_i) < 0$ **entonces**
    $f_i \leftarrow rev(f_i)$ /* Cambio de orientación de los H_j */
  fin
  $f_i \leftarrow part(C_{spl}(f_i), m)$ /* Partición de *m* hitos equidistantes */
  $f_{(r)} \leftarrow align(sim(f_i), \bar{f})$ /* Cálculo y alineamiento de la forma simétrica */
  $\hat{f}_i \leftarrow \hat{f}_i, \bar{f}$
  $\hat{C} \leftarrow \hat{C} \cup \{\hat{f}_i, \hat{f}_i^{\,o}\}$ /* Agrega dos elemento al conjunto de entrenamiento */
fin

Figura 3.15: Algoritmo de pre-procesamiento de la base de datos de entrenamiento $\hat{C}$ de formas vermes representados por hitos de frontera.

3.3. Metodo del serpenteo

El serpenteo es un metodo que permite generar secuencias de siluetas vermes emulando y adoptando de nuevas formas. Cada instancia del conjunto se expresa en vectores de hitos de frontera. Los movimientos continuos se consiguen a traves de caminatas aleatorias entre subdominios vermiformes. En las secciones 3.3.1, 3.3.2 y 3.3.3 se detalla el proceso.

En la gura 3.16 se muestra el diagrama de las etapas del metodo del serpenteo.

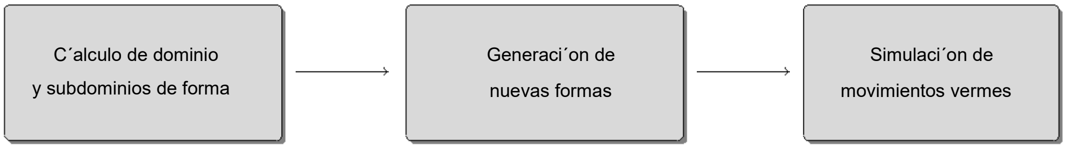


Figura 3.16: *Diagrama de ujo del metodo del serpenteo.*

3.3.1. Dominio y subdominio de formas vermes

El dominio de formas validas vermes, denotado D_f , se de ne como un conjunto no vac o de cardinalidad in nita en el que cada elemento $\hat{f}_i$ 2 D_f es una representacion de una forma verme valida. Cada elemento $\hat{f}_i$ 2 D_f es un elemento de IR^{2m}; ademas, se asume que un dominio de forma D_f es una variedad topologica n dimensional empotrada en el conjunto IR^{2m} con $n < 2m$.
Asimismo, un subdominio de formas permitidas, denotado S_{D_f} , es un conjunto de indi-viduos $\hat{f}_i$ de una particion del dominio de formas vermes D_f que, por medio de algun estad stico de proximidad o distancia, se agrupan por similitud de forma.
As , una forma particular de aproximar subdominios de forma es por medio de hiper-volumenes delimitados por n-s mplexes (seccion 2.7) donde los nodos que engendran di-cho conjunto son formas vermes a nes en su silueta. La region interna de los subdominios contienen in nitas instancias vermes validas generadas a traves de los nodos.

Los nodos $\hat{f}_i$ de los subdominios de forma se obtienen usando el algoritmo de k-medias dado por (Wagsta et al., 2001). Cada nodo $\hat{f}_i$ es un centroide c_i de una particion realizada sobre el dominio de forma D_f . Ademas, cada c_i representa una forma signi cativamente distinta sobre los restantes de las otras particiones del dominio de formas D_f .
En la gura 3.17 se muestra el proceso de obtencion de los subdominios de forma S_{D_f} . La primera parte muestra un conglomerado de puntos, cada uno representa una forma verme $\hat{f}_i$. Luego, se calculan los centroides de los agrupamientos dados por la particion, y nalmente, con estos se construyen los S_{D_f} que representan n-s mplexes en el espacio de datos.

Cabe destacar que todas las formas $\hat{f}_i$ que engendran los subdominios de forma S_{D_f} son alineados a una forma verme $\bar{f}$ segun el metodo descrito en la seccion 2.5. Ademas, si dos o mas subdominios al menos poseen un vertice en comun, se les denomina subdominios vecinos.

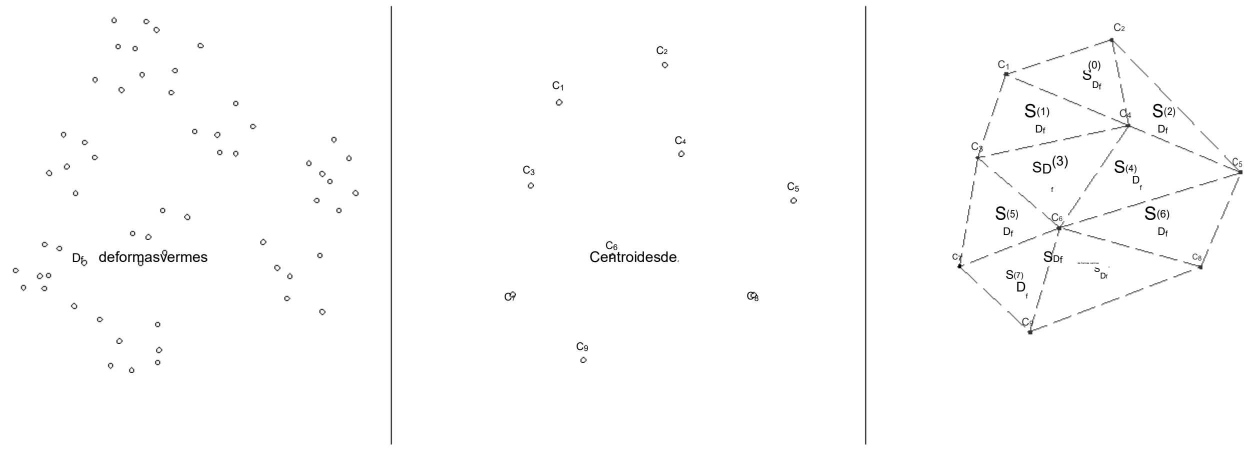


Figura 3.17: Proceso de obtencion de S_{D_f} a traves del algoritmo de k-medias. La primera parte del proceso (a la izquierda) muestra el conglomerado de datos, la segunda exhibe los centroides, y la ultima los subdominios de forma.

3.3.2. Generacion de nuevas formas vermes en un S_{D_f}

Una nueva forma verme $\hat{f}^{(i)}$ pertenece en un subdominio S_{D_f} de nido por un hipervolumen delimitado por un n s mplex de n + 1 nodos a traves de:

$$\hat{f}^{(i)} \in S_{D_f} \quad , \hat{f}^{(i)} = {}_0\hat{f}_0 + {}_1\hat{f}_1 + {}_2\hat{f}_2 + \dots + {}_n\hat{f}_n = \sum_{=0}^{n} {}_i\hat{f}_i \tag{3.13}$$

donde los $\hat{f}_0; \hat{f}_1; \dots; \hat{f}_n$ son vertices del n s mplex dados por los centroides que generan el subdominio S_D (Arroyo y Alvarado, 2020). Ademas, las coordenadas baricentricas $_0; _1; \dots; _n \, 0$ son coe cientes reales tales que $\sum_{j=0}^{n} {}_j = 1$.

3.3.3. Simulacion de movimientos aleatorios vermes

El metodo del serpenteo consiste en movimientos vermes simulados por medio de una caminata aleatoria (Woess, 2000) a traves de una sucesion de pasos aleatorios

$$(P_0; P_1; \dots; P_n)_{n \in \mathbb{N}} \tag{3.14}$$

entre subdominios vecinos que deforman gradualmente la silueta original. Cada paso P_i representa un cambio de subdominio (gura 3.18).

El recorrido entre subdominios vecinos posibilita que cada elemento generado $\hat{f}^{(i)}$ conserve informacion de forma de al menos uno de los vertices en comun permitiendo que la sucesion de las p + 1 nuevas siluetas vermes:

$$\hat{f}^{(0)}; \hat{f}^{(1)}; \hat{f}^{(2)}; \dots; \hat{f}^{(p)} \tag{3.15}$$

cambien de forma paulatina.

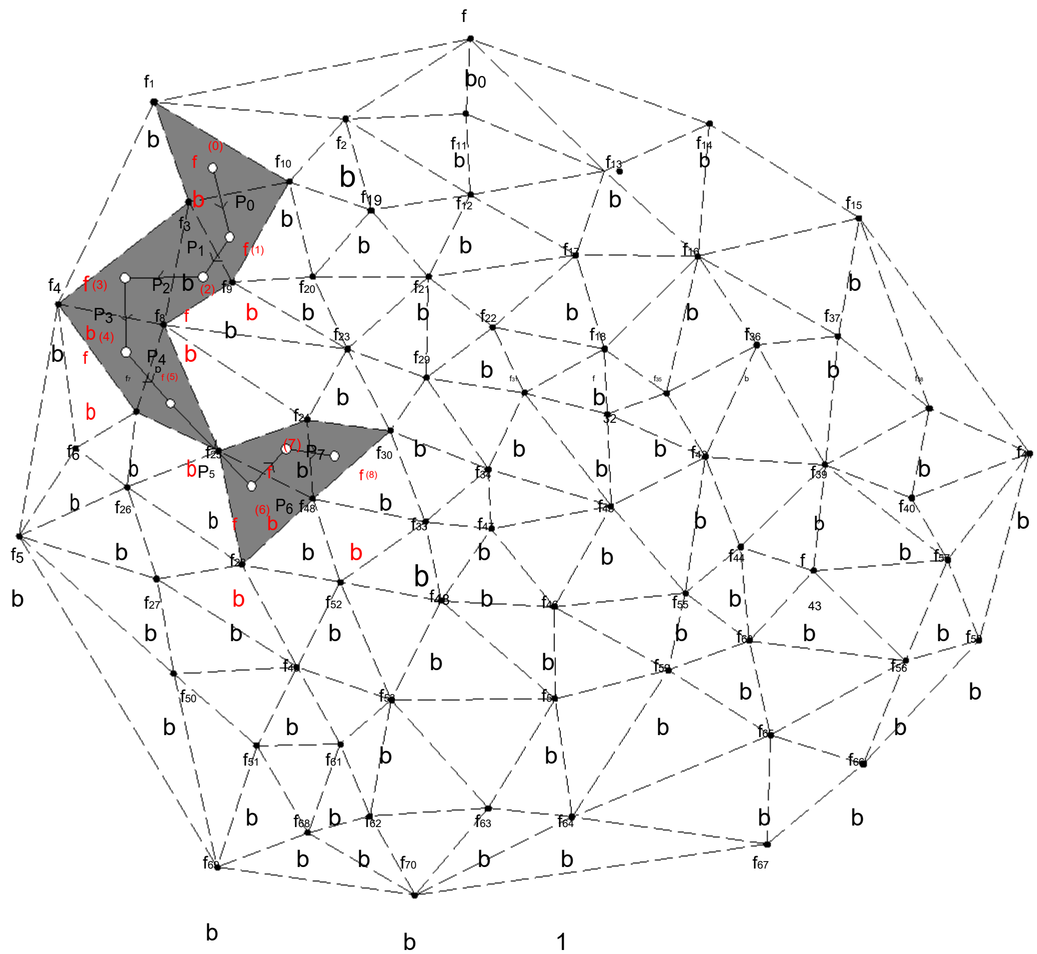


Figura 3.18: Caminata aleatoria entre S_{Df} vecinos. En rojo se muestran las formas $f^{b(i)}$ para los diferentes subdominios vecinos.

El metodo es un heur stico que parte de una forma verme inicial $f^{b(0)}$ para algun $S^{(0)}_{Df}$ elegido aleatoriamente. Los vertices del nuevo subdominio $S_{D_f}^{(0)}$ corresponden al conjunto $V^{(0)}$ de los q vecinos mas cercanos a una forma f^{b}_{i} elegida aleatoriamente o previamente de nida.

La siguiente forma $f^{(1)}$ de esta simulacion se calcula en un nuevo subdominio $S^{(1)}_{Df}$ que se determina por el conjunto de los q centroides mas proximos $V^{(1)}$ de algun nodo f^{b}_{j} seleccionado aleatoriamente en $S^{(0)}_{Df}$. Esta forma $f^{(1)}$ se calcula en $S^{(1)}_{Df}$ a traves del metodo descrito en la seccion 3.3.2. El nuevo b n-s mplex vecino del anterior.

El proceso se repite iterativamente hasta satisfacer un criterio de paro, por ejemplo un numero de iteraciones previamente de nidas. Cabe se~nalar que los centroides se calculan e cientemente con el metodo aproximado expuesto en (Muja y Lowe, 2009).

El procedimiento propuesto permite que las deformaciones se simulen a partir de una instancia inicial y continuen de manera natural conservando informacion de forma entre las iteraciones y logrando el efecto de deformaciones vermes continuas validas.

En la gura 3.19, se muestra un ejemplo del movimiento generado por el calculo iterativo de formas aleatorias entre cambio de S_{D_f} vecinos. Se parte de una forma $\hat{f}^{(0)}$ y se muestra para todas las iteraciones la deformacion verme valida. En esta se usa 10 nodos y 200 centroides y se aprecia que en la sucesion hay un cambio de la forma gradual no abrupto.

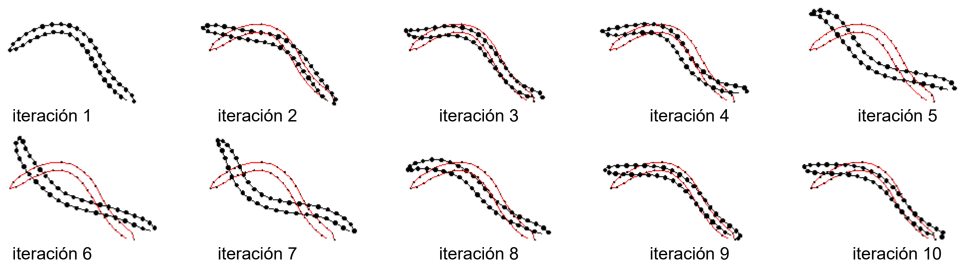


Figura 3.19: Sucesion de movimientos vermes continuos entre S_{D_f} vecinos. El nematodo que se distorsiona esta representado con hitos dibujados con c rculos de relleno negro. En rojo se colorea el borde de la silueta original. Para esta simulacion se utiliza 10 nodos y 200 centroides.

El algoritmo de la simulacion de movimientos aleatorios vermes se presenta en la gura 3.20.

Entrada: $\hat{f}^{(0)}$ ← Forma inicial
D_f ← Dominio de formas permitidas
p ← número de formas vermes deseadas

$C \leftarrow kmeans(D_f)$ /* centroides de D_f calculados por *k*-means */
$M_0 \leftarrow \hat{f}^{(0)}$ /* M_0 vértice inicial */
$i \leftarrow 1$
mientras $i < p$ **hacer**
  $V^{(i)} \leftarrow \mathrm{knn}(C, M_{i-1})$ /* $V^{(i)}$ vecinos más cercanos de M_{i-1} por *k*-nn */
  $S^{(i)}_{D_f} \leftarrow ksimplex(V^{(i)})$ /* i-ésimo subdominio de formas permitidas */
  $\hat{f}^{(i)} \leftarrow random\ S^{(i)}_{D_f}$ /* Cálculo de la i-ésima deformación $\hat{f}^{(i)} \in S^{(i)}_{D_f}$ */
  $M_i \leftarrow random(V^{(i)})$ /* M_i es un vértice al azar de $S^{(i)}_{D_f}$ */
  $i \leftarrow i + 1$
fin

Figura 3.20: Algoritmo de distorsion de un nematodo entre S_{D_f} vecinos.

El algoritmo permite buscar nuevas siluetas de nematodos, s miles en forma, que permitan cambiar las part culas del frente de Pareto por otras que se ajusten mejor. En adelante, la funcion serpenteo se denota snk.

3.4. Metodo de deformacion de nematodos: forrajeo

El metodo de deformacion de nematodos denominado forrajeo es una tecnica de simulacion de la deformacion de organismos vermiformes sobre la super cie terrestre en su desplazamiento. Este consiste en continuas ondulaciones sinusoidales a traves de contrac-ciones del cuerpo, cola y cabeza, y alargamiento.

Con este se logra una sucesion de formas continuas que simulan la locomocion de los nematodos, permitiendo, en primera instancia, generar nuevas formas vermes a partir de cada instancia del conjunto de entrenamiento C^b y, la evolucion de la silueta aplicada en el modelo evolutivo de forma.
En la gura 3.21 se muestra el proceso general que se detalla en las secciones 3.4.1, 3.4.2 y 3.4.3.

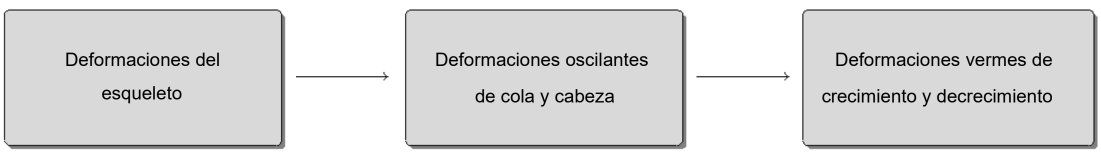


Figura 3.21: *Diagrama de ujo del metodo forrajeo.*

Cabe destacar cada elemento del conjunto de entrenamiento C^b es previamente preprocesado segun la seccion 3.2.1.

3.4.1. Deformaciones del esqueleto

Las nuevas formas de nematodo f^b_i son logradas a traves del movimiento del esqueleto E_i (gura 3.22), de nido como el vector compuesto por las posiciones de las vertebras, cada una dada por el promedio de los dos hitos correspondientes diametralmente opuestos h_{ij} y $h_{i\,m\,j}$.

El conjunto de todos los esqueletos E_i de f^b_i 2 D_f se denota E, y cada uno se expresa vectorialmente por:

$$E_i := \left(E_{i0x} \quad E_{i0y} \quad E_{i1x} \quad E_{i1y} \quad E_{iux} \quad E_{iuy} \right)^T \tag{3.16}$$

donde la posicion de cada vertebra se de ne por:

$$E_{ijx} := \frac{x_{ij} + x_{i(m\,j)}}{2} \qquad E_{ijy} := \frac{y_{ij} + y_{i(m\,j)}}{2} \qquad \text{para } 0 < j < u: \tag{3.17}$$

y u := m=2 con m es el numero de hitos de f^b_i.

Ademas, los hitos sobre la cabeza y cola de la forma verme son elegidos respectivamente como la primera vertebra:

$$V_{i0} := (E_{i0x}\ ; E_{i0y}\ ) := (x_{i\ 0};y_{i\ 0}) \tag{3.18}$$

y ultima vertebra

$$V_{iu} := (E_{iux}\ ; E_{iuy}\ ) = (x_{i\ u};y_{i\ u}) \tag{3.19}$$

La distorsion de la forma de la silueta verme se concentra en dos fragmentos del esqueleto, de nidos por las u_0 vertebras mas cercanas a la cabeza y a la cola de fb_i. Para las vertebras $V_{ij} = (E_{ijx}\ ; E_{ijy}\ )$ del fragmento de la cabeza se cumple 1 j < u_0, y para el fragmento de cola u u_0 j < u.

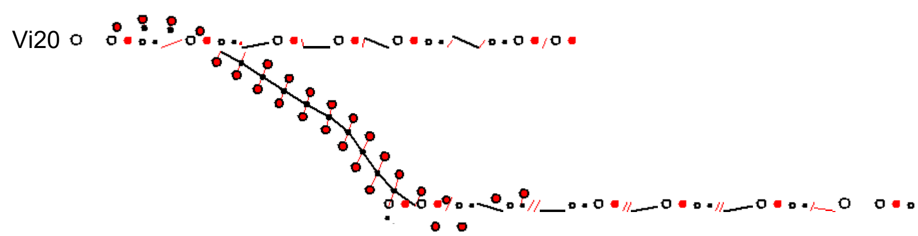


Figura 3.22: Esqueleto de una forma verme f^{b}_{i}. En c rculos con relleno de color negro se visualizan las vertebras. Los c rculos con relleno color rojo corresponden a los hitos de frontera de la forma verme. La forma verme posee 40 hitos de frontera con u = 20.

El proceso de forrajeo esta dado por dos movimientos: oscilaciones laterales y, expansion y contraccion de cola y cabeza. Ambos se explican en la secciones 3.4.2 y 3.4.3 respecti-vamente.

3.4.2. Deformaciones vermes oscilantes de cola y cabeza

La primera parte del metodo del forrajeo es una deformacion del cuerpo del nematodo inspirada en el trabajo de K.-M. Huang et al. (2008), que consiste en oscilaciones laterales de cabeza y cola.

Para realizar estos movimientos, se toman los fragmentos con un numero u^0 vertebras de cola y cabeza, y se manipulan (junto a los hitos correspondientes a cada vertebra) como una cadena cinematica directa (Harary y Yan, 1990). En dicha cadena, cada vertebra cuenta con un grado de libertad rotacional con un angulo de rotacion .

Dado el numero u^0 de vertebras a modi car y un angulo inicial , el algoritmo toma la vertebra del fragmento de cabeza $V_{iv^0} := (t_x; t_y)$ como ja y aplica la transformacion de nida por:

$$M_v \begin{bmatrix} x_{ij} \\ y_{ij} \\ 1 \end{bmatrix} := \begin{bmatrix} \cos & \mathrm{sen} & (t_x \cos + t_y \,\mathrm{sen}) + t_x \\ \mathrm{sen} & \cos & (t_y \cos \;\; t_x \,\mathrm{sen}) + t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ij} \\ y_{ij} \\ 1 \end{bmatrix} \qquad (3.20)$$

al resto de vertebras e hitos en la cadena cinematica. Luego, dada la nueva posicion de la vertebra $V_{i\, u^0\, 1}$ se le aplica M_v al resto de elementos de la subcadena cinematica hacia la cabeza. Posteriormente se repite el proceso a partir de la vertebra $V_{i\, u^0\, 2}$ y as sucesivamente hasta llegar al hito de la cabeza.

Lo anterior provoca una rotacion en secuencia de las vertebras y sus respectivos hitos, que en conjunto, logra una rotacion natural del fragmento cabeza. De forma analoga, se sigue el procedimiento desde la vertebra colocada en la posicion u u^0 + 1 hacia la cola. En ambos fragmentos, el metodo se aplica individualmente.

Un ejemplo de esta deformacion se ilustra en la gura 3.23.

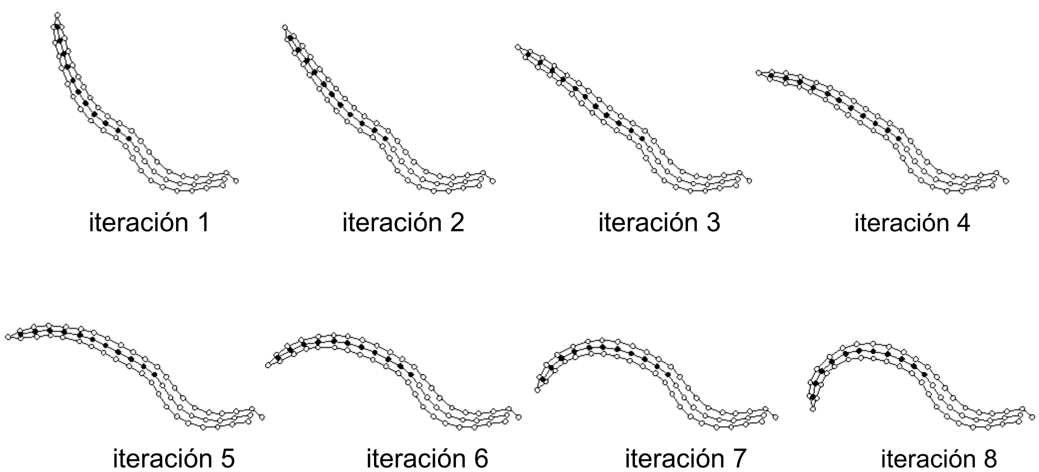


Figura 3.23: Ejemplo del metodo del forrajeo (parte I) para la cola de f^b_i para $v^0 = 11$ y = 0:2 rad. En la gura se muestran algunas iteraciones de las deformaciones.

En la gura 3.24 se muestra el algoritmo del metodo del forrajeo que distorsiona de la vertebra u al hito cola.

Entrada: f_i /* Forma de nematodo de *m* hitos */

u Número de hitos a deformar */

b

ϑ /* /* Ángulo de deformación */

t_x, t_y /* Traslaciones */

para $j \leftarrow 1$ **hasta** u **hacer**

  $V_{ij} \leftarrow$ n E_{ijx} E_{ijy} i /* cálculo de la *j*-ésima vértebra de *i* */ E

fin

para $j \leftarrow u - u^0 + 1$ **hasta** u **hacer**

  $k \leftarrow j$

  mientras $k \leq u$ **hacer**

    $V_{ik} \leftarrow M_V [V_{ik}]$ /* Transformación de la vértebra *k* */

    $h_{yik}^{x} {}^{i} \leftarrow M_V\, h_{yik}^{x}$ i /* Transformación del hito *k* */

    $x_{i(m-k)}$ M $x_{i(m-k)}$ *m* *k*

    $h_{yi(m-k)}$ i $\leftarrow$ $_V\, h^{y}_{i(m-k)}$ i /* Transformación del hito − */

    $k \leftarrow k + 1$

  fin

fin

Figura 3.24: Algoritmo del metodo de forrajeo (parte I) usando la transformacion M_V para el fragmento cola para la vertebra u. La matriz M_V corresponde a (3.20).

3.4.3. Deformaciones vermes de crecimiento y decrecimiento

La segunda parte del movimiento del forrajeo corresponde a una contraccion o extraccion del fragmento cola y/o fragmento cabeza manteniendo la escala y la forma de la silueta original.

El alargamiento del fragmento cabeza se de ne como el desplazamiento d de los hitos

h_{ij} y $h_{i\ m\ j}$ que determinan la vertebra $V_{ij} := (E_{ijx} ; E_{ijy})$ dado por:

$$h_{ij} := h_{ij} + d \qquad h_{i\ m\ j} := h_{i\ m\ j} + d \tag{3.21}$$

donde $d := (d_x ; d_y)$ se calcula por:

$$d_x := (E_{ijx} \quad E_{ij+1x})(s \quad 1) \qquad d_y = (E_{ijy} \quad E_{ij+1y})(s \quad 1) \tag{3.22}$$

con s 2 IR y 1 $j < u^0$.

!

Si $s > 1$, el nematodo crece en el mismo sentido del vector $V_{ij+1}\ V_{ij}$, en caso contrario, se contrae a nivel de cabeza. En caso que $0 < s < 1$, el fragmento cabeza decrece. Con el mismo razonamiento, es calculado el crecimiento y decrecimiento del fragmento de la cola.

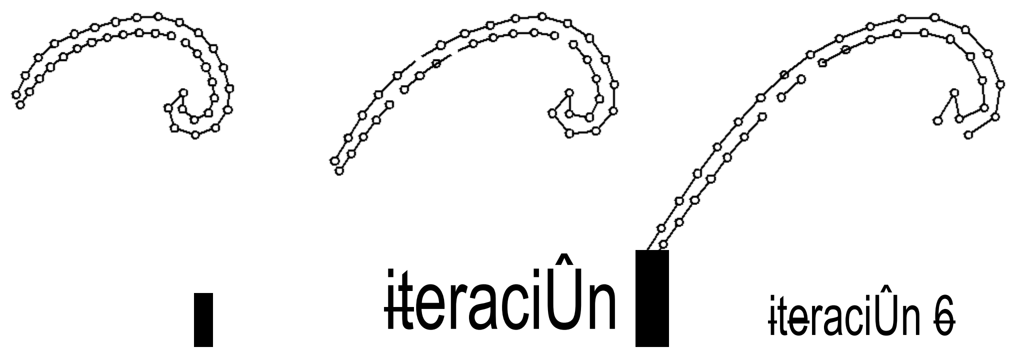


Figura 3.25: Ejemplo del metodo forrajeo (parte II) para el crecimiento de la silueta del fragmento de la cabeza del nematodo usando $u^0 = 5$ y s = 1:3 en cada iteracion.

Luego de cada alargamiento o encogimiento, se equidistan el numero de hitos que moldean la silueta el organismo vermiforme usando el metodo descrito en la segunda fase de la seccion 3.2.1. El algoritmo se presenta en la gura 3.26.

```
Entrada: fbi        /* Forma de nematodo de m hitos */
         s          /* Escala de contracción-dilatación */
         u0         /* Número de vértebras a dilatar-contraer */
para j ← 1 hasta u0 + 1 hacer
    Eij ← ( (xij + xi(m−j)) / 2 , (yij + yi(m−j)) / 2 )   /* Cálculo de la j-ésima vértebra de Ei */
fin
i ← u0
mientras i > 0 hacer
    d ← (Eijx − Eij+1x) (s − 1) (Eijy − Eij+1y) (s − 1)   /* Cálculo de los desplazamientos d */
    hij ← hij + d                                         /* Actualización de las posiciones de los hitos */
    him−j ← him−j + d
    i ← i − 1
fin
```

Figura 3.26: Segunda parte del algoritmo del metodo de forrajeo de crecimiento del fragmento de la cabeza.

Cabe destacar que este metodo no corresponde al escalamiento que se detalla en la seccion 2.5. Se diferencia en que este sufre de un alargamiento o encogimiento fragmento cola yno fragmento cabeza manteniendo el ancho de la silueta de gusano, el otro es un escalamiento de tama~no de la silueta. En adelante, la funcion de forrajeo se denota forrg.

3.5. Truncamiento a un S_{Df}

Una representacion vectorial $\hat{f}_i$ es vermiformemente valida si existe algun subdominio al que pertenezca. En caso contrario, se dice que no es vermiformemente valida y se denota $\tilde{f}_i$ El metodo de truncamiento de part culas a un subdominio S_{Df} consiste en una aproximacion numerica para $\tilde{f}_i$, denotado trun($\tilde{f}_i$), que calcula una representacion vectorial verme $\hat{f}_i$ proyectada en un subdominio engendrado por los vertices mas cercanos a su forma. En la gura 3.27 se muestra $\hat{f}_i$ truncado a un S_{Df} .

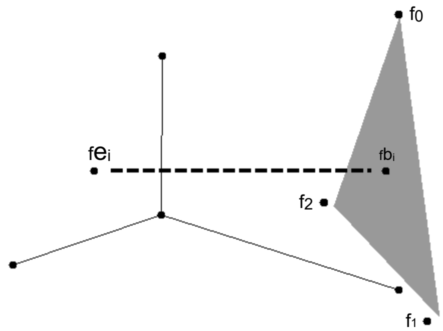


Figura 3.27: Truncamiento $\hat{f}_i$ a un S_{Df} .

El proceso consiste en hallar un S_{Df} generado por los vertices $\hat{f}_0; \hat{f}_1; \ldots; \hat{f}_q$ mas cercanas en forma y los respectivos los coe cientes $\alpha_0; \ldots; \alpha_q$ que cumplan las restricciones de n-s mplexes y que la distancia entre los hitos correspondientes de $\hat{f}_i$ y $\tilde{f}_i$ sea m nima. Esto es:

$$\min_{\{\hat{f}_0;\ldots;\hat{f}_q;\,\alpha_0;\ldots;\alpha_q\}} \|\hat{f}_i - \tilde{f}_i\|_2^2 \tag{3.23}$$

As , $\hat{f}_i$ es el elemento truncado que pertenece a S_{Df} y es mas proximo en distancia a $\tilde{f}_i$ (gura 3.28). Esto es:

$$\mathrm{trun}(\tilde{f}_i) = \hat{f}_i \tag{3.24}$$

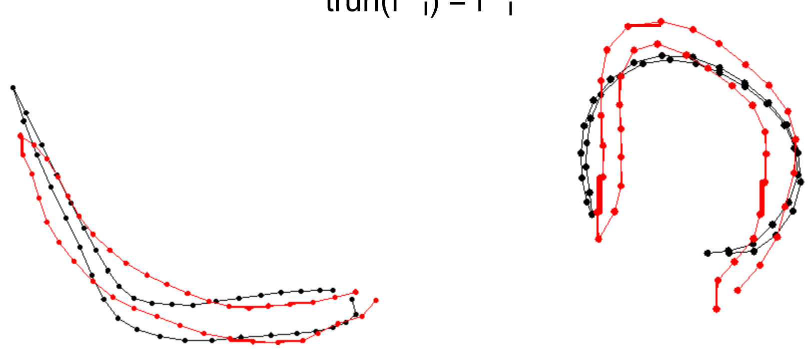


Figura 3.28: Ejemplos del metodo de truncamiento de formas no vermes. En negro se colorea una forma no verme y en rojo su respectivo truncamiento en un S_{Df} de 10 nodos.

3.6. Funciones objetivo

El modelo evolutivo de forma es un proceso basado en optimizacion multiobjetivo de dos funciones que permiten el ajuste de una forma verme a la imagen digital. La primera es una funcion que mide la longitud del esqueleto de f^b_i. La segunda funcion obtiene una metrica de posicionamiento de f^b_i sobre el borde del nematodo en la imagen digital.

Ambas funciones se normalizan en una escala comun a traves de la funcion de sigmoide que se denota F_{Sig} y se de ne por:

$$F_{Sig}(x) := \frac{L}{1 + e^{k(x \quad x_0)}} \tag{3.25}$$

donde L es el maximo valor de la funcion, k es la tasa de crecimiento de la curva y x_0 es el punto medio de la curva sigmoide. En las secciones 3.6.1 y 3.6.2 se detallan ambas funciones.

3.6.1. Funcion verme de distancia

Para regular el crecimiento de la forma verme f^b_i al aplicar el modelo evolutivo de forma, se recurre a una funcion que permita evaluar su tama~no. El objetivo es proveer al optimizador informacion de la extension o longitud de la silueta para que coadyuve en su ajuste.

El tama~no del nematodo f^b_i se de ne como longitud de su esqueleto dado en (3.16). Esta es dada por una funcion que calcula la suma de las distancias entre vertebras consecutivas. Esto se denomina funcion verme de distancia, se denota F_D y se de ne por F_D : E ! IR con:

$$F_D(E_i) := D_{real} \tag{3.26}$$

donde D_{real} es la medida ideal o exacta del esqueleto. La aproximacion de esta metrica es dada por:

$$D_{real} \quad D_{aprox} \tag{3.27}$$

con

$$D_{aprox} = \sum_{j=0}^{\frac{m}{2} \quad 1} \sqrt{(E_{ijx} \quad E_{ij+1x})^2 + (E_{ijy} \quad E_{ij+1y})^2} \tag{3.28}$$

con m es el numero de hitos y $\frac{m}{2}$ el numero de vertebras. Un maximo local de F_D es dado si la silueta de f^b_i se ajusta completamente al nematodo en la imagen digital, en consecuencia, es obtenido su longitud real de su esqueleto.

El maximo local es determinado por:

$$j D_{aprox} \quad D_{real} j < \tag{3.29}$$

para un previamente de nido. En la gura 3.29 se muestra el esqueleto de una forma verme ajustado al nematodo en una imagen digital.

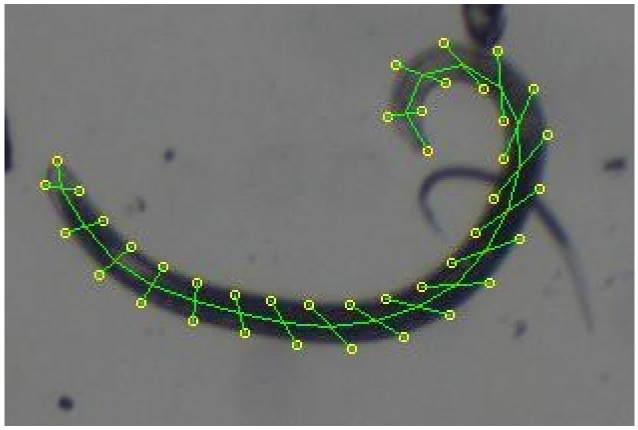


Figura 3.29: Ejemplo de esqueletizacion de un nematodo a partir de los hitos de frontera.

Para usarla en el modelo evolutivo de forma, la funcion F_D es normalizada al intervalo de [0 ; L] usando la funcion sigmoide dada en (3.25) dando como resultado:

$$\hat{F}_D (E_i) := \frac{L}{1 + e^{k(F_D(E_i) \; x_0)}} \tag{3.30}$$

para algun k 2 IR determinado previamente. En los experimentos siguientes se toma emp ricamente $k = 0{:}2\ 10^{4}$ con el n coadyuvar al M P SO en la determinacion de las part culas del frente de Pareto.

3.6.2. Funcion verme de contorno

Para determinar el ajuste de la silueta $\hat{f}_i$ sobre en la imagen digital, y en particular, el amojonamiento de los hitos a lo largo del contorno del nematodo, se propone la funcion denominada funcion verme de borde denotada F_B. Esta funcion provee una metrica de ubicacion de la silueta verme $\hat{f}_i$ a partir del conjunto de hitos que la conforman.

La funcion F_B se calcula a partir de los vectores tangenciales de borde, ortogonales de borde y los gradientes. En las siguientes secciones se describen estos vectores y la funcion F_B.

Vectores tangenciales de borde

Los vectores tangenciales de borde, denotados V_B, son vectores tangentes a la curva C_{spl} determinada por los puntos de control h_j de f_i.

El conjunto de los vectores tangenciales de borde de f_i se denota B y vectorialmente son de nidos por:

$$V_B := h \quad b \quad i \quad c$$

Cada componente V_{Bj} de V_B corresponde al j esimo vector tangente en el hito h_j, de nido por:

$$V_{Bj} := \begin{bmatrix} V_{Bxj} \\ V_{Byj} \end{bmatrix} = \begin{bmatrix} \frac{1}{m_{xy}} V_{Bxj} \\ \frac{1}{m_{xy}} V_{Byj} \end{bmatrix} \tag{3.32}$$

donde x_j y y_j son las coordenadas h_j, $m_{xy} = \sqrt{V_{Bjx}^2 + V_{Bjy}^2}$ es la magnitud del vector y, V_{Bxj} y V_{Byj} son las componentes normalizadas de V_{Bj}.

Para lograr que los vectores tangenciales de borde V_{Bj} suavicen el sentido hacia el contorno de la curva C_{spl} de f, se deriva parcialmente el resultado de la convolucion entre cada hito de H_j y la funcion Gaussiana G para la aproximacion del gradiente en la imagen digital. En este se indica la direccion del gradiente en la que ocurre el mayor cambio que ocurre en el borde del nematodo. Es dada por:

$$G(x;\ ;\ ) := \frac{1}{\sqrt{2\ ^2}} e^{-\frac{(x\ )^2}{2\ ^2}} : \tag{3.33}$$

El resultado de este calculo es:

$$V_{Bjx} := \frac{\partial}{\partial x}\left(H_j \circledast G\right) = H_j \circledast \frac{\partial G}{\partial x} \tag{3.34}$$

$$V_{Bjy} := \frac{\partial}{\partial y}\left(H_j \circledast G\right) = H_j \circledast \frac{\partial G}{\partial y} \tag{3.35}$$

donde $\circledast$ denota la operacion de convolucion. Las componentes $\frac{\partial G}{\partial x}$ y $\frac{\partial G}{\partial y}$ se denominan las derivaciones orientadas de la funcion Gaussiana (OGD por sus siglas en ingles) en x y en y (Alvarado et al., 2001). La gura 3.30 muestra el calculo de los vectores tangenciales de borde sobre el contorno nematodo en una imagen digital.

Vectores ortogonales de borde

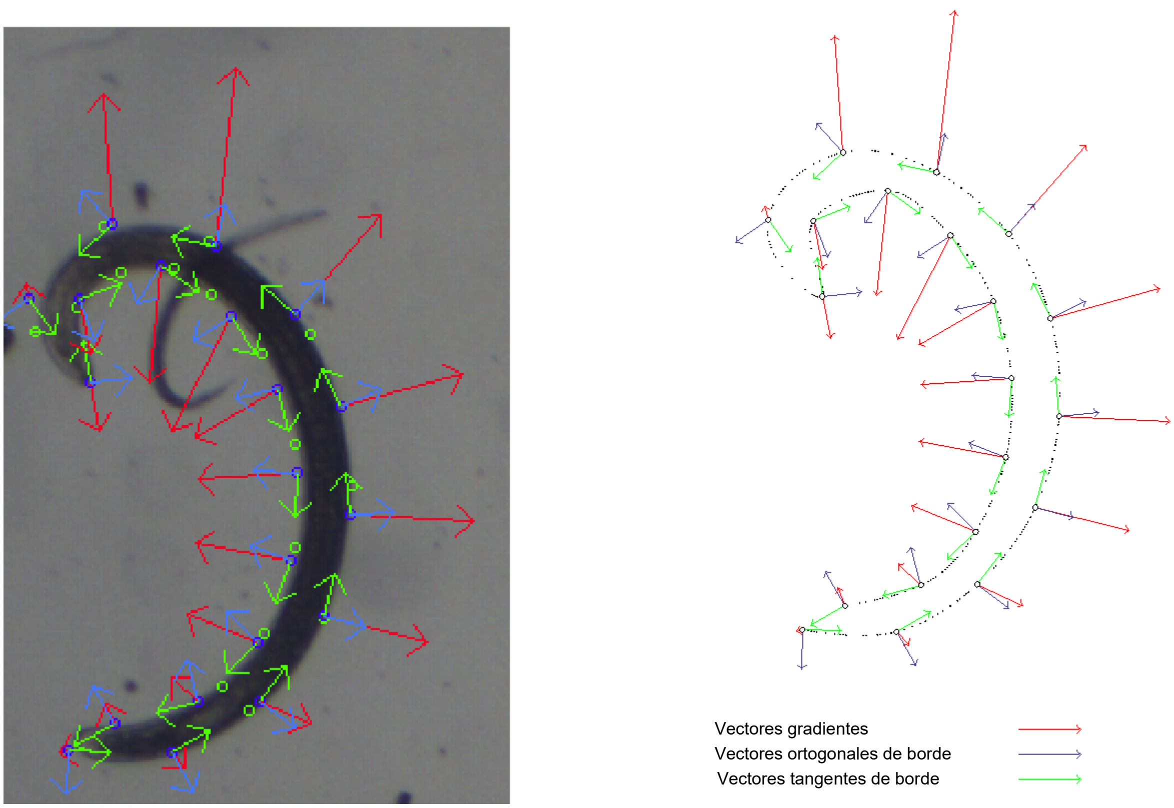
Los vectores ortogonales de borde de cada f_i, denotados V_O, son el conjunto de vectores perpendiculares a cada V_{Bj} en cada hito de ${}^bH_j$. Vectorialmente, $\vec{V}_O$ se expresa por:

$$\vec{V}_O := \left[\vec{V}_{O0} \quad \vec{V}_{O1} \quad \ldots \quad \vec{V}_{Om-1} \right] \tag{3.36}$$

Cada vector $\vec{V}_{Oj}$ de V_O se calcula por:

$$V_{Oj} := \vec{V}_{Byj} \quad V_{Bxj} \tag{3.37}$$

En la figura 3.30 se muestra los vectores ortogonales de borde.



(a) Imagen de un nematodo con los vectores tangenciales y ortogonales de borde coloreados en verde y rojo respectiva-mente.

(b) Extraccion de los vectores.

Figura 3.30: En verde se representan los vectores tangenciales, en rojo se representa los vectores gradiente de borde, y en azul se representan ortogonales de borde de un nematodo y su respectiva silueta conformada por 20 hitos de frontera H_j.

Funcion verme de borde F_B

La funcion verme de borde F_B provee una metrica de posicionamiento de las formas $\hat{f}_i$ sobre los nematodos en las imagenes digitales.

Si se toman los hitos como puntos de anclaje de los vectores ortogonales de borde y gradiente, entonces se garantiza que entre mayor sea la suma de sus productos escalares de los vectores V_{O_j} (3.37) y los $\nabla I(x_j;y_j)$ (2.2) respectivamente, entonces todos los hitos en conjunto estan ubicados con mayor proximidad a los bordes del nematodos en las imagenes digitales.

La funcion $F_B : D_f \to \mathbb{R}$ viene dada por:

$$F_B\left(\hat{f}_i\right) := \sum_{j=0}^{m-1} V_{O_j} \cdot \nabla T(I(x_j;y_j)) = \sum_{j=0}^{m-1} \begin{bmatrix} V_{Bx_j} & V_{By_j} \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} T\left(I_{x_j;y_j}\right) \\ \frac{\partial}{\partial y} T\left(I_{x_j;y_j}\right) \end{bmatrix} \tag{3.38}$$

donde T es la funcion que ltra la imagen a niveles de gris. Para determinar el maximo local de F_B, el conjunto de vectores tangentes sobre h_j de $\hat{f}_i$ deben seguir la trayectoria del contorno de la forma verme, y el conjunto de vectores ortogonales a cada vector tangente en h_j de $\hat{f}_i$ debe tener el mismo sentido que el vector gradiente.
Si el vector ortogonal de borde y el vector gradiente de la imagen tienen el mismo sentido, se asegura que:

- El producto escalar de los vectores ortogonales de borde y gradiente es maximo.
- En los h_j y h_{m-j} de $\hat{f}_i$, sus vectores ortogonales de borde respectivos tienen sentido aproximadamente opuesto. Lo que garantiza que estos hitos se ubiquen en puntos sobre el borde que sean opuestos en su silueta y, en conjunto, cubran la totalidad del nematodo.

Para usarla en el modelo evolutivo de forma, la funcion F_B se normaliza al intervalo de $[0 ; L]$ empleando la funcion sigmoide dada en (3.25) dando como resultado:

$$\widehat{F_B}\left(\hat{f}_i\right) := \frac{L}{1 + e^{k\left(F_B\left(\hat{f}_i\right)\right)}} \tag{3.39}$$

para algun $k \in \mathbb{R}$ previamente dado.

3.7. Modelo evolutivo de forma

En esta seccion se desarrolla el modelo evolutivo de forma (MEF) parametrico basado en hitos de frontera. Este metodo es capaz de adoptar, a partir de un conjunto de part culas $\hat{f}_i$, la silueta del nematodo en la imagen digital atendiendo simultaneamente un doble criterio de decision: forma y tama~no. El modelo propuesto es iterativo y se inicializa con:

- Un conjunto de datos entrenamiento $\hat{C}$ normalizado (segun la seccion 3.2.1).
- Una imagen digital que contenga un nematodo.
- Una forma $\hat{f}_0$ seleccionada aleatoriamente del conjunto de entrenamiento $\hat{C}$. Esta puede ser generada de manera sintetica usando los resultados de la seccion 3.3.2.
- Una base datos vermes BD_0 generada a partir de $\hat{f}_0 \in \hat{C}$ expresada por (3.12).

Para la aplicacion del modelo, se elimina la forma de inicio $\hat{f}_0$. Rede niendo $\hat{C}$ de la forma:

$$\hat{C} := \hat{C} \cap \{\hat{f}_0\}^c$$

En esta fase, la base de datos BD_0 se calcula por el metodo de caminatas aleatorias para un numero v de individuos. Cada subdominio de forma S_{Df} esta constituido por un numero w de vertices cercanos en forma. El modelo sigue las fases que se describen en la gura 3.31 para un numero N_{iter} de iteraciones prede nidas. En la seccion 3.7.1 se describen cada una de estas etapas.

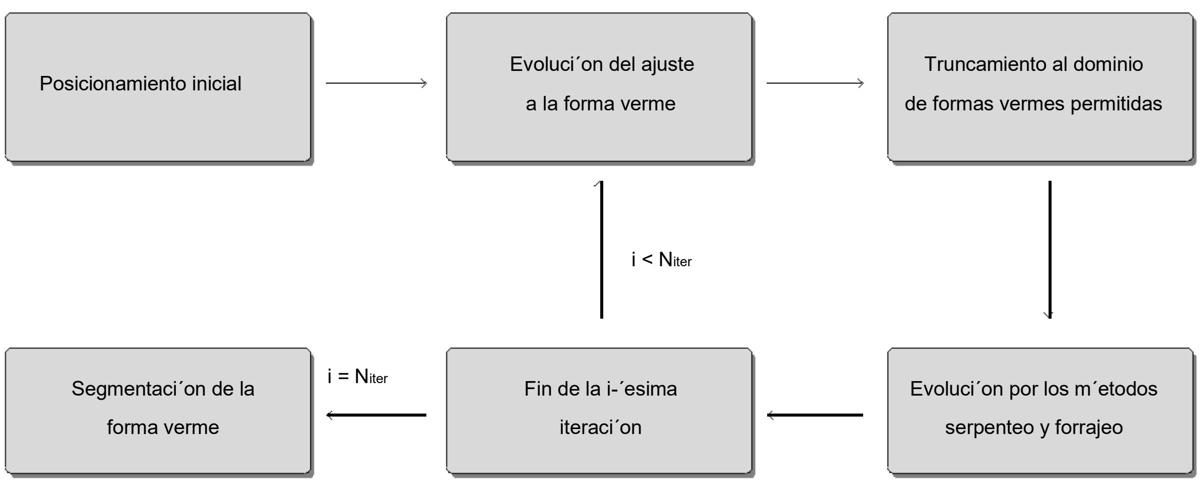


Figura 3.31: Diagrama de ujo del modelo evolutivo de forma para N_{iter} iteraciones.

La forma verme de inicio $\hat{f}_0$ se alinea respecto al centro de masa de la imagen digital. Ademas, es escalada para que la longitud del esqueleto E no exceda la mitad del m nimo entre las dimensiones de la imagen digital.

Cabe destacar que el MEF es invariante durante todo el proceso de su ejecucion. Lo anterior se debe a que el algoritmo trabaja en un sistema de coordenadas determinado previamente por los centroides calculados de $\hat{C}$. Estos, que conforman los subdominios de forma, no var an en traslacion, tama~no y rotacion durante la ejecucion del algoritmo.
Asimismo, las part culas del frente de Pareto se alinean a dicho sistema coordenado cuando pasan por los procesos del serpenteo y truncamiento. Despues de nalizado, se recalculan las nuevas formas correspondientes a la escala previa. Las alineaciones son realizadas utilizando el metodo de Procrusto explicado en la seccion 2.5.

Por otro lado, a nivel de imagen digital, los elementos del frente de Pareto se alinean a $\hat{f}_0$ para ser aplicados en el MOPSO y el metodo del forrajeo. Posteriormente, evolucionan en forma y tama~no segun el ajuste correspondiente.

3.7.1. Ajuste de forma usando M OP SO

En la primera etapa del modelo, se inicia el ajuste por medio del metodo de optimizacion multiobjetivo M OP SO (descrito en la seccion 2.8) en un espacio de busqueda dado por el dominio de formas permitidas D_f .
En cada iteracion, los hitos de las formas $\hat{f}_i$ son ajustados individualmente dando al modelo un criterio de decision para aproximarlas a las siluetas en las imagenes digitales. Este criterio se denominar en adelante la medida optima de borde-longitud y se re ere a que todos los hitos h_j de la forma vectorial verme $\hat{f}_i$ esten ubicados sobre el borde del nematodo cubriendo su silueta. El criterio satisface si se cumplen las siguientes dos premisas:

- Todos los hitos son equidistantes y correspondan a los puntos de anclaje de los vectores ortogonales de borde V_{Bj} de un nematodo en la imagen digital.
- La longitud del esqueleto del nematodo en la imagen digital debe corresponder a la longitud del esqueleto de la forma $\hat{f}_i$.

Numericamente, las premisas borde-longitud se satisfacen con las funciones $\hat{F}_D$ y $\hat{F}_B$ dadas por las ecuaciones (3.30) y (3.39) respectivamente. Ademas, a traves de dicho criterio y M OP SO se logra determinar el j-esimo conjunto optimo de Pareto que esta dado por:

$$M_j := \left[p_{0j} \;\; p_{1j} \;\; \ldots \;\; p_{q-1\,j} \right]^T \tag{3.40}$$

que corresponde a las q part culas $\hat{f}_i$ que compiten por el mejor ajuste en longitud y en borde respecto al nematodo en la imagen digital.

Cabe se~nalar que la forma media inicial $\hat{f}_0$ se conserva durante todo el proceso para usarse en la normalizacion de los elementos del frente de Pareto usados en los metodos de truncamiento y serpenteo. En el M OP SO y el metodo del forrajeo, las formas vermes evolucionan en tama~no y forma.

La metodolog a de adaptacion de la silueta $\hat{f}_i$ a la del nematodo en la imagen digital tiene dos estadios: el de posicionamiento inicial y la evolucion de ajuste. Ambas se describen a continuacion.

Etapa de posicionamiento inicial

La etapa de posicionamiento inicial requiere que las primeras iteraciones del modelo evolu-tivo de forma se enfoquen en hallar la mejor ubicacion de la forma $\hat{f}_0$ respecto al nematodo en la imagen digital usando M OP SO. Las formas vermes del frente de Pareto se escalan al tama~no de la forma verme inicial, y evolucionan de acuerdo al tama~no del nematodo presente en la imagen digital.

La mejor ubicacion inicial se re ere a que el area de la silueta descrita por $\hat{f}_i$ se traslape con la mayor parte del nematodo en la imagen digital. Ademas, que la mayor cantidad de hitos del fragmento central de $\hat{f}_i$ se ubiquen en el contorno del nematodo.
Para esto, se utiliza la estrategia de darle un peso w_j distinto a la evaluacion en (3.39) de los hitos localizados en la parte central de la silueta en relacion con los fragmentos cola y cabeza en las primeras M iteraciones del MEF.

Esto permite al M OP SO decidirse por incluir en el conjunto optimo de Pareto las siluetas que se han adaptado en su parte central primeramente, sin dejar de lado la evolucion de la forma a la existente en las trozos cabeza y cola en la imagen digital, para posteriormente, darle paso al metodo del serpenteo.

Para una forma verme $\hat{f}_i$ normalizada a m hitos, la asignacion los pesos w_j se hace a traves de un mapeo lineal del j esimo hito h_j al intervalo [a;a], evaluados en la funcion gaussiana G(x; ;) (gura 3.32), dada por:

$$w_j := \begin{cases} G\left(\frac{4aj}{m-1} - a; 1; 0\right) & \text{si} \quad 0 \le j < \frac{m-1}{2} \\ G\left(\frac{2aj - 2am}{m-1} - a; 1; 0\right) & \text{si} \quad \frac{m-1}{2} \le j \le m-1 \end{cases} \tag{3.41}$$

As de (3.38), las evaluaciones para hitos para las primeras i iteraciones se ponderan con w_j de la forma:

$$F_B\left(\hat{f}_i\right) := \sum_{j=0}^{m-1} w_j V_{Oj} \nabla I(x_j;y_j) \tag{3.42}$$

Etapa de evolucion de ajuste a la forma verme

El proceso de reemplazo interno o actualizacion de part culas contenidas en el conjunto optimo de Pareto M_j permite que se deformen hacia la del nematodo presente en la

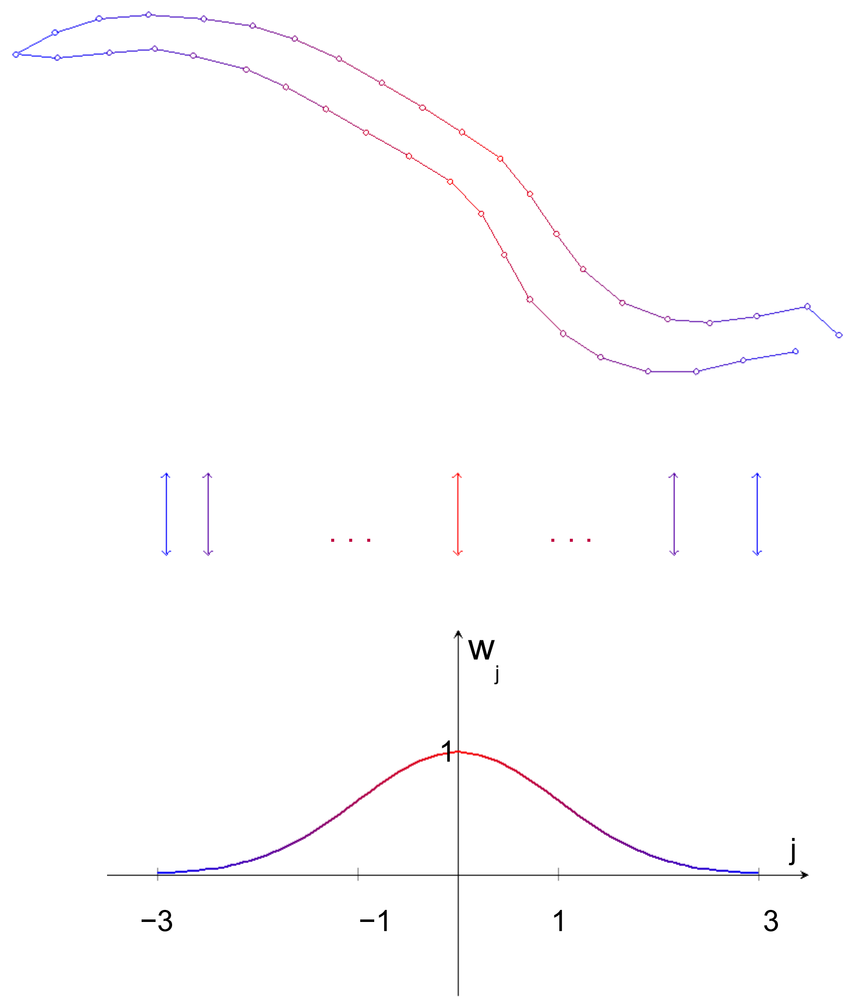


Figura 3.32: Mapeo de los pesos w_j para cada hito de $\hat{f}^{1}_{i}$ usando la funcion gaussiana en el intervalo de [3;3]. La gra ca presenta un degradado de colores, en la cual la parte coloreada en rojo de la silueta se le asigna un maximo w_j, y que se degrada hasta los hitos en azules a los cuales se les asigna un peso menor segun corresponda.

imagen digital. El numero de part culas q que contiene el frente de Pareto puede variar de acuerdo al resultado o salida del algoritmo.

Asimismo, las nuevas soluciones del conjunto optimo de Pareto contienen formas no do-minadas por todos los miembros del enjambre y las part culas dominados son eliminadas de dicho conjunto.

Cada iteracion del modelo evolutivo de forma, genera un sucesion de conjuntos optimos de Pareto que se aproximan a la solucion, y que representan un subdominio de formas permitidas.

Truncamiento a un dominio de formas permitidas

Despues de cada iteracion del M OP SO y debido al desplazamiento independiente de cada hito h_j , existe la posibilidad de que algunas part culas p_{ij} del conjunto optimo de Pareto puedan perder su vermiformidad. Esto es, que no pertenezcan al dominio de forma permitido D_f , o en otras palabras, que son siluetas vermiformemente no validas.
Para subsanar esta situacion, al subconjunto de elementos de M_i que no pertenezcan a un dominio D_f , es decir los $\tilde{f}$ se les calcula una forma proyectada $\hat{f}_i$ en el S_{D_f} mas cercana en forma (gura 3.28).
Cabe destacar que para aplicar este proceso, los elementos actuales del frente de Pareto son normalizados a la misma escala, rotacion y traslacion a los centroides. Luego de ser truncados, se devuelven a la escala original.

El proceso de obtencion de la nueva forma verme (dada en la seccion 3.5) es tal que:

$$\mathrm{trun}\, \tilde{f} = \hat{f}_i \quad \mathrm{con} \quad \hat{f}_i \in S_{D_f} \tag{3.43}$$

Cada forma truncada se encuentra en un n-s mplex de p centroides mas cercanos en forma a pe_{rj}. El conjunto total de centroides que determina el subdominio de formas permitido es dado a priori. Con el cambio propuesto en (3.43), se garantiza que el conjunto M_i posea unicamente elementos pertenecientes a D_f dados por:

$$M_i := \left\langle \hat{f}_{0j} \; \hat{f}_{1j} \; \ldots \; \hat{f}_{q-1j} \right\rangle^T \tag{3.44}$$

De igual forma, si la forma resultante tiene mejor la evaluacion tanto en longitud y en posicionamiento sobre la original de siluetas, se sustituye por la original, inclusive si la que se sustituyo es forma valida. Las part culas sustituidas son desechadas por el metodo y esto hace que el modelo no dependa de una cantidad exponencial de datos.

Despues de la aplicacion de M OP SO, cada forma $\hat{f}_i$ del frente de Pareto ha mejorado su posicionamiento y la forma respecto a la del nematodo en la imagen digital.

Evolucion del ajuste de la forma a traves del metodo del forrajeo

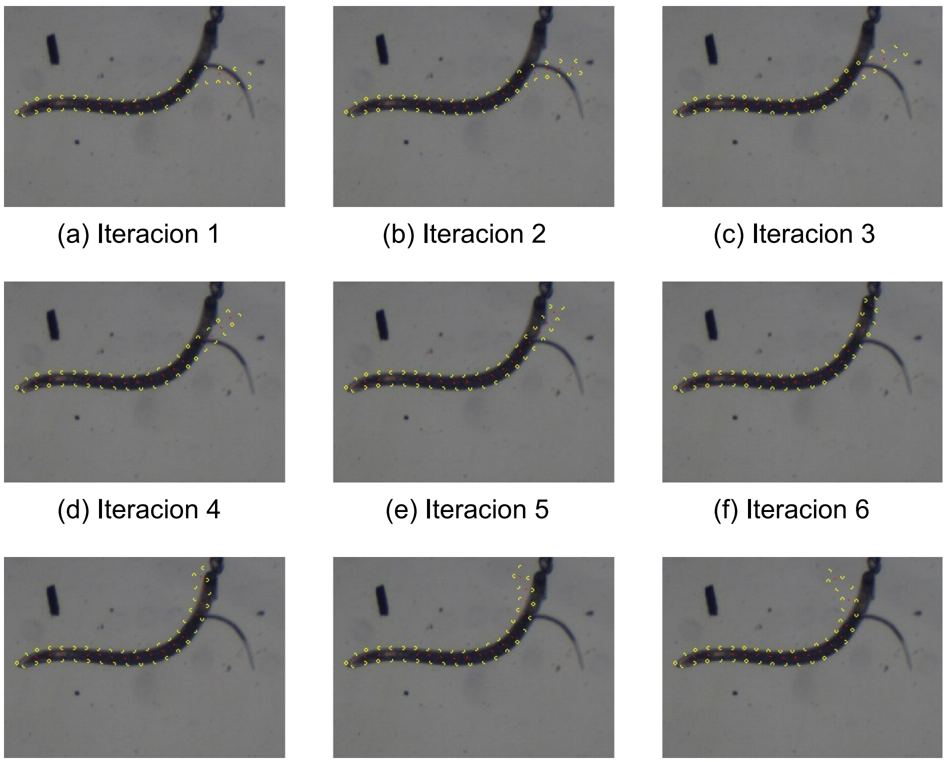
Subsiguientemente, a cada part cula $\hat{f}_i$ del frente de Pareto M_i se aplica el metodo del forrajeo con el objetivo de mejorar las respectivas postura de ajuste respecto al nematodo en la imagen digital.
A partir de la aplicacion de este metodo, se le permite a la siguiente iteracion del M OP SO buscar en un subdominio restringido de formas vermes o espacio muestral mas proximo a la forma verme buscada en la imagen digital.

Luego, a cada elemento de este conjunto se le aplican las dos partes del algoritmo del forrajeo (la busqueda lateral y el crecimiento), a los fragmentos cola y cabeza de manera independiente un numero de veces predeterminado.

La primera parte del metodo del forrajeo, ajusta las formas del frente de Pareto por las torsiones laterales independientes de los fragmentos de la cabeza y la cola de $\hat{f}_i$ (detallado en la seccion 3.4.2).
Con estos movimientos, se genera una sucesion de q formas vermes (gura 3.33), de las cuales es elegida la que posea mejor metrica de posicionamiento, esto es:

$$\max\left\{F_B\left(\hat{c},\hat{f}_0\right); F_B\left(\hat{c},\hat{f}_1\right); \ldots; F_B\left(\hat{c},\hat{f}_{q-1}\right)\right\} \tag{3.45}$$

para luego reemplazar a $\hat{f}_i$ en el frente de Pareto.



(a) Iteracion 1 (b) Iteracion 2 (c) Iteracion 3

(d) Iteracion 4 (e) Iteracion 5 (f) Iteracion 6

(g) Iteracion 7 (h) Iteracion 8 (i) Iteracion 9

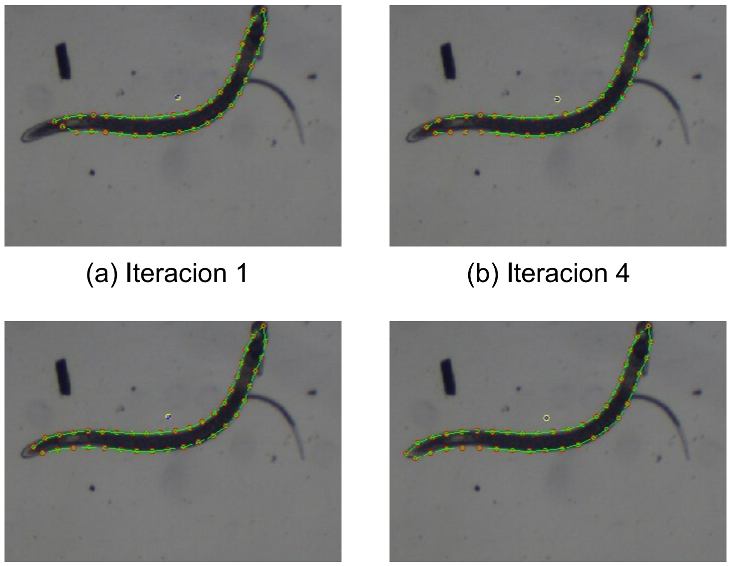
Figura 3.33: Iteraciones de la primera parte del metodo forrajeo para el fragmento cabeza con torsiones laterales. En la iteracion 6, se aprecia que la forma $\hat{f}$ posee mejor metrica de posicionamiento.

La segunda fase del forrajeo es el crecimiento o decrecimiento de la forma $\hat{f}$ dado por el algoritmo que se presenta en la seccion 3.4.3. Para determinar si $\hat{f}$ crece en el fragmento cabeza o cola, se calcula el valor de intensidad del p xel segun la posicion de los hitos y respectivas las vertebras[2]
Posteriormente, se comparan contra un umbral dado a priori que determine un valor

[2] El valor de intensidad del p xel determina si un hito h_j se ubica en la region interna del cuerpo del nematodo, en el borde o no esta dentro de esta region.

l mite promedio entre el fondo de la imagen digital y cuerpo del objeto a segmentar. Este corresponde a un numero real asociado al nivel de intensidad del p xel que provee un valor l mite entre el fondo de la imagen y el cuerpo del nematodo, y depende del conjunto de imagenes de entrenamiento y experimentacion. En este documento se toma = 0:3, en el rango de [0;1], como valor promedio de los p xeles del fondo y el cuerpo del nematodo.

Si el valor de intensidad de al menos alguno de los hitos o la vertebras del fragmento es menor al umbral, signi ca de ^{b}f se ubica en la region interna del cuerpo del nematodo o en su contorno, por lo que se debe crecer. En caso contrario, no esta en la region interna del cuerpo del nematodo, y debera decrecer.



(a) Iteracion 1 (b) Iteracion 4

(c) Iteracion 6 (d) Iteracion 9

Figura 3.34: Iteraciones de la segunda parte del metodo forrajeo para el crecimiento del fragmento de la cabeza. Se observa que a partir de la iteracion 6, hay un ajuste de forma ideal respecto al crecimiento.

Fin de la i-esima iteracion del modelo evolutivo de forma

El algoritmo inicia la segunda iteracion, al seleccionar al azar algunas part culas M_i y se construye una base datos vermes BD_2. Luego, se reinicia el proceso con M OP SO y hasta un numero de iteraciones N_{iter} de nido previamente, en la que el metodo haga converger la deformacion de la silueta al nematodo en la imagen digital.

El algoritmo nal del proceso completo de modelo evolutivo de forma usando M OP SO se muestra en la gura 3.35.

La siguiente iteracion de modelo evolutivo de forma, elige del conjunto optimo de Pareto una forma f^{b}_{i} aleatoria para generar una nueva base datos vermes por medio del metodo del serpenteo. Los nodos de los subdominios son las instancias vermes contenidas en el frente de Pareto, garantizando que los elementos resultantes en la caminata aleatoria

Entrada: C /* Conjunto de datos de entrenamiento */

I /* Imagen digital */

$i \leftarrow 0$ /* El iterador i */

$f_i \leftarrow random$ (C) /* Elección de f aleatorio */

C ← C r f_i /* Elimina f de C */

b (b b */

adjust ,

BD$_i \leftarrow rw$ b /* BD$_i$ por caminatas aleatorias */

mientras $i < N_{iter}$ **hacer**

c c b

M$_i$ ← MOPSO I , F$_D$, F$_B$, BD$_i$, f_i /* Cálculo del frente de pareto */

Para cada e$_p$ *de* M$_i$ **hacer**

$f_j \leftarrow trun$ C p e u w /* Proyectar e$_p$ al /* S$_{Df}$ de w nodos */ más cercano de los u */

fin

Para cada f_j *de* M$_i$ **hacer**

b ← *forrg* b b I b

fin

$i \leftarrow i + 1$

b b b

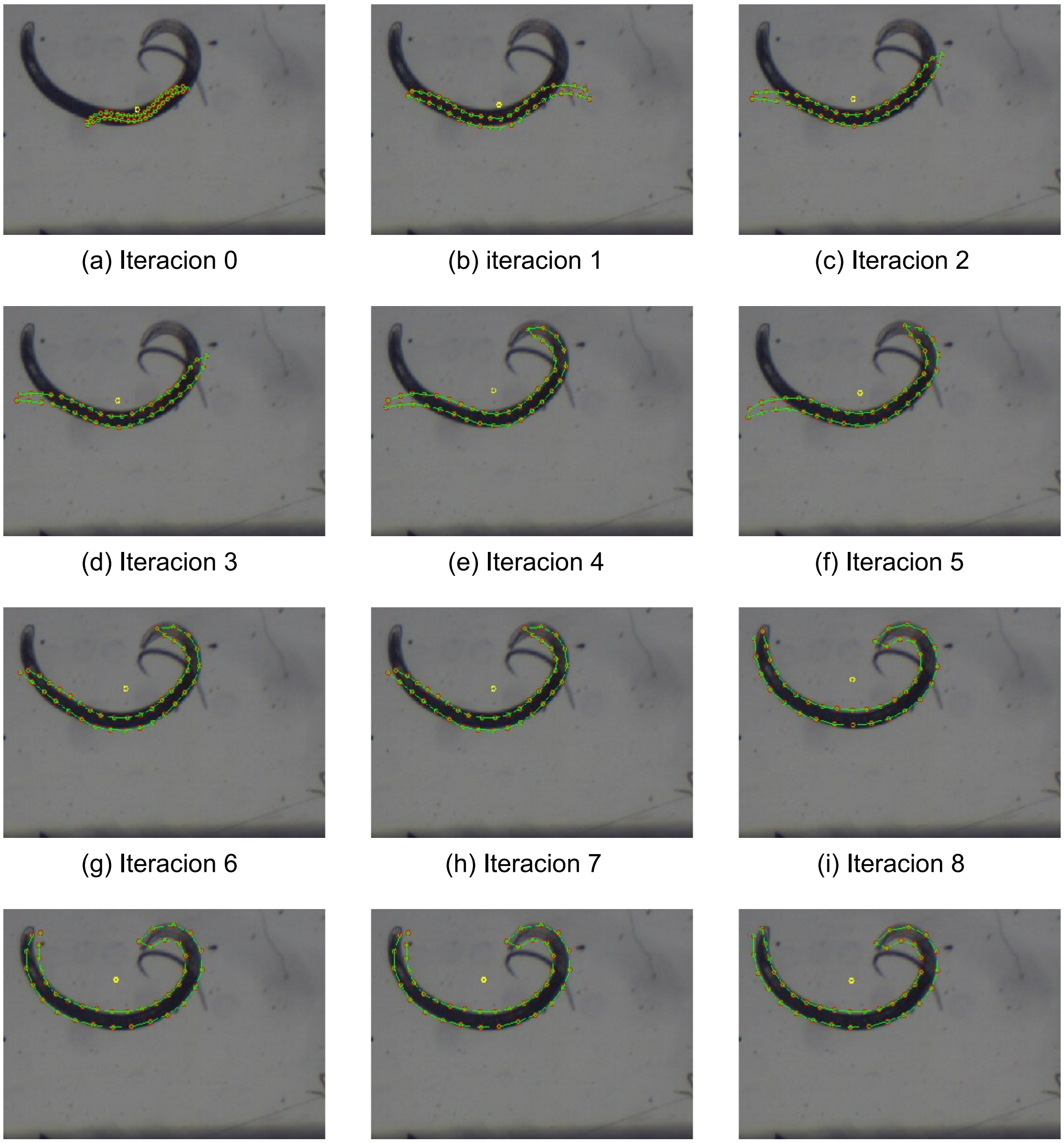
$f_i \leftarrow random$ (M$_i$) /* Elección de f_i */

BD$_i \leftarrow snk$ f_i /* BD$_i$ por el método del serpenteo */

fin

Figura 3.35: Algoritmo del Modelo adaptativo de forma usando M OP SO.

sean a nes en forma al nematodo en la imagen digital. Luego, se reinicia el proceso con M OP SO y hasta un numero de iteraciones N$_{iter}$ de nido previamente en el que converja la deformacion de la silueta. En la gura 3.36 se muestran 12 iteraciones del modelo de ajuste propuesto de un nematodo en una imagen digital.

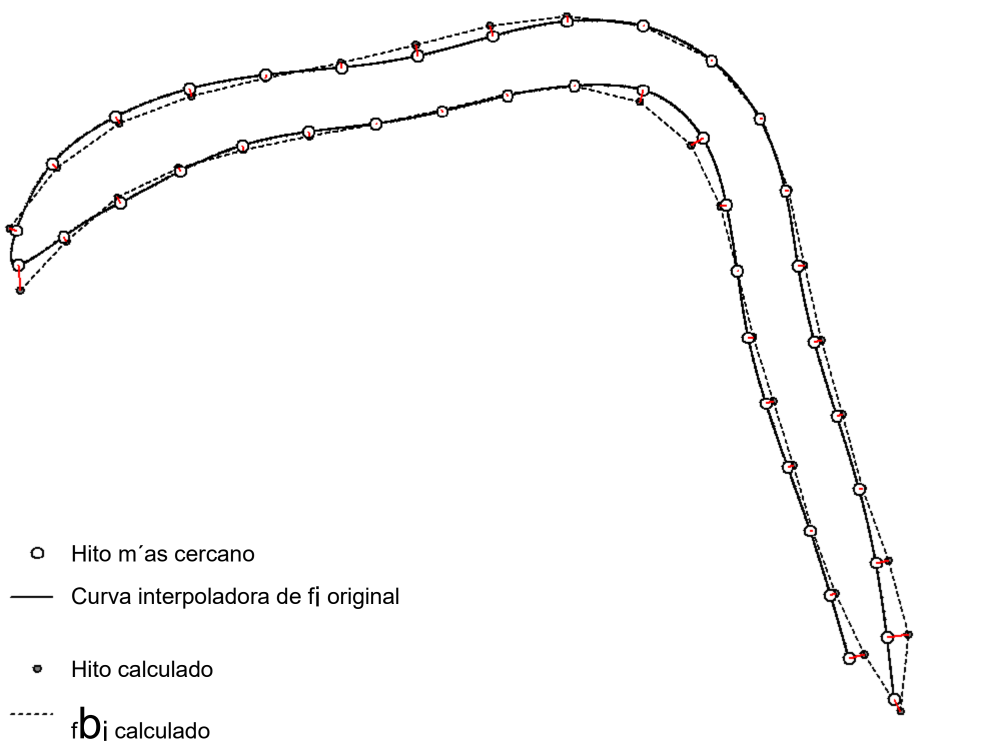


(a) Iteracion 0
(b) iteracion 1
(c) Iteracion 2
(d) Iteracion 3
(e) Iteracion 4
(f) Iteracion 5
(g) Iteracion 6
(h) Iteracion 7
(i) Iteracion 8
(j) Iteracion 9
(k) Iteracion 10
(l) Iteracion 11

Figura 3.36: Iteraciones del algoritmo Modelo adaptativo de forma usando M OP SO.

3.8. Medida de ajuste a la forma verme

Para medir el ajuste de la silueta $\hat{f}_i$ al nematodo en la imagen digital, se utiliza el promedio de las distancias ℓ_2 entre los hitos de forma verme calculado por MEF y el hito mas cercano de la curva interpolante C_{sp} generada por los hitos anotados originalmente.
Se selecciona como el hito referencia el mas cercano perteneciente a la curva trazadora interpolante C_{sp} de los f_i originales, excepto para los hitos numerados 0 y $\frac{n}{2}$ (correspon-diente a cabeza y cola) de la forma obtenida $\hat{f}_i$, a los cuales se les calcula la distancia a los hitos numerados 0 y $\frac{n}{2}$ (correspondiente a cabeza y cola) o $\frac{n}{2}$ y 0 (correspondiente a cola y cabeza) respectivamente de f_i original, elegidos segun su cercan a: los mas proximos (ver gura 3.37).



1

Figura 3.37: Representacion de la medicion de los hitos mas cercanos al $\hat{f}_i$ a la silueta original para 40 hitos anotados. Los segmentos coloreados en rojo proveen la distancia entre la curva interpolante dada por los hitos originales y los hitos calculados por MEF (Medida de ajuste a la forma verme).

Cabe destacar que el criterio de cercan a esta dado por la menor distancia ℓ_2. En adelante, a esta metrica se le denominar medida de ajuste a la forma verme.

Cap tulo 4

Resultados experimentales y analisis

En este cap tulo se presentan los resultados del metodo propuesto. En primera instancia, se describen los datos experimentales. Luego, a traves la experimentacion con distintas variantes del modelo, se valida el comportamiento del modelo de ajuste de la forma a traves de los parametros de con guracion del MEF (numero de coordenadas baricentricas, numero de centroides, numero de hitos por instancia, adaptabilidad y variabilidad de nuevas formas y numero de iteraciones del algoritmo) por medio de la tecnica de validacion cruzada para k-iteraciones para $k = 10$ en todos los casos, y el ndice de Jaccard.

Asimismo, la metrica dada en la seccion 3.8 se aplica para la evaluacion del ajuste de las formas vectoriales vermes a las contenidas en las imagenes digitales. En los experimentos referentes a las distancias, estas se mensuran tomando el p xel como unidad basica de medida. Las bases de datos de imagenes utilizadas con nematodos se detallan en la seccion 3.1, y la de formas vermes expresadas vectorialmente en hitos se pre-procesaron de acuerdo a los metodos explicados en la seccion 3.2. En esta misma seccion se detalla la distribucion de las longitudes originales de las formas vermes.

Ademas, en todas las fases experimentales, la forma inicial $\hat{f}_0$ es generada de forma sinteti-ca y centrada respecto al centro de masa de la imagen digital. Por consiguiente, $\hat{f}_0$ se escalo a un tama~no de manera que su longitud de esqueleto E no exceda la mitad del m nimo entre las dimensiones (largo y ancho) de la imagen digital del nematodo a la que se le aplico el algoritmo.

La evaluacion del algoritmo se presenta cuatro fases: primeramente, en la seccion 4.1 se analiza el numero de hitos optimo para la representacion vectorial de las formas vermes. En la seccion 4.2 se analiza el alcance del algoritmo propuesto para generar y emular nuevas formas vermes. En la seccion 4.3 se explora el numero de centroides del dominio de formas y el numero de coordenadas baricentricas necesarias para engendrar los subdominios de forma. Finalmente, en la seccion 4.4 se determina la e cacia del algoritmo a traves del ndice de Jaccard, donde se provee el porcentaje de exito del algoritmo propuesto. La secuencia del proceso evaluativo se ilustra en la gura 4.1.

Cabe destacar que las pruebas realizadas a nivel de modelo de forma se hicieron con

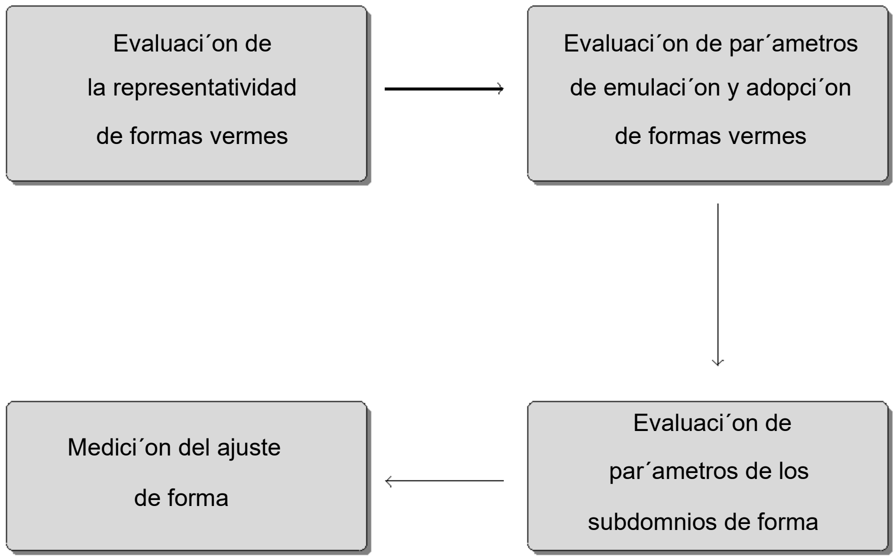


Figura 4.1: Diagrama de ujo de evaluacion de las partes del algoritmo MEF.

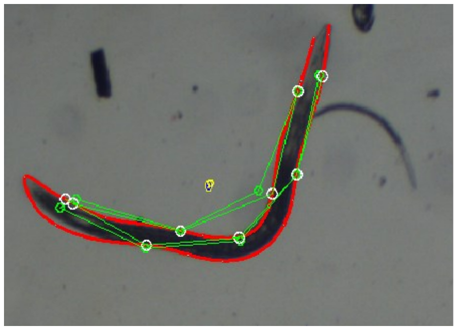
una base de datos normalizadas de 2744 instancias de nematodos expresadas en hitos de frontera. Asimismo, las pruebas del modelo a nivel de imagen se ejecutaron con una base de datos de 667 imagenes digitales que cumplen las restricciones explicadas en la seccion 3.1.

4.1. Representacion vectorial de la forma verme a traves de hitos

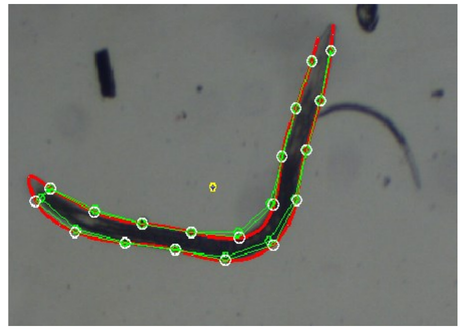
A partir del proceso de anotacion de los hitos en las 667 imagenes digitales, se obtuvieron las siluetas dadas por los $\hat{b}_{f_i}$. Cada una de estas se caracteriza por tener la misma orientacion en la secuencia de su anotacion (sentido contrario a las manecillas del reloj) y su respectiva forma simetrica.

Para las fases experimentales del MEF, y con el n de satisfacer las premisas de represen-tatividad propuestas en la seccion 3.2.1, se debe determinar el numero m nimo de hitos que en conjunto cubran la totalidad de la silueta y extraigan la mayor informacion posible del contorno del nematodo.

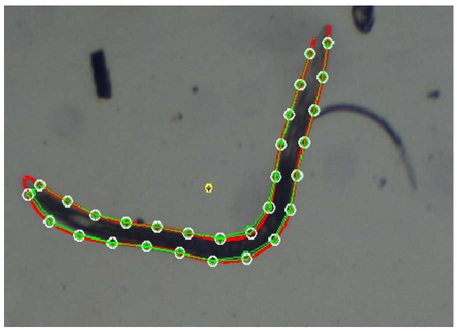
Para determinar dicho numero, se usa la base de datos de formas vermes y la base de datos de imagenes digitales. Se particiona la curva de las silueta vermes en 10, 20, 30, 40, 50 y 60 hitos equidistantes usando el metodo de trazadores cubicos interpolantes C_{sp} para los hitos anotados manualmente. Un ejemplo de las diferentes particiones se aprecia en la gura 4.2.



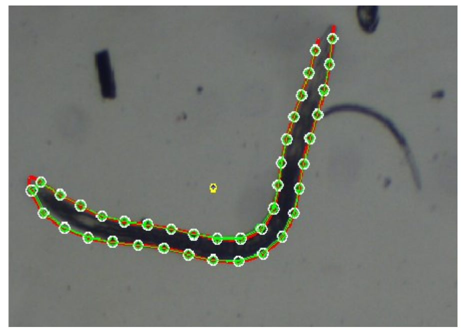
(a) Silueta representada por 10 hitos



(b) Silueta representada por 20 hitos



(c) Silueta representada por 30 hitos



(d) Silueta representada por 40 hitos

Figura 4.2: Siluetas de un nematodo segmentado por numero de hitos anotados en color verde y su respectiva curva de representacion en color rojo dada por C_{sp}.

El numero de hitos de f^{b}_{i} ideal se mide a traves dos pruebas y tomando como referencia el punto intermedio de cada dos hitos consecutivos h_j y h_{j+1} anotados manualmente.
En la primera prueba, se promedia la medida de ajuste a la forma verme a cada hito intermedio. En la segunda, se aplica la metrica de bordicidad F_B con el n de determinar el mejor ajuste de acuerdo con el numero de hitos. Con estas medidas se determina la cercan a de los hitos intermedios al borde del nematodo en la imagen digital, y en consecuencia, toda la silueta original de f^{b}_{i}.
Los resultados en la gura 4.3 muestran que a menor cantidad de hitos anotados, la distancia entre los hitos intermedio y la curva de representacion verme es en promedio mayor.

Es decir, las representaciones vermes vectoriales con 10, 20, 30 hitos muestran un descenso en la curva de error medio cuadratico de las distancias promedios entre los hitos de la curva C_{sp} y los calculados. A partir de los 30, se estabiliza la curva y para 40, la distancia promedio es aproximadamente menor que 1 p xel por hito.

En la gura 4.4 se muestran los resultados de la segunda prueba, dejando en evidencia que para 30 hitos o menos, hay perdida de informacion de bordicidad. A partir de los 40 hitos de representacion de la forma verme, la informacion obtenida del contorno es la misma. Para un numero mayor a 40 hitos, proporcionar a informacion redundante.

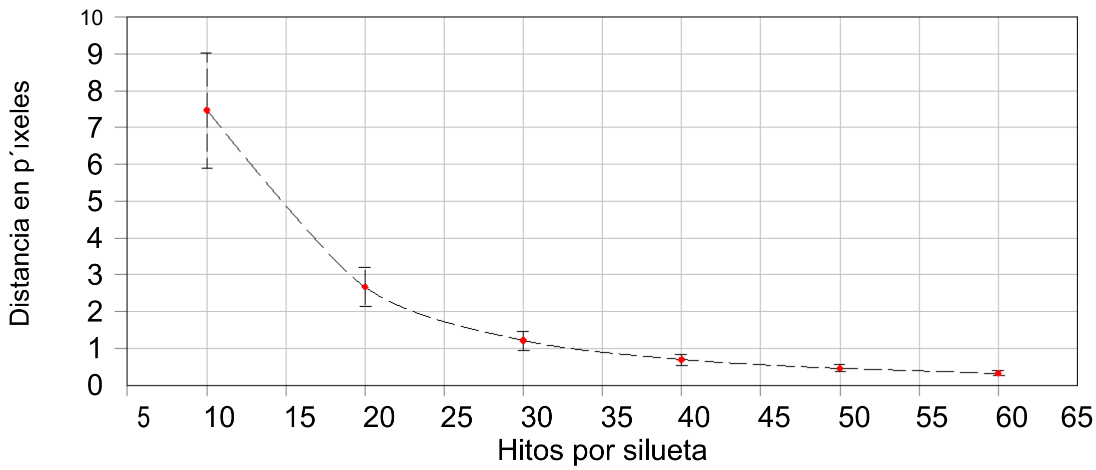


Figura 4.3: Distancias de los hitos promedio a la curva trazadora. La barra vertical denota la desviacion estandar de cada medicion promedio.

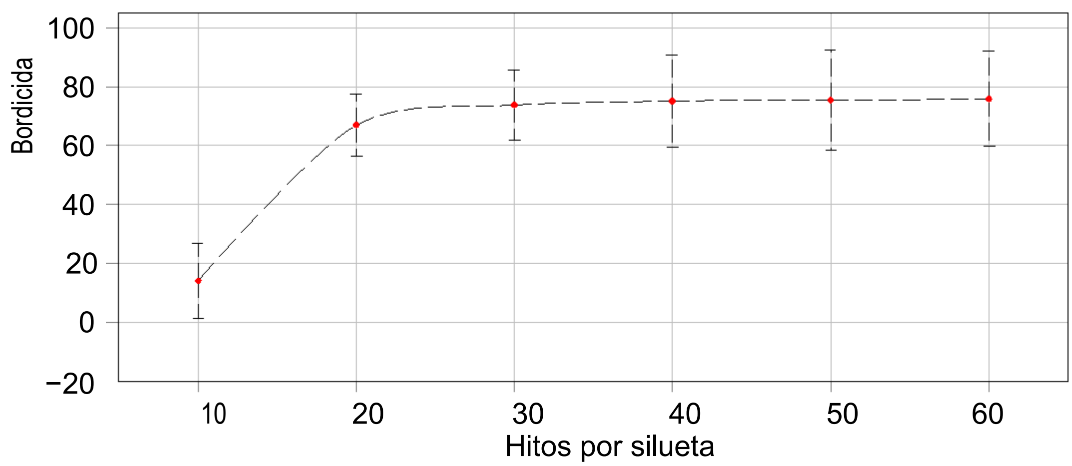


Figura 4.4: Promedio de la evaluacion de la funcion de bordicidad por hito. La barra vertical denota la desviacion estandar de cada medicion promedio.

De ambas pruebas, se evidencia que con 40 hitos se logra una representacion verme de contorno de compromiso aceptable y sin redundancia (gura 4.5). Para las pruebas experimentales subsiguientes, sera tomado dicho numero de hitos por ser el que provee mejor representatividad sin perdida o exceso de informacion.

Cabe destacar el costo asociado al recurso computacional de memoria y complejidad algor tmica crece al menos de forma lineal con el numero de hitos seleccionado. Es decir, a mayor cantidad de hitos se tiene un mayor costo computacional. En consecuencia y segun los resultados obtenidos, la leve mejora que se obtiene para 60 hitos por silueta no justi ca su uso, aspecto por el cual se decide trabajar con 40 hitos.

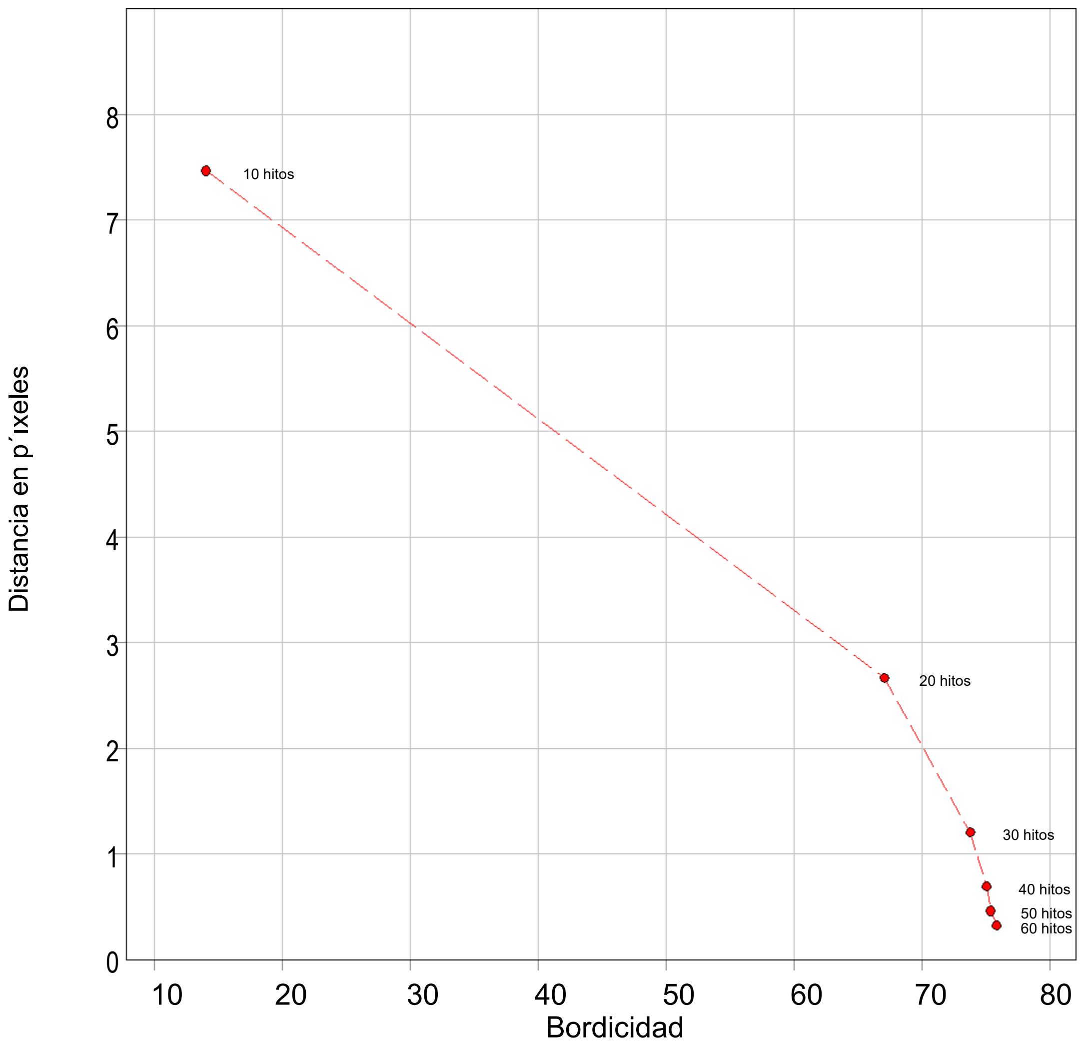


Figura 4.5: Promedio de la evaluacion de la funcion de bordicidad por hito y medida de ajuste a la forma verme al hito intermedio.

4.2. Emulacion y adopcion de nuevas formas vermes usando ACP K y MEF

En este apartado, se busca determinar la capacidad del MEF de emular o adoptar cualquier forma de gusano en un dominio de formas validas. El experimento se divide en dos partes.

En la primera se realiza un comparativo entre el modelo ACP K y estimaciones de las coordenadas baricentricas dentro de un n-s mplex a traves del metodo del serpenteo. El objetivo es analizar la capacidad de los metodos de adoptar cualquier deformacion de gusano y mantener su vermiformidad. El proceso se explica en la seccion 4.2.1.

Ademas, El numero de hitos por silueta es 40 debido a los resultados obtenidos en la seccion 4.1.

En la seccion 4.2.2 se detalla la segunda parte de este experimento donde el objetivo es determinar valores espec cos de los parametros del metodo de simulacion de movimientos aleatorios vermes para generar nuevas siluetas vermes a traves de caminatas aleatorias, calculadas a traves de sucesiones de saltos aleatorios entre subdominios vecinos S_{Df} .

En ambas partes de la experimentacion se utiliza una medida de similitud entre siluetas vermes denominada ndice de cercan a denotado . Esta mide la distancia ℓ_2 promedio entre el desplazamiento de hitos respectivos de dos formas vermes $\hat{f}_p$ y $\hat{f}_q$. Viene dada por:

$$d\left(\hat{f}_p, \hat{f}_q\right) \tag{4.1}$$

con

$$d\left(\hat{f}_p, \hat{f}_q\right) := \frac{1}{n}\sum_{j=0}^{n-1} \left\| \hat{h}_{pj} - \hat{h}_{qj} \right\|_2 \tag{4.2}$$

En este ndice de cercan a se asume que $\hat{f}_p$ y $\hat{f}_q$ son previamente normalizadas segun el metodo dado en la seccion 3.2.1.

4.2.1. Aproximacion de siluetas vermes

El objetivo de esta seccion es evaluar la capacidad del modelo de adoptar cualquier silueta de nematodo. Para esta prueba, se parte de la premisa que $\hat{f}_a$ es proxima o cercana a la forma de la silueta $\hat{f}_c$ si su ndice de cercan a dada en (4.1) es menor a un previamente dado. En los experimentos de esta seccion se tomar como 1 p xel.

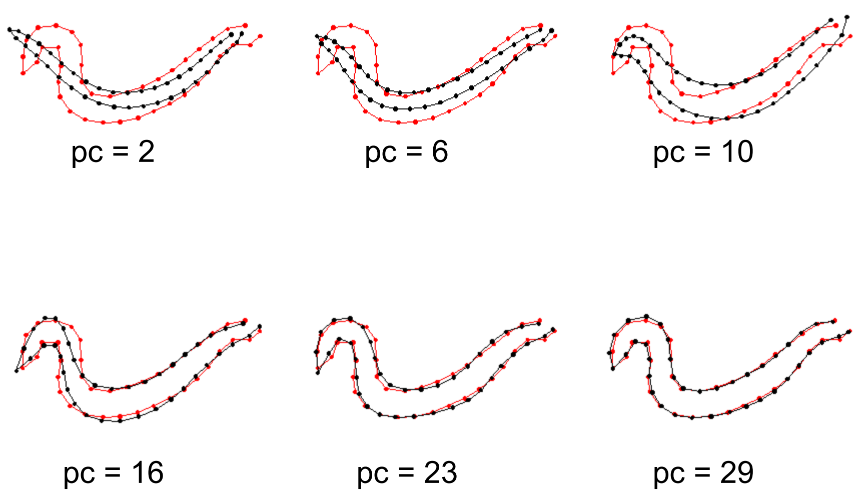
Para este experimento, se realiza un comparativo entre dos modelos de reconstruccion de formas. La primera es mediante metodos de reduccion de dimensionalidad, y la segunda es mediante combinaciones baricentricas en s mplexes geometricos por medio del metodo del serpenteo.

Para los metodos de reduccion de dimensionalidad, se analiza la cantidad de componentes principales necesarios para reconstruir una silueta verme usando ACP y ACP K. En el caso de las combinaciones baricentricas en s mplexes geometricos, se hace la reconstruccion de las siluetas vermes en subdominios de forma engendrados por nodos calculados por el metodo de los k vecinos mas cercanos.

En ambos experimentos, se utiliza una muestra de 1000 formas vermes f^{b}_{a} generadas aleatoriamente usando el metodo del serpenteo descrito en la seccion 3.4, normalizadas segun el metodo descrito en 3.2.1, independientes a los individuos que conforman la base de datos de entrenamiento. Los resultados se muestran en las siguientes subsecciones.

Aproximacion de siluetas vermes usando ACP y ACP K

La reconstruccion de siluetas vermes usando los metodos de reduccion de dimensionalidad explicados en la seccion 2.3 permiten que en un dominio de formas permitidas se pueda reconstruir las siluetas vermes usando una cantidad menor de parametros que moldean su forma dados por los componentes principales. En la gura 4.6 se muestra la reconstruccion de una forma verme usando el metodo ACP K para un numero de componentes principales diferente en cada iteracion.



1

Figura 4.6: Ejemplo de reconstruccion de una silueta de un nematodo usando ACP K con una funcion kernel radial gaussiana. En negro se dibujan las siluetas calculadas en relacion con la cantidad de componentes principales (pc), respecto a la silueta roja.

Cabe resaltar que para la reconstruccion de las siluetas de gusano es necesario que a lo sumo se logre un p xel de distancia promedio por hitos correspondiente para que el modelo evolutivo de forma no se vea inuenciado por error en el desplazamiento.

En el caso de ACP , se reconstruye la silueta $\hat{f}_c$ utilizando n componentes principales obtenidos de la matriz de representacion de la transformacion lineal ortogonal dada por (2.5).

Para el metodo de ACP K, una silueta verme $\hat{f}_c$ se mapea al espacio caracter stico F a traves de una funcion kernel dada por $(\hat{f}_c)$. Esta a su vez se proyecta a un espacio de menor dimension. El vector $(\hat{f}_c)$ corresponde al vector proyectado en el espacio de menor dimension.

La preimagen de $(\hat{f}_c)$ en el espacio de entrada es $\hat{f}_a$. Esta se calcula a traves de la variacion del metodo por mapeos conformes explicado en la seccion 2.3.2. El proceso completo del ACP K para las siluetas $\hat{f}_c$ y $\hat{f}_a$ se ilustra en la gura 4.7.

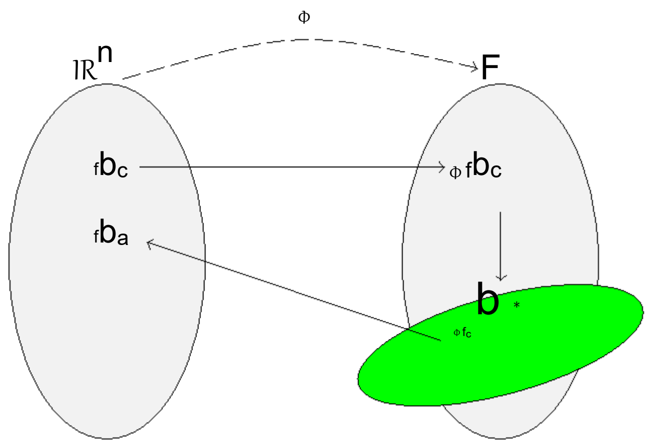


Figura 4.7: Diagrama del metodo de ACP K para $\hat{f}_c$ y $\hat{f}_a$. En verde se colorea el hiperplano de proyeccion para $(\hat{f}_c)$ para un numero n de componentes principales. $(\hat{f}_c)$ corresponde al valor proyectado en el dicho hiperplano correspondiente a $(\hat{f}_c)$.

Para realizar el analisis comparativo de ambos metodos, es menester primeramente indagar sobre los resultados producidos por las variantes de hiperparametros y funciones kernel en el ACP K. En las funciones kernel polinomial, se estudia el efecto de los hiperparametros como el grado del polinomio n y el sesgo c, en el caso de las funciones radial gaussiana y exponencial se hizo con diferentes valores de .

En la gura 4.8 se muestran los resultados de la reconstruccion de las formas respecto a la cantidad de componentes principales y el valor de la funcion kernel proyectiva exponencial. Segun los resultados de la gra ca, la reconstruccion de las formas vermes mejoran con el aumento de la cantidad de componentes principales y del valor . A partir de los 25 componentes principales y $> 4\ 10^3$, se logra reconstruir la forma verme con una distancia menor a 2 p xeles promedio entre hitos correspondientes de las formas calculadas y aproximada.

En la gura 4.9 se muestran los resultados del metodo ACP K con la funcion kernel radial gaussiana. Al igual que el caso anterior, se muestra una tendencia de reconstruccion menor a un pixel de distancia entre hitos correspondientes a partir de los 23 componentes principales y para > 3:5 10^3.

En el caso de las funciones kernel polinomiales, se evalua el comportamiento de la reconstruccion de las siluetas vermes para los hiperparametros c y n respecto al numero de las componentes principales. Los resultados mostrados en 4.10 indican que el polinomio con n = 1 aproxima mejor que el caso de n = 2. El valor de c no presenta variacion.

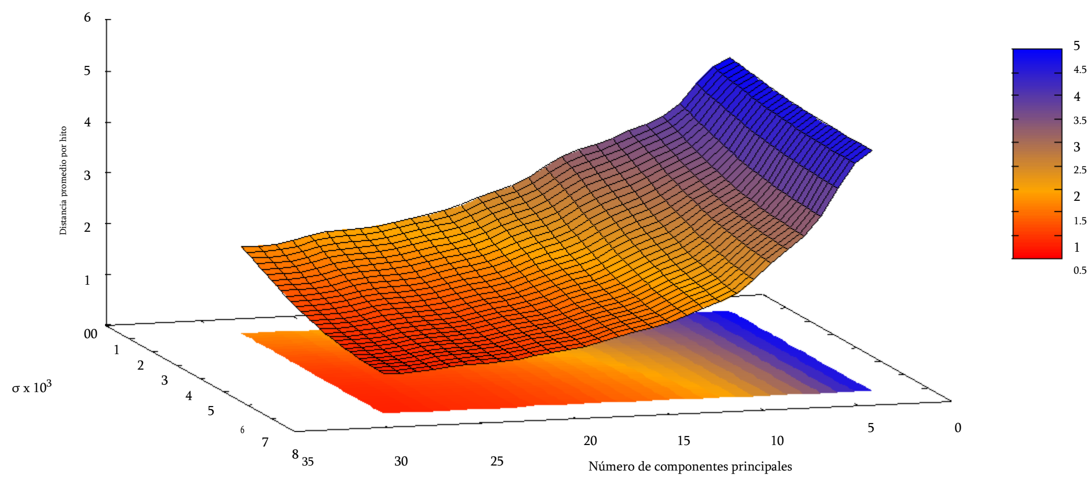


Figura 4.8: Medicion de los componentes principales usando como kernel la funcion proyectiva exponencial exp(hx_i; x_ji=(2 ^{2})) para valores de .

Para todas la variantes del ACP K analizadas, se logra una reconstruccion de la forma verme a partir de 23 componentes principales. El mejor resultado se obtuvo para la fun-cion kernel polinomial usando n = 1, con menos de dos p xeles de distancia entre hitos respectivos.

Consiguientemente, los resultados de la aplicacion de los metodos ACP y ACP K para la reconstruccion de la forma verme se muestran en la gura 4.11. En esta se muestra el metodo ACP tiene un leve mejor a que el ACP K para la reconstruccion de las formas vermes cuando pc < 10; sin embargo, la distancia entre sus hitos respectivos es mayor que 4 p xeles.

Por otra parte, tanto para ACP como para ACP K se necesitan al menos 20 componentes principales para reconstruir la silueta con al menos 1 p xel de distancia entre sus hitos, es decir, al menos un 25 % de los componentes principales son necesarias para reconstruir las siluetas en forma de gusano.

En este mismo analisis, es preciso resaltar que a partir de los 20 componentes principales no hay diferencias signi cativas entre ambos metodos pues aproximadamente hay un p xel de diferencia entre los hitos respectivos, haciendo que los metodos se comporten de manera similar sin diferencias signi cativas.

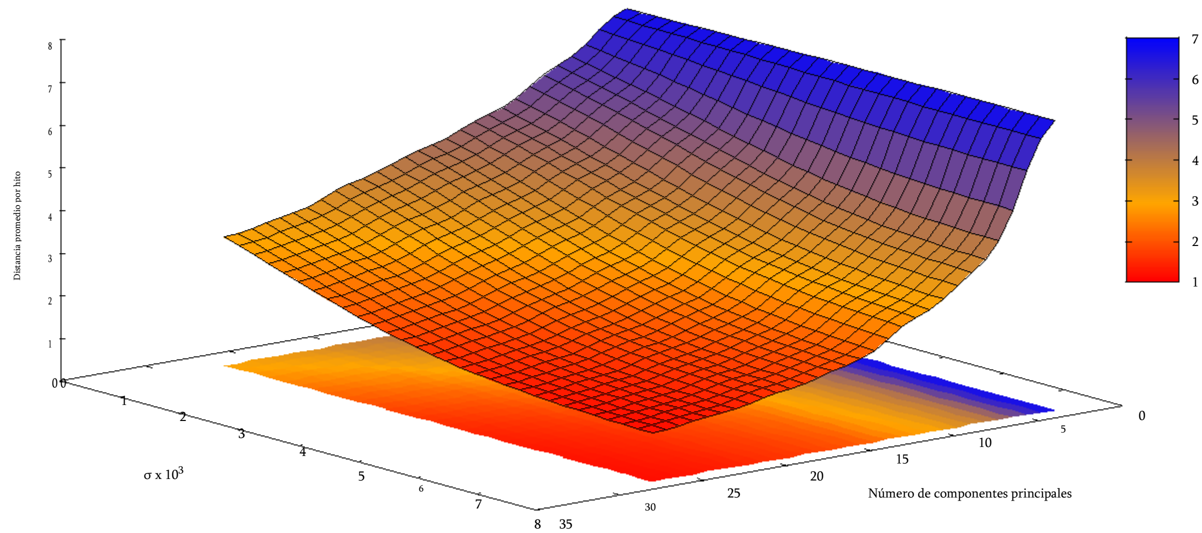


Figura 4.9: Medicion de los componentes principales usando como kernel la funcion radial gaussiana exp(kx_i x_jk^2=(2 ^{2})) para valores de .

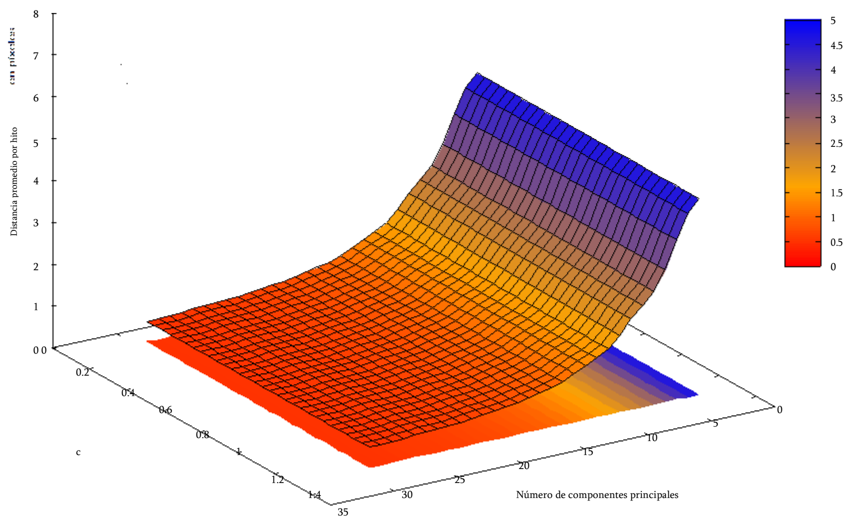
Asimismo, tanto en ACP como ACP K, la reconstruccion de la silueta depende de un numero de componentes principales. Esto es, que entre mayor sea el numero de componentes entonces mejor sera el numero su reconstruccion.

Por otra parte, para el ACP K se experimenta sobre el efecto que produce el error intr nseco inducido por la busqueda de la preimagen en el proceso iterativo de ajuste respecto al cambio en la forma del nematodo o la perdida de la vermiformidad. El objetivo es analizar la estabilidad del metodo. Para esto se mide la distancia promedio por hito de una forma inicial $\hat{f}_c$ a cada elemento de la sucesion $\hat{f}_0$; $\hat{f}_1$; ; $\hat{f}_{m\ 1}$.

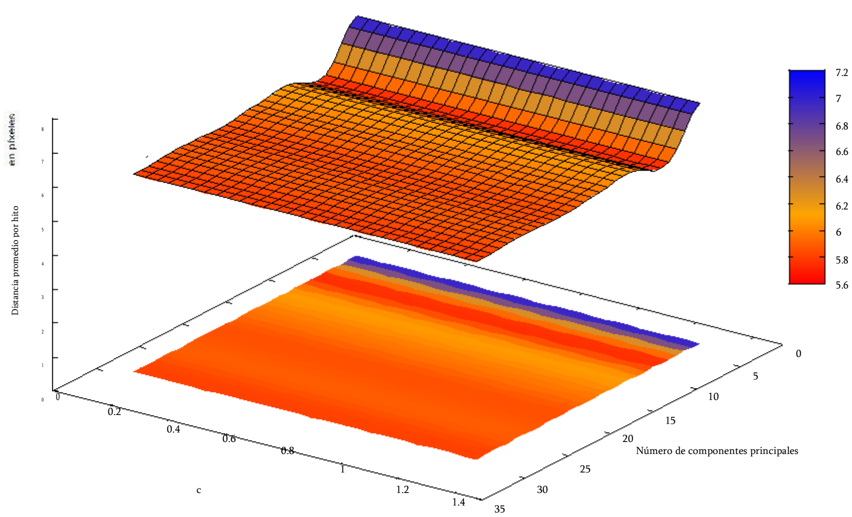
La sucesion de formas vermes es obtenida a traves un proceso iterativo del ACP K. El pro-ceso inicia eligiendo una forma verme calculada $\hat{f}_c$ la cual corresponde al primer termino de la sucesion $\hat{f}_0$. A esta se le calcula su imagen en el espacio caracter stico F a traves de la funcion kernel polinomial. Luego, se proyecta al espacio de menor dimension y posteriormente se calcula su preimagen $\hat{f}_a$ en el espacio de entrada para n componentes principales.

La forma $\hat{f}_a$ se asigna a $\hat{f}_1$ que corresponde al segundo termino de la sucesion . El proceso se repite sustituyendo $\hat{f}_1$ por $\hat{f}_c$, por n iteraciones hasta obtener la sucesion deseada. Los resultados se exhiben en la gura 4.12 y muestran el promedio de las distancias entre hitos correspondientes para $\hat{f}_i$ por $\hat{f}_c$ con 1 i 7.

Los resultados muestran que el ndice de cercan a de la forma calculada y cada una de las formas aproximadas de la sucesion generada por el ACP K tienen un aumento abrupto.



(a) n = 1.



(b) n = 2.

Figura 4.10: Distancia promedio entre hitos respectivos usando la funcion kernel polinomial $(hx_i; x_ji + c)^n$ para valores de c.

Es decir, en la sexta iteracion para 25 componentes principales la distancia entre hitos se ha degenerado por mas de 34 p xeles. Un efecto similar sucede para las otras mediciones de componentes principales mostradas en la gra ca. Lo anterior permite concluir con el ACP K, que el error derivado del calculo de la preimagen provoca una perdida de forma

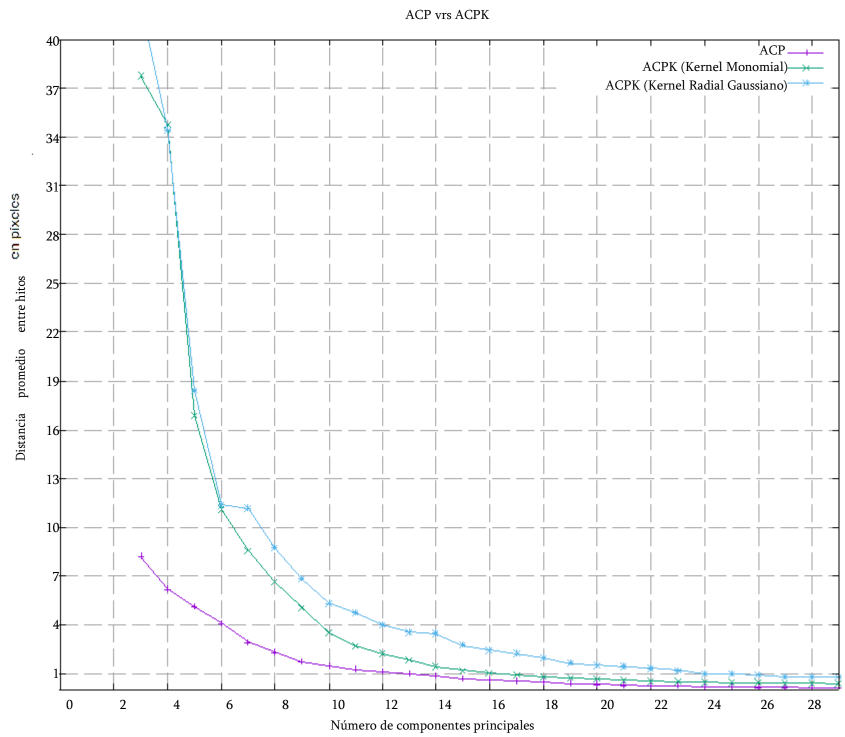


Figura 4.11: Distancia promedio por hito entre f bc y f b a usando ACP K

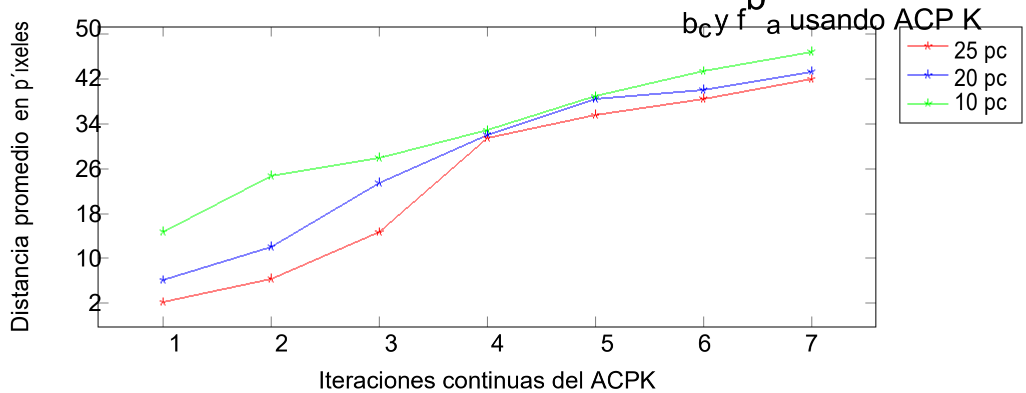


Figura 4.12: Iteraciones continuas del ACP K para componentes principales con el kernel polinomial. Cada iteracion corresponde a una forma verme calculada.

de la silueta original.

Finalmente, se evaluo respecto el metodo del ACP K y del serpenteo sucesiones de formas y su ndice de cercan a respecto al nematodo mas cercano contenido en el subdominio de forma mas cercano. Se utilizo el ndice de cercan a para la forma calculada y la pertene-ciente al dominio de forma mas cercano. Los resultados se muestran en la gura 4.13.

En este se muestra que el metodo del serpenteo promedia entre 3 y 5 p xeles de ndice de cercan a a partir de la quinta iteracion. Contrariamente, el metodo de ACP K aumenta la distancia promedio a mas de 30 p xeles entre hitos respectivos. En consecuencia, muchas de las part culas no son vermiformente validas.

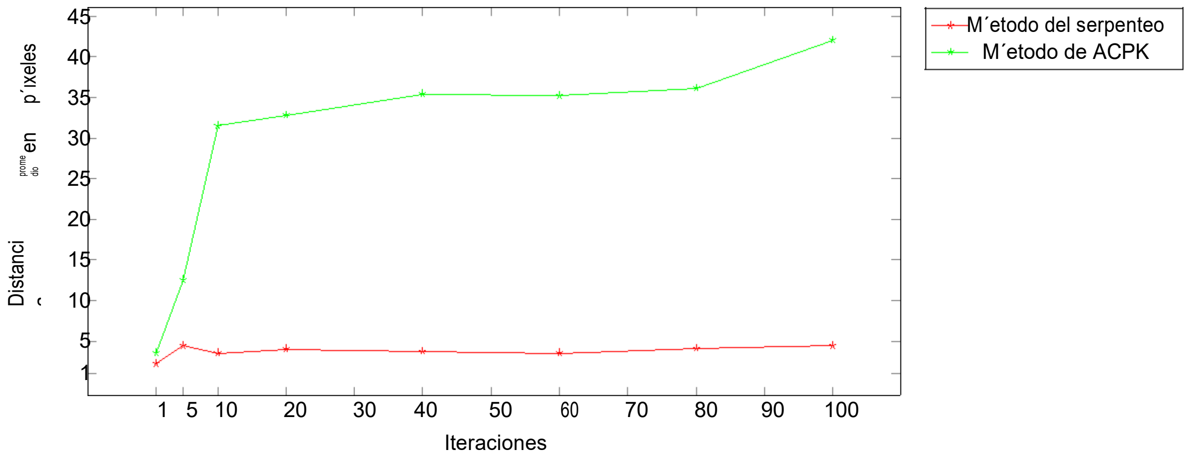


Figura 4.13: Promedio del ndice de cercan a entre formas calculadas con los metodos ACP K con kernel polinomial y el metodo del serpenteo, respecto a la forma verme calculada perteneciente al subdominio de forma mas cercano. Los subdominios de forma se calcularon con 10 nodos.

Finalmente, se concluye que el metodo del serpenteo s logra conservar la vermiformidad entre sucesiones a traves de caminatas aleatorias. En la aplicacion sucesiva del metodo ACP K, se degrada la vermiformidad de las siluetas llegando a perderla.

Aproximacion de siluetas vermes a traves del metodo Serpenteo

La segunda prueba muestra que cualquier forma verme $\hat{f}_a$ puede ser aproximada en algun subdominio S_{D_f}. Se parte de la premisa que $\hat{f}_a$ es vermiformente valida si existen $f_0; f_1; \ldots; f_k \in D_f$ tales que la engendran o la aproximan a un S_{D_f} como una combinacion lineal baricentrica basada en la ecuacion de los s mplexes geometricos dada en (2.38).

El experimento de aproximacion de cualquier forma verme $\hat{f}_a$ se divide en dos etapas. En la primera, se determina el S_{D_f} al cual pertenece. Y la segunda etapa calcula las coordenadas baricentricas que mejor aproximan la forma en dicho subdominio.

El proceso se inicia calculando el conjunto de los vecinos mas cercanos $\{\hat{f}_0; \hat{f}_1; \ldots; \hat{f}_{k+1}\}$ a $\hat{f}_a$ utilizando FLANN (Muja y Lowe, 2009), ordenados segun su lejan a. Luego, se centra el conjunto sobre $\hat{f}_0$ rede niedo elemento de la forma:

$$\hat{f}_i := \hat{f}_i - \hat{f}_0 \qquad (\forall i; 0 \leq i \leq n+1) \tag{4.3}$$

El S_{D_f} es generado por los k vecinos resultantes:

$$\{\hat{f}_1; \hat{f}_2; \ldots; \hat{f}_{k+1}\} \tag{4.4}$$

La segunda parte es fundamentada por una variacion del metodo dado por matching pursuit usado en este contexto (Donoho et al., 2012). En esta se aproximan las coordenadas baricentricas:

$$= [\ {}_1;\ {}_2;\quad;\ {}_{k+1}] \tag{4.5}$$

que determinan el S_{Df} a traves de la ecuacion:

$$\hat{f}_a \quad {}_1\hat{f}_1 + {}_2\hat{f}_2 + \ + \ {}_k\hat{f}_k + {}_{k+1}\hat{f}_{k+1} = \hat{f}_c \tag{4.6}$$

donde cada coordenada baricentrica se trunca al intervalo [0;1[mediante:

$${}_i = \mathrm{m\ n}(1; \max(0; {}_i)) \tag{4.7}$$

bajo la restriccion:

$$\sum_{i=1}^{k+1} {}_i = 1 \tag{4.8}$$

Ademas, para la caracterizacion de las formas $\hat{f}_c$ y las respectivas $\hat{f}_a$ se utiliza la distancia ‘$_2$ resultante del promedio entre las de los hitos respectivos de ambas siluetas. Esta carac-terizacion mide las distancias de similitud entre $\hat{f}_c$ y $\hat{f}_a$. Se evalua la distancia promedio en funcion de la cantidad de centroides con el n de analizar la mejora en la aproximacion de la silueta. Los resultados se muestran en la gura 4.14.
En el caso de 100 centroides para 1000 formas vermes, la curva muestra que el ajuste presenta un comportamiento erratico: la curva no sigue una tendencia de nida entre creciente y decreciente. Esto indica que en algunos casos se ajusto mejor y en otros no, independientemente del numero de nodos que conforman el n s mplex.

Lo anterior se debe a que en esta iteracion se cuenta con una menor cantidad de centroides, haciendo que cada $\hat{f}_a$ sea engendrada por nodos mas lejanos entre s . Si los nodos son mas lejanos, seran menos s miles en su forma, y por lo tanto, se provoca que la forma aproximada dependa de que tan cercano sean a uno o varios nodos del n s mplex.
Asimismo, hay una tendencia de mejora en la aproximacion $\hat{f}_c$ sobre $\hat{f}_a$ cuando aumenta el numero de centroides que determinan los S_{Df} . Desde una particion con 300 centroides del D_f para formar los S_{Df} , se muestra que aproximadamente la distancia por hito entre las siluetas original y calculada es menor a 1 pixel promedio. As , si hay un numero mayor de nodos, estos seran mas cercanos en forma, y por lo tanto, mejorar el calculo de su aproximacion.

Se concluye que, es posible aproximar cualquier silueta verme a traves de este metodo, usando un numero de centroides representativo con 10 nodos para engendrar los subdo-minios de forma.

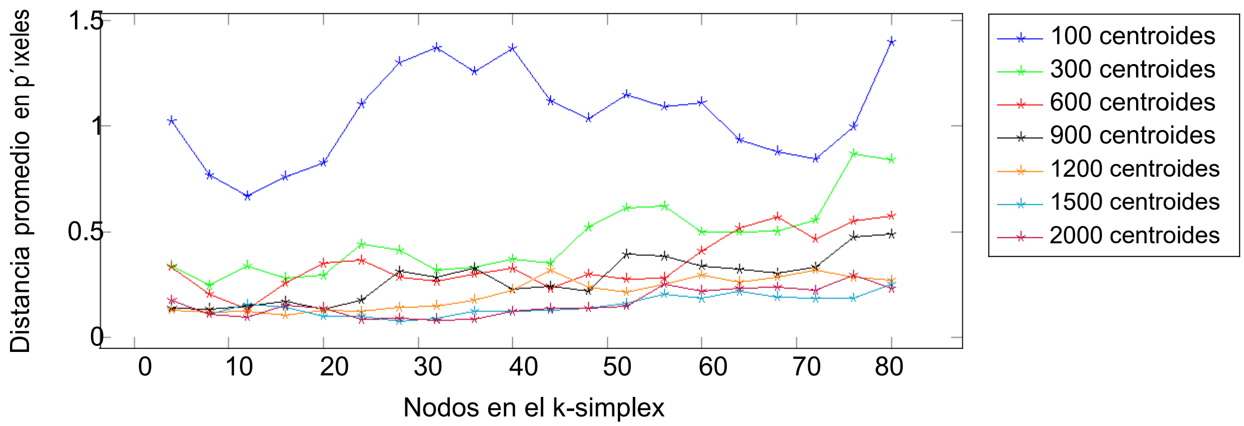


Figura 4.14: Distancia promedio por hito entre $f^b{}_c$ y $f^b{}_a$ en un S_{Df} de k nodos.

Por otro lado, aunque ambos metodos: por reduccion de dimensionalidad (ACP y ACP K) y por aproximacion de siluetas vermes en un subdominio de forma, logran reconstruir las formas vermes, este ultimo reconstruye la forma con errores de aproximadamente 1 p xel por hito para al menos 100 centroides y 10 nodos para la generacion de subdominios de forma.

4.2.2. Variabilidad de formas

En la aplicacion del metodo de emulacion y adopcion de nuevas formas vermes se logran conseguir nuevos conjuntos de siluetas de gusano de n individuos cercanos en forma y pertenecientes a subdominios S_{Df} que pueden competir contra otros no dominados por el frente de Pareto en el MEF.

Para obtener un nuevo conjunto de siluetas vermes con dicha caracter sticas, se analiza la generacion de nuevas formas vermes a traves de cambios de la forma entre saltos sucesivos de subdominios S_{Df} para un numero de centroides y nodos dados, con el ndice de cercan a dado en (4.1).

Tanto el numero de centroides, nodos y el intervalo de restriccion I, se obtienen de los resultados de la gura 4.15. En esta se usa N_{iter} = 1000 iteraciones del algoritmo dado en la gura 3.20 para S_{Df} de nidos para k = 5; 10; 20; ; 100 vertices de dimension de la variedad topologica 80 dimensiones y centroides de nidos para un numero total de multiplos de 100 calculados sobre la base de datos de 2744 formas vermes usando el metodo k vecinos mas cercanos.

El resultado con rma la correlacion existente entre el numero de vertices que conforman los S_{Df} y los cambios de forma, es decir, a mayor numero de nodos las deformaciones son mayores.

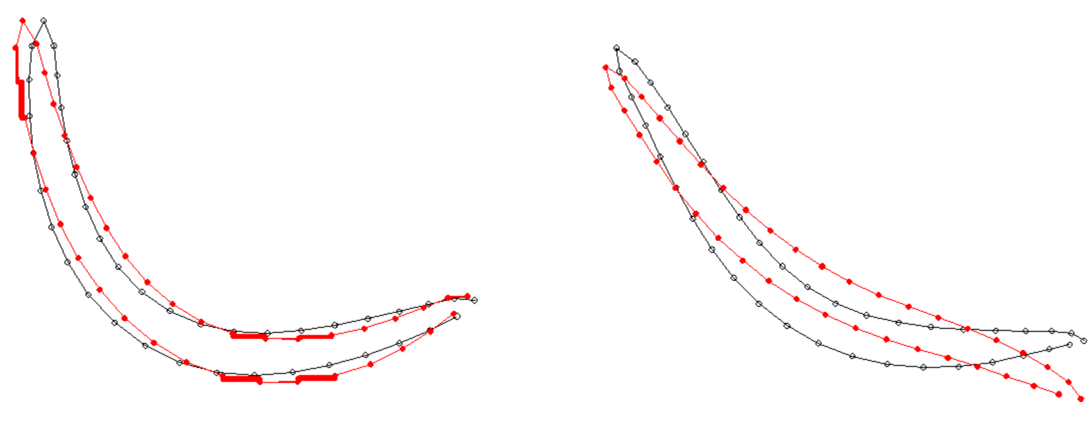
Entre mayor es la cantidad de vertices que pueda seleccionar el algoritmo en forma alea-toria para construir el siguiente subdominio S_{Df} , entonces la distancia entre los hitos correspondientes a dos deformaciones sucesivas es en promedio mayor.

C_j \ Nodos	5	10	20	30	40	50	60	70	80	90	100
200	9.234	10.912	12.290	14.024	13.299	14.149	15.740	14.748	16.111	17.904	16.833
300	9.772	11.263	12.270	12.142	13.477	15.999	16.252	16.175	17.001	17.576	18.475
400	8.734	10.261	11.417	12.571	13.813	13.415	15.108	15.911	16.614	17.124	17.395
500	8.870	10.999	11.530	11.955	12.223	13.122	14.145	14.925	16.685	18.693	18.583
600	7.647	9.9830	11.314	12.416	13.591	14.164	14.560	15.969	16.823	19.612	20.405
700	7.683	9.5160	11.193	12.497	12.639	13.226	12.569	14.097	15.279	16.482	19.400
800	6.232	9.0780	10.791	12.318	12.559	13.558	13.182	13.921	14.517	14.568	17.801
900	6.283	9.3140	11.217	12.206	13.095	14.642	14.992	15.347	15.927	16.359	19.029
1000	6.483	9.2060	10.794	12.207	13.324	14.218	15.074	15.902	16.522	16.791	18.385

Figura 4.15: Distancias promedio de hitos respectivos entre f^b_i y f^b_{i+1} de una serie simulada de movimientos vermes segun los C_j y nodos del S_{Df} , usando 1000 iteraciones.

Aunado a esto, si el numero de vertices del S_{Df} es 5, el desplazamiento de los hitos es menor a 8 p xeles en promedio, lo que exhibe baja variabilidad promedio de cambios independientemente del numero de centroides.

Si se elige un numero mayor a 40 nodos para formar los S_{Df} , las deformaciones seran abruptas y se perdera la nocion de continuidad de la simulacion, pues se puede llegar a un desplazamiento aproximado de 13 o mas p xeles por hito, independientes del numero de centroides.



(a) Distancia de 7.3 p xeles promedio entre hitos respectivos usando 600 centroides y 5 nodos del s mplex.

(b) Distancia promedio de 17.8 p xeles en pro-medio entre hitos respectivos usando 500 cen-troides y 100 nodos del s mplex.

Figura 4.16: En la imagen se presenta la silueta de dos formas de gusano de dos iteraciones se-guidas extra das de una sucesion simulada de formas vermes. Se mide la distancia entre los hitos respectivos.

Un conjunto de deformaciones recorridas en secuencia y que permitan la simulacion en tiempo real oscila entre 10 y 30 nodos por S_{Df} cuya distancia entre hitos sera aproxi-madamente de 10 a 12 p xeles en promedio, por lo que l = [10;12].

4.3. Ajuste de los parametros del MEF

A partir de la forma inicial $\hat{f}_0$, MEF aproxima las siluetas pertenecientes al conjunto optimo de Pareto a la forma verme deseada en el D_f . Dicha convergencia es determinada por dos variables asociadas a la constitucion de los subdominios de formas:

- Numero de centroides del dominio de formas D_f .
- Numero de nodos que determinan cada subdominio S_{Df} .

Ambas variables en conjunto, determinan la variabilidad de las formas en los subdominios, y por ende, la generacion de nuevas formas vermes dadas por el modelo. Las secciones 4.3.1 y 4.3.2 muestran el efecto de la distancia entre los vertices que determinan cada subdominio, y el numero de nodos que los generan.

Primeramente, se analiza cada una de las variables dejando ja la otra variable, para determinar el efecto aislado de cada una en el modelo. En la seccion 4.3.3 se explora de manera conjunta el efecto de ambas variables.

En todos los experimentos de esta seccion, se utilizo la tecnica de validacion cruzada para k = 10 particiones (seccion 2.9.1).

4.3.1. Numero de particiones sobre el dominio de forma D_f

En este experimento, se evalua el efecto en el MEF cuando se aumenta o disminuye la cantidad de centroides calculados sobre el dominio D_f , con los cuales se obtienen los nodos de los n-s mplexes.

Se desea determinar un numero adecuado de centroides optimo sobre el dominio de formas vermes para ser aplicados en los metodos de truncamiento (ver seccion 3.5) y serpenteo (ver seccion 3.4). En el MEF, se experimenta con un numero jo de 10 nodos por subdo-minio de forma S_{Df} y para 50, 100, 150, 250 y 300 centroides que engendran el dominio de forma.

Se tom el promedio de las distancias de la medida de ajuste a la forma verme entre los hitos calculados y anotados. Los resultados se muestran en la gura 4.17.

Los resultados muestran que a partir de 8 iteraciones del MEF, el modelo converge a un promedio entre 4 y 6 p xeles aproximadamente de la silueta calculada respecto a la que pertenece la imagen digital.

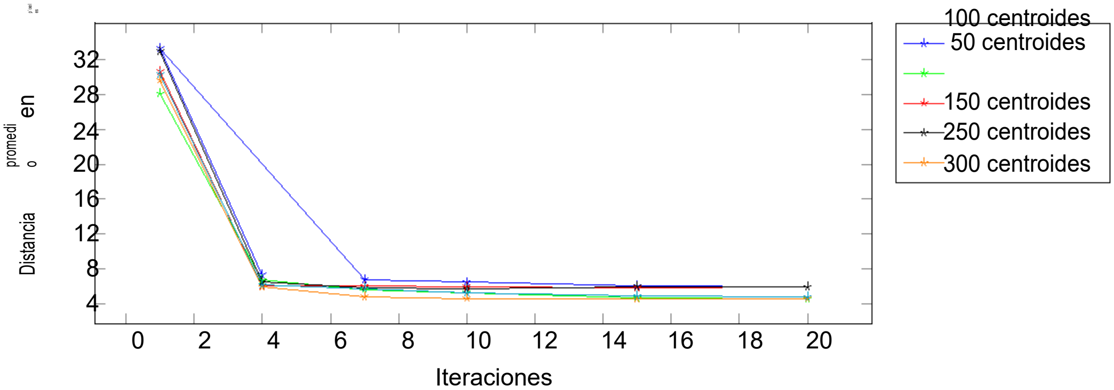


Figura 4.17: Iteraciones del MEF por numero de centroides para engendrar subdominios de n = 10 nodos en cada subdominio de forma. La distancia promedio por hito es menor que 8 a partir de 10 iteraciones.

Debido a lo anterior, el algoritmo muestra que para 10 iteraciones del MEF y 100 centroi-des, el algoritmo provee una distancia promedio entre los hitos dados y calculados dentro del rango posicionamiento valido.

4.3.2. Dimension de los subdominios de forma S_{Df}

En este apartado se evalua el numero optimo de nodos que conforman los subdominios de forma S_{Df} para ser aplicados en el MEF.

Para este experimento se utiliza un numero jo de 100 centroides y se analiza el compor-tamiento para subdominios conformados por 5, 10, 15, 20 y 25 nodos. Ademas, se usa la medida de ajuste a la forma verme dada en la seccion 3.8 entre las distancias entre los hi-tos de las siluetas de formas vermes calculadas por el MEF y las anotadas respectivamente de manera manual. Asimismo, se calcula la ra z del error medio cuadratico (REMC) de las medidas obtenidas para este analisis. Los resultados se muestran en la gura 4.18 con un numero jo de 10 iteraciones del MEF.

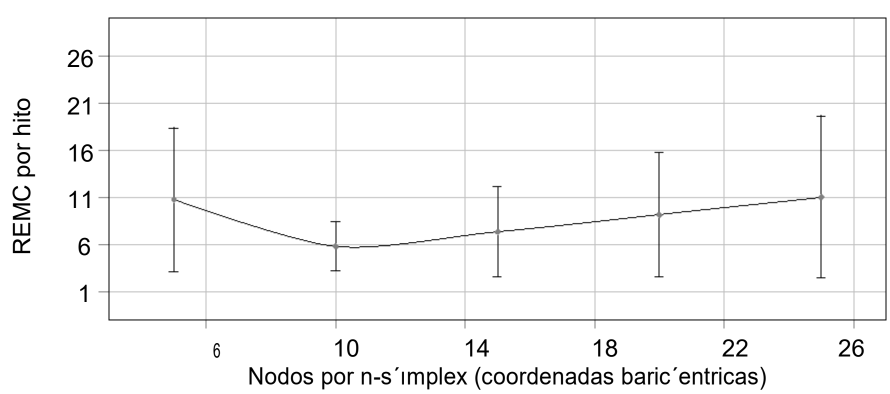


Figura 4.18: Numero de nodos de cada subdominio de forma.

Los resultados muestran que para subdominios conformados por 10 nodos disminuye el promedio hasta alcanzar el m nimo de las distancias por hito entre la silueta calculada y la curva de representacion.

Ademas, estos mismos reejan que si se disminuye o aumenta el numero de coordenadas baricentricas que conforman cada subdominio S_{Df} aumenta dicho promedio, y por lo tanto, el ajuste a la forma verme empeora.

Lo anterior se explica debido a que, al utilizar un numero menor de nodos, estos son mas cercanos en forma entre s dos a dos, con lo cual en el espacio de conformacion de nuevas formas vermes se generan siluetas de mayor similitud, o bien, el subdominio tiende a disminuir el rango de variabilidad, aspecto por el cual, los metodos de truncamiento y serpenteo trabajaran con espacios de menor variabilidad y no logran alcanzar el objetivo de ajuste.

En caso contrario, si la cantidad de nodos de cada subdominio de forma es un numero mayor a 10, entonces aumenta el promedio de las distancias entre las formas generadas en dichos subdominios. Por consiguiente, los subdominios estan conformados por los nodos mas lejanos entre s , aumentando la variabilidad en la generacion de formas vermes.

Una consecuencia directa de estos resultados en el MEF, es que los metodos del serpenteo y truncamiento trabajaran en espacios de mayor variacion, lo que desemboca en una mayor cantidad de iteraciones de busqueda y ajuste para lograr el objetivo nal.

En ambos casos con 10 iteraciones del MEF y con un numero mayor o menor a 10 coor-denadas baricentricas no se alcanza el objetivo de ajuste.

4.3.3. Vertices y particiones sobre el dominio de forma

Para analizar el efecto en conjunto sobre el numero de centroides del dominio de formas y numero de nodos que determinan cada subdominio, respecto a los resultados obtenidos en las secciones 4.3.2 y 4.3.1, se ejecuta MEF para medir el ajuste a la forma verme de los nematodos en las imagenes digitales.

Se mide la distancia entre la curva de representacion verme y los hitos calculados usando 10 iteraciones y el metodo validacion cruzada. Los resultados se exhiben la gura 4.19.

Los resultados con rman que la mejor asignacion es 10 coordenadas baricentricas para conformar los subdominios de forma S_{Df} y una particion de 100 centroides, pues se logra que en promedio haya una distancia de ajuste sobre la curva de representacion verme de aproximadamente 6 p xeles por hito que esta dentro del rango de ajuste optimo.

4.4. Medicion del ajuste de la forma verme

La medicion del ajuste nal de la forma verme al nematodo en la imagen digital se realizo mediante el ndice de Jaccard (seccion 2.9.2). Este coe ciente permite determinar una

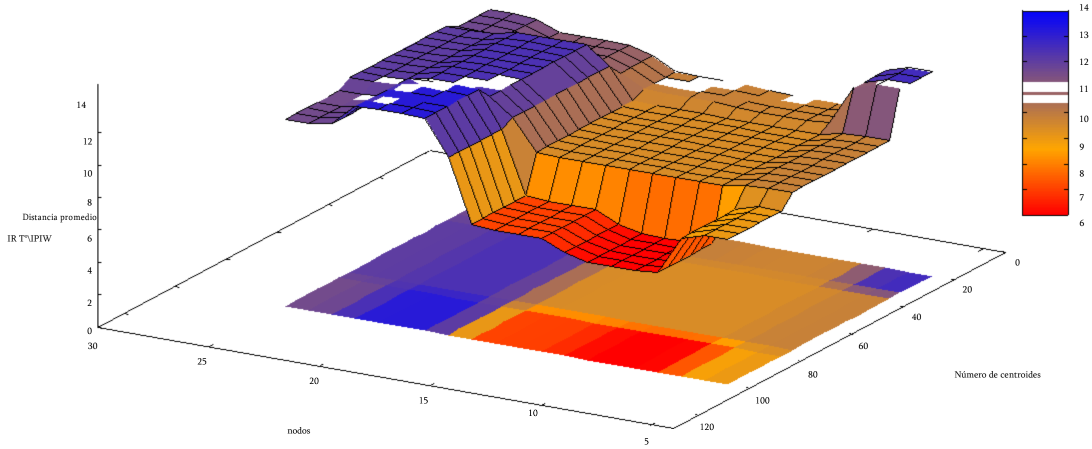


Figura 4.19: Resultado de 10 iteraciones, variando el numero de coordenadas baricentricas que determinan los subdominios de forma y el numero de centroides de particion del dominio de forma. Se midio la distancia entre la curva de representacion y los hitos calculados por el MEF.

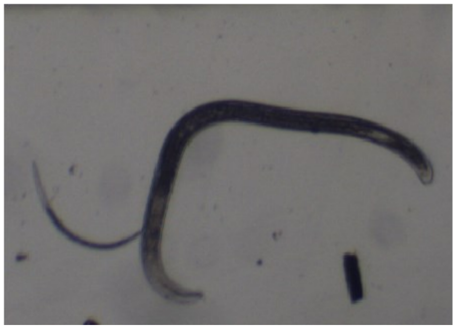
metrica de acierto sobre ajuste de las siluetas $\hat{f}_i$ respecto a las referencias dadas. Se denota $J(A_a;A_c)$ y se calcula como el cociente de las cardinalidades de los conjuntos de las areas de interseccion y union de A_a y A_c, dada por:

$$J(A_a;A_c) := \frac{card(A_a \cap A_c)}{card(A_a \cup A_c)} \tag{4.9}$$

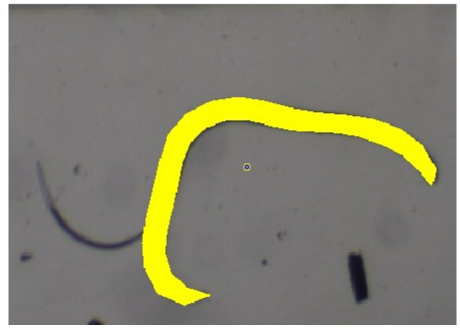
Para realizar dicha medicion, se tom como region verdadera y calculada las areas de las siluetas de las formas vermes delimitadas por los hitos anotados A_a (gura 4.20b), y la de los hitos calculados A_c (gura 4.20c). Se resalta que el area de cada region corresponde a la cantidad de p xeles que contiene. Los resultados obtenidos se promediaron y clasi caron segun el coe ciente $J(A_a;A_c)$ en clases segun la tabla 4.1.

Asimismo, los resultados obtenidos en las secciones 4.3 y 4.1 fueron aplicados: siluetas formadas por 40 hitos de frontera, particion de 100 centroides del $\hat{C}$ para engendrar los subdominios S_{Df}, 10 vertices o nodos para generar S_{Df} y $N_{iter} = 10$ iteraciones del algoritmo del MEF.

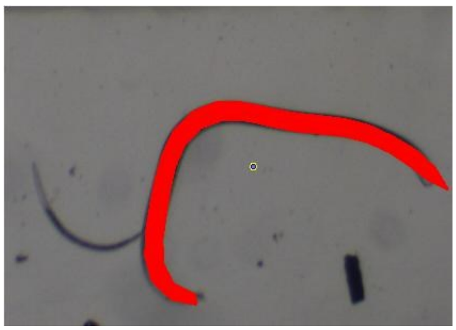
Tabla 4.1: Tabla de particiones para ndice de Jaccard $J(A_a;A_c)$.



(a) Imagen de un nematodo.



(b) Area del nematodo delimitada por hitos anotados A_a.



(c) Area del nematodo delimitada por hitos calculados A_c.

(d) Interseccion entre las areas.

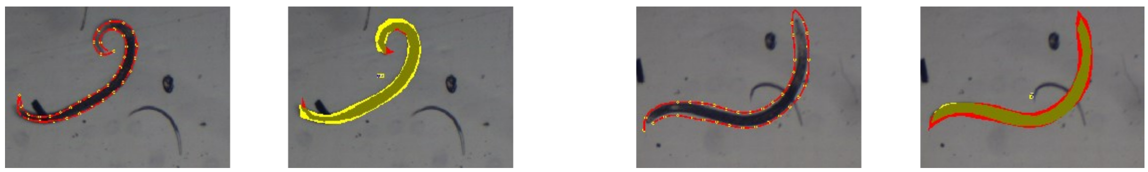
Figura 4.20: Regiones verdadera, calculada y su interseccion segun el resultado nal de la aplicacion del MEF utilizadas para el calculo del J $(A_a;A_c)$.

Clases	J $(A_a;A_c)$
a_1	J <0:2
a_2	0:2 J < 0:4
a_3	0:4 J < 0:6
a_4	0:6 J < 0:8
a_5	0:8 J

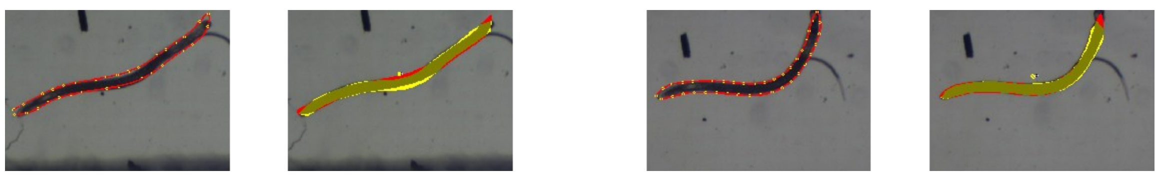
En la gura 4.21 se ilustran los resultados con imagenes de las areas resaltadas y sus respectivos ndices de Jaccard. Se observa tambien que a partir de de la clase a_3, el MEF ajusta de forma verme similar a la silueta de nematodo a la que se encuentra en la imagen digital.

Los resultados obtenidos en este experimento se muestran en el histograma normalizado de la gura 4.22. Las muestras clasi cadas en las clases a_4 y a_5 muestran un ajuste aceptable respecto a la silueta del nematodo, aunque de igual forma es afectada por el posicionamiento de los hitos anotados. No obstante, las formas vermes contenidas en la clase a_2 en su mayor a recuperan parte de la forma del nematodo que son afectadas por detritos presentes en la imagen digital, aunque parcialmente se recupera su forma. Las formas vermes pertenecientes a la c_1 son fracasos del algoritmo (gura 4.23).

(a) $J(A_a;A_c)$ 0:48 (b) $J(A_a;A_c)$ 0:58



(c) $J(A_a;A_c)$ 0:65 (d) $J(A_a;A_c)$ 0:71



(e) $J(A_a;A_c)$ 0:77 (f) $J(A_a;A_c)$ 0:81

Figura 4.21: Ejemplos de la aplicacion del MEF con su respectivo ndice de Jaccard. En cada caso, la imagen digital de la izquierda muestra el resultado nal del MEF y, en la de la derecha se resalta en amarillo el area de la forma verme de hitos anotados as como con rojo el area de la forma obtenida por el modelo propuesto. En el color verde musgo se muestran las intersecciones entre las areas resaltadas por rojo y amarillo.

Cabe destacar que los hitos anotados que forman las siluetas vermes de la base de datos de entrenamiento se ubican en el rango de posicionamiento valido de nido en la seccion 4.1 el cual indica que puede existir hasta 6 p xeles de diferencia entre posiciones dadas por el algoritmo MEF y la manual. Ademas, existe un error de posicionamiento asociado al error humano.

El modelo elige otras posiciones de los hitos af n en ubicacion pero no coincidente con la de la base de datos de entrenamiento. Debido a esto, los resultados son afectados por dicha diferencia a nivel de contorno. En la sub guras 4.21a y 4.21b se observan estos errores, aunque los resultados nales coinciden con la silueta del nematodo en la imagen digital.

Finalmente, del modelo evolutivo de forma constituido por el M OP SO y los metodos del forrajeo y serpenteo logran un porcentaje de exito para el ndice de Jaccard sobre las clases a_3, a_4 y a_5 del 90.56 %. Las instancias que parcialmente capturan su forma asciende

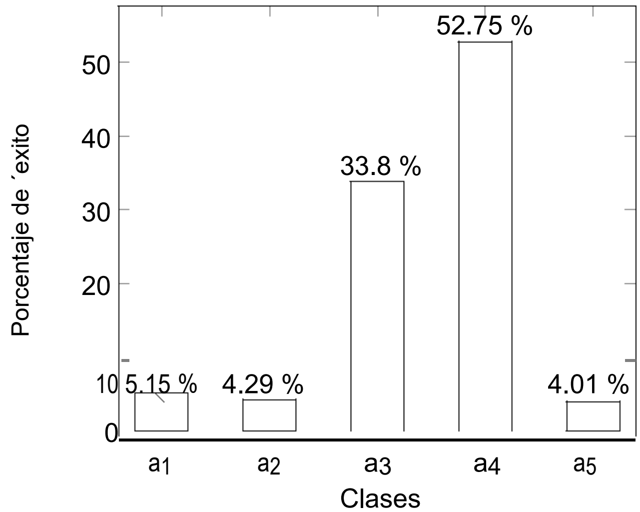
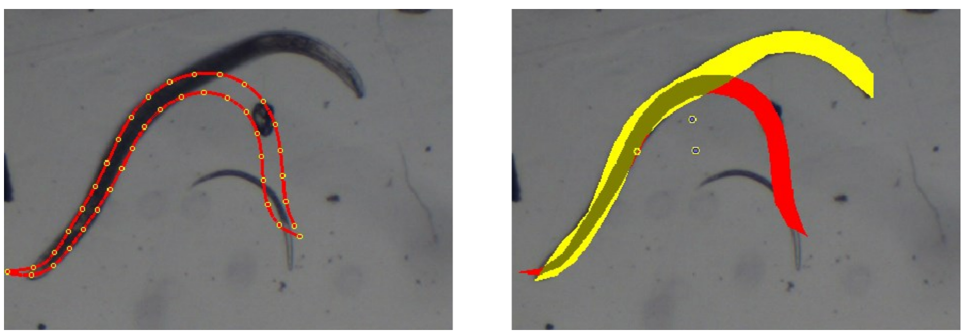


Figura 4.22: Resultado del aplicacion del MEF por clases segun la tabla 4.1.



(a) J (Aa;Ac) 0:28

Figura 4.23: Imagenes del resultado nal del MEF de la clase a_2.

a un 4.29 % del total. El algoritmo no logra su cometido en un 5.15 % de las instancias vermes utilizadas.

Aunque los tiempos de respuesta no se evaluaron, como referencia en promedio un ajus-te de forma del MEF con 5 iteraciones dura aproximadamente 72.2 segundos con una desviacion estandar de 6.83.

Cap tulo 5

Conclusiones

Este trabajo propone un metodolog a de ajuste de formas vermes expresadas en vectores de hitos en imagenes digitales. Consiste en un proceso de optimizacion multiobjetivo apoyado de dos procesos intermedios de deformacion: serpenteo y forrajeo. El serpenteo emula formas vermes usando caminatas aleatorias entre subdominios de forma validos y el forrajeo contorsiona los fragmentos cola y cabeza de las siluetas de los nematodos. Estos procesos permiten un aceleramiento de busqueda de las part culas del frente de Pareto que son determinados por dos funciones objetivo: una ajuste de forma y otra de tama~no de la silueta. Ambas normalizadas con una funcion sigmoide. Finalmente, se utiliza los metodos de validacion cruzada e ndice de Jaccard para la evaluacion del metodo propuesto.

La principal contribucion de este trabajo es la propuesta de una estrategia de ajuste de siluetas vermes basado en un proceso iterativo de optimizacion que permite hallar, en el dominio de formas permitidas vermes, la mejor adaptacion de la silueta del nematodo en las imagenes digitales.

Asimismo, la estrategia utilizada se fundamento en dos criterios de adaptacion de las formas. El primer criterio de mejor adaptacion es referido a la obtencion de la forma a traves de un metodo adaptativo de movimientos corporales vermes. Este fue logrado mediante el metodo de optimizacion M OP SO asistido por dos subfases: el metodo del serpenteo y el metodo del forrajeo.

El segundo criterio es dado el truncamiento de las formas vermes resultantes a dominios de forma permitidos. El modelo de optimizacion se encarga de ajustar en forma y en posicion las part culas de forma de gusano. La decision es guiada por las funciones de bordicidad y longitud verme que evaluan y determinan la mejor posicion y postura de cada una de las part culas del frente de Pareto.

Asimismo, el metodo del forrajeo interviene despues de cada iteracion del M OP SO como ayudante en el ajuste de forma a las part culas. Los trozos cabeza y cola son tratados en este metodo individualmente permitiendo un mejor ajuste. Las elongaciones y deforma-ciones laterales son evaluadas tambien por la funcion de bordicidad. De forma paralela, el metodo de truncamiento no permite la perdida de la vermiformidad en el proceso evolutivo de ajuste de forma.

En este mismo criterio, se enfatiza que la descripcion y deformacion de las siluetas debe

ser a traves de un conjunto limitado de parametros de forma. Aqu el objetivo es cumplido, primeramente por la estrategia de representacion y luego por el metodo de emulacion de nuevas formas vermes.

Cada nematodo es representado e cientemente por un conjunto reducido de 40 hitos de frontera que, en conjunto, moldean su silueta y logran extraer toda la informacion a nivel de contorno requerida para el modelo.
Asimismo se comprueba que, a traves de la seleccion de caminatas aleatorias entre S_{Df}

vecinos y un numero adecuado de coordenadas baricentricas dado por el metodo del serpenteo, se logr emular cualquier secuencia de movimientos vermiformes de forma aleatoria, permitiendo y generando representaciones de nuevas formas vermes.

De igual manera, este metodo permite generar, e inclusive, describir de forma parametrica formas vermes permitidas en S_{Df} no determinados por centroides de la base de datos.

Otra contribucion de este trabajo es el metodo del serpenteo. Este permite la generacion de movimientos vermes simulados y continuos a traves de secuencias de siluetas de nematodos creados arti cialmente por medio de caminatas aleatorias entre subdominios de forma.

El tercer criterio de mejor adaptacion indica que el modelo sea capaz de ajustarse a la informacion presente en la imagen digital. Bajo las restricciones dadas, y por la metodo-log a empleada en el MEF, se logr que la inuencia de todo el ruido presente no fuera determinante en el modelo. A pesar de que en este trabajo, las imagenes digitales con las que se experimento s contienen escenas ruidosas, llenos de arti cios e incluso presencia de
multiples individuos vermes en algunos casos, se logra ajustar la silueta al nematodo en
imagenes digitales. Aunado a esto, la unica operacion de imagenes empleada es el calculo

del gradiente por medio del metodo OGD que determina la razon de cambio entre el fon-do de la imagen digital y el cuerpo del nematodo. La evaluacion de la medida optima de borde-longitud permite al M OP SO el posicionamiento y adopcion de la silueta de ajuste nal de la forma de nematodo.

Es preciso destacar que no se utilizo ninguna operacion morfologica de imagenes digita-les ni operaciones asociadas a los canales de colorimetr a. Esto permite que el modelo sea adaptable a casi cualquier tipo de imagen digital que cumpla con las restricciones mencionadas.

Finalmente, el cuarto criterio se re ere a que el algoritmo sea capaz de discernir si la silueta ajustada a la imagen digital corresponde efectivamente a un nematodo. Para esto, se evaluaron todas las partes y procedimientos asociados al algoritmo MEF dando certeza del cumplimiento de este en todos los objetivos planteados inicialmente.

Ademas, esta etapa permitio ajustar los criterios sobre los valores parametrizables, tales
como el numero de coordenadas baricentricas que determinan los subdominios de forma,
numero de centroides, numero de coordenadas que representan cada deformacion verme.

Por ultimo, la evaluacion realizada sobre los conjuntos de entrenamiento y validacion de 667 imagenes digitales determinan un 90.56 % de exito en el ajuste de la forma, asegu-rando que el desarrollado es robusto y ajustable a cualquier silueta verme en una imagen digital segun las restricciones a lo largo de esta investigacion. Lo anterior, es sujeto a la distribucion de formas que se describieron en la seccion 3.2.

Como resultado nal del MEF, se favorece la compresion de los datos permitiendo describir de forma parametrica un nematodo a traves de sus hitos de frontera. Ademas, este modelo logra eliminar la redundancia del conjunto de datos y permite mejorar procesos de visualizacion de los datos a un menor costo computacional.

Finalmente, la evaluacion del algoritmo evidencia de manera objetiva, que la solucion propuesta es satisfactoria ante el problema en cuestion, y con esto, se logra demostrar la hipotesis de investigacion as como el cumplimiento a cabalidad de los objetivos principal y secundarios planteados.

Trabajo futuro

A futuro, se recomienda continuar el trabajo en otro tipo de deformaciones ajustables para clasi cacion y posicionamiento en v deo digital, y con imagenes digitales con diferentes tipos de resoluciones. Asimismo, queda como trabajo venidero, modi car el metodo con otros procesos de optimizacion multiobjetivo con el n de comparar cual se ajusta mejor a la metodolog a empleada en esta investigacion.

Ademas, por la forma de trabajo empleada en el modelo evolutivo de forma, este puede ser replicado a otros tipos de elementos estructurales en imagenes digitales cuyas deformaciones no sigan distribuciones normales con el n de robustecer y ampliar el espectro de este metodo. Ademas, aplicarle el metodo con otras distribuciones de longitudes de nematodos diferentes a las que se realizaron en este trabajo.

Aunque el objetivo principal de este trabajo es la elaboracion del modelo de forma, ser a importante replicar el MEF con otros conjuntos de imagenes digitales que se incluyan diferentes resoluciones de las imagenes digitales de nematodos con el n ejempli car y profundizar el analisis de los resultados.

Por otra parte, el metodo de trabajo de simulacion de movimientos vermiformes mediante traslado aleatorio entre S_{Df} vecinos funciona como generador de formas, por lo que puede ser replicado, integrado y hasta extendido a distintos escenarios y modelos aplicados a la segmentacion y clasi cacion de cuerpos que son alterados por transformaciones continuas en espacios topologicos.

Bibliograf a

Abbass, H. A., Sarker, R. & Newton, C. (2001). PDE: a Pareto-frontier di erential evo-lution approach for multi-objective optimization problems, En *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*. IEEE.

Albrecht, T., Luthi, M. & Vetter, T. (2009). Deformable Models. En *Encyclopedia of Biometrics* (pp. 210-215). Boston, MA, Springer US. https://doi.org/10.1007/978-0-387-73003-5 88

Alvarado, P., Doerer, P. & Wickel, J. (2001). Axon2{a visual object recognition system for non-rigid objects, En *Proceedings International Conference on Signal Proces-sing, Pattern Recognition and Applications (SPPRA)*.

Arroyo, J. (2016). Dimensionality Reduction Methods: Comparative Analysis of methods PCA, PPCA and KPCA. *Uniciencia*, *30* (1), 115-122. https://doi.org/10.15359/ ru.30-1.7

Arroyo, J. & Alvarado, J. (2020). Simulacion de movimientos vermiformes mediante ca-minatas aleatorias entre n-s mplex vecinos. *Revista Tecnolog a En Marcha*, *33* (2), 54-66.

Arroyo, J. & Alvarado, J. (2014). A new variant of Conformal Map Approach method for computing the preimage in Input Space. *Recent Advances in Computer Enginee-ring, Communications and Information Technology*, 301-304.

Association, A. W. W. Et al. (2004). Problem Organisms in Water: Identi cation and Treatment{Manual of Water Supply Practices M7. American Water Works Asso-ciation, Denver.

Atkinson, K. E. (2008). *An introduction to numerical analysis*. John Wiley & Sons.

Berman, M., Rannen Triki, A. & Blaschko, M. B. (2018). The lovasz-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks, En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Bernard, G. C., Egnin, M. & Bonsi, C. (2017). The Impact of Plant-Parasitic Nematodes on Agriculture and Methods of Control. En Nematology-Concepts, Diagnosis and Control. Intechopen.

Berry, M. W., Mohamed, A. & Yap, B. W. (2019). Supervised and Unsupervised Learning for Data Science. Springer.

Bogale, M., Baniya, A. & DiGennaro, P. (2020). Nematode Identi cation Techniques and Recent Advances. Plants, 9 (10), 1260.

Bongers, T. & Ferris, H. (1999). Nematode community structure as a bioindicator in environmental monitoring. Trends in Ecology & Evolution, 14 (6), 224-228.

Boyd, S. & Vandenberghe, L. (2004). Convex optimization. Cambridge University Press.

Brown, S., Pedley, K. C. & Simcock, D. C. (2016). Estimation of surface area and volume of a nematode from morphometric data. Scienti ca, 2016.

Carr, H., Snoeyink, J. & van de Panne, M. (2010). Flexible isosurfaces: Simplifying and displaying scalar topology using the contour tree. Computational Geometry, 43 (1), 42-58.

Chen, M., Correa, C., Islam, S., Jones, M. W., Shen, P.-Y., Silver, D., Walton, S. J. & Willis, P. J. (2007). Manipulating, deforming and animating sampled object representations, En Computer Graphics Forum. Wiley Online Library.

Chew, L. P. (1989). Constrained delaunay triangulations. Algorithmica, 4 (1-4), 97-108.

Coello, C. A. C., Pulido, G. T. & Lechuga, M. S. (2004). Handling multiple objectives with particle swarm optimization. IEEE Transactions on Evolutionary Computation, 8 (3), 256-279.

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. (1995). Active Shape Models - Their Training and Application. Computer vision and image understanding, 61 (1), 38-59.

Cootes, T. F. & Taylor, C. J. (1993). Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation., En BMVC. Citeseer.

Costa, C., Antonucci, F., Pallottino, F., Aguzzi, J., Sun, D.-W. & Menesatti, P. (2011). Shape analysis of agricultural products: a review of recent research advances and potential application to computer vision. Food and Bioprocess Technology, 4 (5), 673-692.

Coto, G. R. (2007). Conceptos introductorios a la topatolog a. Costa Rica, EUNED.

Currie, M., Cantarelli, M., Idili, G. & Larson, S. (2014). OpenWorm: an open-science approach to modeling Caenorhabditis elegans. Frontiers in Computational Neuros-cience, 8 (137), 2194-2232.

Cury, C. (2015). Statistical shape analysis of the anatomical variability of the human hip-pocampus in large populations. (Tesis doctoral). Universit Pierre et Marie Curie-Paris VI.

Delingette, H. (1999). General object reconstruction based on simplex meshes. Interna-tional Journal of Computer Vision, 32 (2), 111-146.

Donoho, D. L., Tsaig, Y., Drori, I. & Starck, J.-L. (2012). Sparse solution of underde-termined systems of linear equations by stagewise orthogonal matching pursuit. IEEE Transactions on Information Theory, 58 (2), 1094-1121.

Dorer, P. & Alvarado, P. (2006). LTI-Lib { a C++ Open Source Computer Vision Library. En K.-F. Kraiss (Ed.), Advanced Man-Machine Interaction. Fundamentals and implementation (pp. 399-421). Berlin Heidelberg, Springer Verlag.

Duran-Mora, J. (2018). Estudio poblacional y alternativas de control para el nematodo nodulador en tomate (solanum lycopersicum l.) en la zona occidental de Costa Rica (Tesis para optar el grado de Doctor en Ciencias Naturales para el Desarrollo, con enfasis en Sistemas de Produccion Agr cola). Doctorado en Ciencias Naturales para el Desarrollo. https://repositoriotec.tec.ac.cr/handle/2238/10656.

Eberhart, R. & Kennedy, J. (1995). A new optimizer using particle swarm theory, En MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. IEEE.

El as, I. (2014). La estrategia competitiva del sector agrario a traves de la innovacion y desarrollo. Cuadernos de Investigacion EPG.

Esquivel, A. (2011). Nematodos como indicadores ambientales. Costa Rica: Universi-dad Nacional. https : / / repositorio . una . ac . cr / bitstream / handle / 11056 / 7460 / Nematodos%20como%20indicadores%20ambientales.pdf?sequence=1

Esquivel, A. & Peraza, W. (2009). Nematodos toparasitos asociados a cultivos agr colas en Costa Rica. Escuela de Ciencias Agrarias, Universidad Nacional.

Farin, G. (1983). Algorithms for rational Bezier curves. Computer-aided design, 15 (2), 73-77.

Foley, T. A. & Nielson, G. M. (1989). Knot selection for parametric spline interpolation. En Mathematical methods in computer aided geometric design (261-CP4). Elsevier.

Friedman, J., Hastie, T. & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer series in statistics New York.

Gallagher, T., Bjorness, T., Greene, R., You, Y.-J. & Avery, L. (2013). The geometry of locomotive behavioral states in C. elegans. PloS one, 8 (3), e59865.

Garza-Hume, C., Jorge, M. C. & Olvera, A. (2018). Areas and Shapes of Planar Irregular Polygons, En Forum Geometricorum.

Geng, W., Cosman, P., Berry, C., Feng, Z. & Schafer, W. (2004). Automatic tracking, fea-ture extraction and classi cation of C. elegans phenotypes. IEEE Trans. Biomed. Eng., 51, 1811-1820.

Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Gomez, O. (2009). Segmentacion de nematodos en imagenes digitales usando redes neuro-nales (Tesis para optar al grado de Magister Scientiae en Computacion). Instituto Tecnologico de Costa Rica. Costa Rica.

Gonzalez, R. & Woods, R. (2008). Digital Image Processing (Third Edition). Pearson Pretince Hall.

Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning [http://www.deeplearningbook.org]. MIT Press.

Gruner, M. (2015). Active Dictionary Models: A framework for non-linear shape modeling (Tesis para optar al grado de Magister Scientiae en Computacion). Escuela de Ingenier a Electronica. Instituto Tecnologico de Costa Rica. Cartago, Costa Rica.

Harary, F. & Yan, H. (1990). Logical foundations of kinematic chains: Graphs, line graphs, and hypergraphs. Journal of Mechanical Design - Transactions of the ASME, 112, 79-83.

Hernandez, R. & Torres, C. P. M. (2018). Metodolog a de la investigacion (Vol. 4). McGraw-Hill Interamericana.

Hidalgo, M. G., Torres, A. M. & Gomez, J. V. (2012). Deformation Models: Tracking, Animation and Applications (Vol. 7). Springer Science & Business Media.

Holden, D., Saito, J. & Komura, T. (2016). A deep learning framework for character motion synthesis and editing. ACM Transactions on Graphics (TOG), 35 (4), 1-11.

Honeine, P. & Richard, C. (2011). Preimage Problem in Kernel-Based Machine Learning. IEEE Signal Process, 77-88.

Huang, K.-M., Cosman, P. & Schafer, W. (2008). Automated Detection and Analysis of Foraging Behavior in C. elegans. Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on.

Huang, Q.-X., Wicke, M., Adams, B. & Guibas, L. (2009). Shape decomposition using modal analysis, En Computer Graphics Forum. Wiley Online Library.

Jahne, B. (2005). Digital Image Processing 6th Edition. Berlin [u.a.], Springer.

Jimenez, J. (2019). SegNema: Nematode segmentation strategy in digital microscopy ima-ges using deep learning and shape models (Tesis para optar al grado de Magister Scientiae en Computacion). Escuela de Ingenier a en Computacion. Tecnologico de Costa Rica. Cartago, Costa Rica.

Juhasz, Z. & Zelei, A. (2013). Analysis of worm-like locomotion. Periodica Polytechnica Mechanical Engineering, 57 (2), 59-64.

Jukna, S. (2011). Extremal combinatorics: with applications in computer science. Springer Science & Business Media.

Kamilaris, A. & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90.

Kenney, E. & Eleftherianos, I. (2016). Entomopathogenic and plant pathogenic nematodes as opposing forces in agriculture. International Journal for Parasitology, 46 (1), 13-19. https://doi.org/10.1016/j.ijpara.2015.09.005

Lawler, G. F. & Limic, V. (2010). Random walk: a modern introduction (Vol. 123). Cambridge University Press.

Lee, H., Grosse, R., Ranganath, R. & Ng, A. Y. (2009). Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, En Proceedings of the 26th annual international conference on machine learning.

Lee, J. A. & Verleysen, M. (2007). Nonlinear dimensionality reduction. Springer Science & Business Media.

Levinson, N. & Redhe er, R. M. (1975). Curso de variable compleja. Revert .

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18 (8), 2674.

Lindenbaum, O., Stanley, J., Wolf, G. & Krishnaswamy, S. (2018). Geometry Based Data Generation, En Advances in Neural Information Processing Systems.

Litany, O., Bronstein, A., Bronstein, M. & Makadia, A. (2018). Deformable shape comple-tion with graph convolutional autoencoders, En Proceedings of the IEEE conference on computer vision and pattern recognition.

Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C. & Wang, J. (2018). Applications of deep learning to MRI images: A survey. Big Data Mining and Analytics, 1 (1), 1-18.

Livingstone, D. J. (2008). Arti cial neural networks: methods and applications. Springer.

Loop, C., Schaefer, S., Ni, T. & Casta~no, I. (2009). Approximating subdivision surfaces with gregory patches for hardware tessellation. En ACM SIGGRAPH Asia 2009 papers (pp. 1-9).

Mar n, J. (2009). Ubicacion de un Nematodo en Imagenes Digitales Utilizando el Algorit-mo modelos activos de forma (Tesis para optar al grado de Licenciatura). Instituto Tecnologico de Costa Rica. Costa Rica.

Marler, R. T. & Arora, J. S. (2004). Survey of multi-objective optimization methods for engineering. Structural and multidisciplinary optimization, 26 (6), 369-395.

Masci, J., Meier, U., Ciresan, D. & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction, En International conference on arti - cial neural networks. Springer.

Mesejo, P., Iba~nez, O., Cordon, O. & Cagnoni, S. (2016). A survey on image segmentation using metaheuristic-based deformable models: state of the art and critical analysis. Applied Soft Computing, 44, 1-29.

Mora, S. (2019). Informe Comercio Exterior del Sector Agropecuario (2017 { 2018) (inf. tec.). SEPSA. Costa Rica. http://www.sepsa.go.cr/

Muja, M. & Lowe, D. G. (2009). Fast Approximate Nearest Neighbors with Automatic Al-gorithm Con guration, En International Conference on Computer Vision Theory and Application VISSAPP'09), INSTICC Press.

Nalepa, J. & Kawulok, M. (2014). Fast and accurate hand shape classi cation, En Inter-national conference: beyond databases, architectures and structures. Springer.

Nutanong, S., Zhang, R., Tanin, E. & Kulik, L. (2010). Analysis and evaluation of v*-knn: an e cient algorithm for moving knn queries. The VLDB Journal, 19 (3), 307-332.

Piegl, L. & Tiller, W. (1995). The NURBS book (monographs in visual communications)(1997, 646 p). New York: Springer, ISBN.

Ponce, J., Chelberg, D. & Mann, W. B. (1989). Invariant properties of straight homogeneous generalized cylinders and their contours. IEEE Transactions on Pattern Analysis & Machine Intelligence, (9), 951-966.

Pratt, L. & Rivera, L. (2003). Tendencias en el Desarrollo de la Agricultura en Centroamerica: Nuevos Retos para el Sector Privado y el Dise~no de Pol ticas Publicas (inf. tec.). Centro Latinoamericano de Competitividad y Desarrollo Sostenible CLACDS { INCAE.

Pu, Y., Gan, Z., Henao, R., Yuan, X., Li, C., Stevens, A. & Carin, L. (2016). Variational autoencoder for deep learning of images, labels and captions, En Advances in neural information processing systems.

Puckering, T., Thompson, J., Sathyamurthy, S., Sukumar, S., Shapira, T. & Ebert, P. (2017). Automated wormscan. F1000Research, 6.

Qing, F., Yu, X. & Jia, Z. (2019). A robust MOF method applicable to severely deformed polygonal mesh. Journal of Computational Physics, 377, 162-182.

Rahman, M. A. & Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation, En International symposium on visual compu-ting. Springer.

Raiko, T., Valpola, H. & LeCun, Y. (2012). Deep learning made easier by linear transfor-mations in perceptrons, En Arti cial intelligence and statistics.

Real, R. & Vargas, J. M. (1996). The Probabilistic Basis of Jaccard's Index of Similarity. Systematic Biology, 45 (3), 380-385.

Refaeilzadeh, P., Tang, L. & Liu, H. (2009). Cross-Validation. En L. LIU y M. T. OZSU (Eds.), Encyclopedia of Database Systems (pp. 532-538). Boston, MA, Springer US. https://doi.org/10.1007/978-0-387-39940-9 565

Restif, C. & Metaxas, D. (2008). Tracking the swimming motions of C. elegans worms with applications in aging studies, En International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer.

Rizvandi, N., Pizurica, A. & Philips, W. (2008). Machine Vision Detection Of Isolated And Overlapped Nematode Worms Using Skeleton Analysis. Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, 2972-2975.

Rizvandi, N., Pizurica, A., Philips, W. & Ochoa, D. (2008). Edge Linking Based Met-hod to Detect and Separate Individual C. Elegans Worms in Culture. Computing: Techniques and Applications, 2008. DICTA ’08.Digital Image, 65-70.

Sadrnia, H., Rajabipour, A., Jafary, A., Javadi, A. & Mosto , Y. (2007). Classi cation and analysis of fruit shapes in long type watermelon using image processing. Int. J. Agric. Biol, 1 (9), 68-70.

Scholkopf, B., Mika, S., Smola, A., Ratsch, G. & Muller, K.-R. (1998). Kernel PCA pattern reconstruction via approximate pre-images, En International Conference on Arti cial Neural Networks. Springer.

Scholkopf, B., Smola, A. & Muller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10 (5), 1299-1319.

Sedgewick, R. & Wayne, K. (2011). Algorithms. Addison-Wesley Professional.

Sethian, J. (1999). Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science. Cambridge University Press.

Shlens, J. (2009). A Tutorial on Principal Component Analysis. Center for Neural Scien-ce, New York University y Systems Neurobiology Laboratory, Salk Insitute for Biological Studies.

Sonka, M., Hlavac, V. & Boyle, R. (2014). Image processing, analysis, and machine vision. Cengage Learning.

Suprem, A., Mahalik, N. & Kim, K. (2013). A review on application of technology systems, standards and interfaces for agriculture and food sector. Computer Standards & Interfaces, 35 (4), 355-364.

Szeliski, R. & Tonnesen, D. (1992). Surface modeling with oriented particle systems, En Proceedings of the 19th annual conference on Computer graphics and interactive techniques.

Terzopoulos, D. & Metaxas, D. (1991). Dynamic 3D models with local and global de-formations: Deformable superquadrics. IEEE Transactions on Pattern Analysis & Machine Intelligence, (7), 703-714.

Terzopoulos, D., Platt, J., Barr, A. & Fleischer, K. (1987). Elastically Deformable Models. SIGGRAPH Comput. Graph., 21 (4), 205-214. https://doi.org/10.1145/37402. 37427

Van Bezooijen, J. (2006). Methods and techniques for nematology. Wageningen University Wageningen, The Netherlands.

Van Der Maaten, L., Postma, E. & Van den Herik, J. (2009). Dimensionality reduction: a comparative. J Mach Learn Res, 10 (66-71), 13.

Wagsta , K., Cardie, C., Rogers, S., Schrodl, S. Et al. (2001). Constrained k-means clus-tering with background knowledge, En ICML.

Wilder, J., Feldman, J. & Singh, M. (2011). Superordinate shape classi cation using natural shape statistics. Cognition, 119 (3), 325-340.

Willmott, C. J. & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30 (1), 79-82.

Wilson, M. & Kakouli-Duarte, T. (2009). Nematodes as enviromental indicator. CABI.

Woess, W. (2000). Random walks on in nite graphs and groups (Vol. 138). Cambridge University Press.

Apendice A

Herramienta de segmentacion manual para hitos de frontera

Herramienta de segmentacion de hitos fue implementada para el proceso de investigacion. En las etapas del proceso, se ha llevado numerosas tareas agrupadas con el n de cubrir las especi caciones descritas en el anteproyecto de investigacion.

Para satisfacer los requerimientos para la construccion de la base de datos, la herramienta permite cargar una imagen digital y anotar los hitos de frontera sobre el contorno del elemento estructural de forma que en conjunto simulan la silueta del organismo.

Despues del posicionamiento del primer hito en la imagen, se esbozan segmentos de aproximacion continuos entre cada par de puntos automaticamente. Cada punto se une al anterior y con el sucesor a traves de dichos segmentos. Cada hito puede ser movido para ajustar el contorno, y si es necesario, se puede incluir nuevos hitos entre los ya colocados con el n de ajustar mejor la silueta formada. Ademas, la herramienta permite anotar los hitos mas proximos a la cabeza y cola.

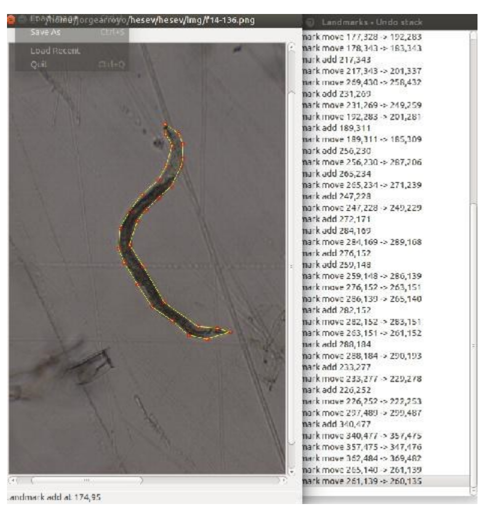


Figura A.1: Secuenciacion de hitos en una imagen usando HESEV.

Indice alfabetico