



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

INDIVIDUALIZACIÓN DE FILAMENTOS EN UNA RED EXTRAÍDA DESDE UNA
IMAGEN DE MICROSCOPIA, MEDIANTE OPTIMIZACIÓN

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN CIENCIAS, MENCIÓN COMPUTACIÓN

LEONARDO ENRIQUE PIZARRO BAEZA

PROFESOR GUÍA:
MAURICIO CERDA VILLABLANCA
JACQUES DUMAIS

MIEMBROS DE LA COMISIÓN:
ALEXANDRE BERGEL
JOSÉ MANUEL SAAVEDRA RONDO
VIOLETA CHANG CAMACHO

Este trabajo ha sido parcialmente financiado por el Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 1180906 y 1190806) y la ICM (P09-015-F)

SANTIAGO DE CHILE
2021

Resumen

En la naturaleza, es posible encontrar de forma ubicua, estructuras alargadas (filamentos), las que conforman redes entre sí. La conformación de estas estructuras complejas y dinámicas se puede observar en ejemplos particulares como en una red de proteínas de una célula eucariota, así como en bacterias, ya que a pesar que pertenecer a distintas familias, ambas tienen estructuras formada por filamentos. La individualización de filamentos permite cuantificar las propiedades de la red tales como número de filamentos, largo de estos, volumen, o curvatura, y puede ser categorizado en: basado en procesamiento de imágenes de bajo nivel o como un problema de optimización.

El problema de identificar filamentos en imágenes de microscopía está limitado por la resolución, y los problemas de múltiples parámetros a ajustar, para los métodos basados en procesamiento de imágenes de bajo nivel, el costo computacional en los métodos basados en optimización, y falta de descriptores cuantitativos en ambas. La revisión bibliográfica da cuenta también de pocas herramientas disponibles. Todo lo anterior implica que parte del análisis deba ser manual, lo que para grandes cantidades de datos, hace los estudios más propensos a errores.

Esta investigación se centra en el desarrollo de un algoritmo que pueda resolver un modelo de optimización para la individualización de filamentos a partir de un grafo que representa una red de filamentos, utilizando múltiples características. El uso de diversas características permite acotar el espacio de búsqueda, el que inicialmente consiste en todas las combinaciones posibles de aristas, lo que permite categorizar este problema como NP-Hard.

Como resultado, se obtiene un algoritmo flexible que individualiza filamentos de forma automática, lo que a su vez permite la incorporación de nuevas características y/o restricciones. La flexibilidad se demuestra mediante las diferentes configuraciones de parámetros disponibles, así como la posibilidad de personalizarlos. El algoritmo propuesto se evalúa en 12 imágenes, entre las que se tienen casos sintéticos, así como en microtúbulos de planta y neuronas de ratón, logrando obtener casos donde no existe diferencia estadísticamente significativa con lo indicado por un experto, y existiendo una diferencia menor en otros casos. Finalmente, el algoritmo propuesto realiza la individualización de los casos más complejos en un tiempo de cómputo inferior a los 40 segundos.

A mis padres, a mi amada Camila y a mis perros

Agradecimientos

Al finalizar este extenso recorrido quisiera agradecer profundamente a mis padres, cuyas enseñanzas y esfuerzo me han entregado enormes oportunidades en la vida. A mi pareja, Camila, por la paciencia, el apoyo incondicional y la motivación durante todo este proceso.

También deseo agradecer a los profesores y los equipos de distintos laboratorios que me ayudaron a construir este trabajo, así como a conocer nuevas áreas de investigación y colaboración. En particular destacar el gran apoyo de mi profesor guía, Mauricio Cerda, que me permitió conocer y aprender de Jacques Dumais y su gran equipo, así como haberme permitido ser parte del proyecto anillo DevMech.

Tabla de Contenido

1. Introducción	1
1.1. Hipótesis y Objetivos	2
1.1.1. Objetivo General	4
1.1.2. Objetivos Específicos	4
1.2. Estructura de esta tesis	4
2. Material Suplementario	5
3. Antecedentes	6
3.1. Métodos basados en procesamiento de imágenes de bajo nivel	6
3.1.1. SIFNE	6
3.1.2. FibrilTool	7
3.1.3. IFS	8
3.1.4. Marco de Análisis de Imágenes de Filamentos de Actina	10
3.2. Métodos basados en optimización	12
3.2.1. Restricciones Geométricas para Segmentar Estructuras Alargadas	12
3.2.2. SOAX	13
3.2.3. DeFiNe	16
3.3. Generación de un grafo desde una imagen	19
3.3.1. Extracción de un Grafo mediante Esqueletonización	21
3.3.2. Obtención de Información Adicional	22
3.4. Metaheurística ACO	23
3.4.1. Solución de un modelo COP mediante la metaheurística ACO	24
4. Algoritmo Propuesto	27
4.1. Inicialización de la metaheurística ACO	28
4.2. Método de construcción de caminos de cada hormiga	29
4.3. Método de actualización de feromonas	31
4.4. Criterios para la actualización de anti-feromonas	37
4.4.1. Curvatura de una solución	37
4.4.2. Magnitud de desplazamiento entre segmentos	37
4.4.3. Criterio específico para neuronas	39
4.5. Método de búsqueda no local	40
4.6. Extracción de información para individualizar filamentos	40
4.7. Ponderación de Características	42
4.8. Complejidad computacional del algoritmo propuesto	42

5. Metodología de evaluación	45
5.1. Métricas y Medidas	45
5.1.1. Índices Rand y Jaccard	45
5.1.2. Variation of Information	46
5.1.3. Mediciones adicionales	47
5.2. Imágenes Sintéticas	47
5.3. Imágenes Reales	49
6. Resultados	57
6.1. Extracción de un grafo desde una imagen	57
6.2. Imágenes Sintéticas	58
6.3. Imágenes Reales	63
6.3.1. Individualización de diferentes expertos sobre la misma red de filamentos	63
6.3.2. Microtúbulos de planta	63
6.3.3. Neuronas	76
6.3.4. Métrica VI e índices Rand y Jaccard	77
6.4. Resultados Generales	79
7. Conclusiones	81
Bibliografía	83
Apéndice A. Material Suplementario	87
A.1. Recuperación de Información mediante Superpixels	88
A.2. Individualización manual de filamentos	89
A.3. Resultados algoritmo propuesto en Imágenes Sintéticas	89

Índice de Tablas

4.1. Complejidad computacional de las etapas del algoritmo propuesto. n representa el número de nodos mientras que s representa el número de hormigas utilizadas.	43
5.1. Equivalencia entre casos para un par de aristas con su respectiva clasificación	46
5.2. Configuración de parámetros predefinidos por tipo de célula.	49
6.1. Resultado de la individualización de filamentos de la Figura 5.1.	59
6.2. Resultado de la individualización de filamentos para la Figura 5.2.	61
6.3. Resultados de individualización de filamentos para la muestra MT-A en la Figura 5.3.	65
6.4. Resultados de la individualización de filamentos para la muestra MT-A2 (Figura 6.3b).	65
6.5. Resultados de la individualización de filamentos para la muestra MT-A3 (Figura 6.3c).	66
6.6. Resultados de la individualización de filamentos para la muestra MT-B (Figura 5.4).	67
6.7. Resultados de individualización de filamentos para Figura 5.5, de la muestra MT-C.	67
6.8. Resultados de individualización de filamentos para Figura 6.9a, de la muestra MT-D.	73
6.9. Resultados de individualización de filamentos para la muestra MT-E en la Figura 6.10a.	75
6.10. Número de filamentos propuestos y tiempos de ejecución para las muestras N1, N2 y N3.	76
6.11. Resultados de la individualización de filamentos mediante el algoritmo propuesto para las muestra N1, N2 y N3.	77
6.12. Resultados de la métrica VI y los índices Rand y Jaccard para las evaluaciones del algoritmo propuesto.	79
6.13. Análisis estadístico de 10 iteraciones del algoritmo propuesto.	80
1. Resultados de 5 iteraciones de algoritmo propuesto para individualizar filamentos en la figura 5.1, con la configuración Modo 2.	90
2. Resultados de 5 iteraciones de algoritmo propuesto para individualizar filamentos en la figura 5.2, con la configuración Modo 3.	90

Índice de Ilustraciones

1.1.	Procedimiento para obtener un grafo que representa la red, utilizando segmentación y esqueletonización.	2
1.2.	Posibles resultados de la identificación de filamentos.	3
3.1.	Búsqueda por líneas con radio r para diferentes ángulos, según filtro LFT.	7
3.2.	Criterios del filtro OFT.	7
3.3.	Área de interés seleccionada manualmente en amarillo, y resultado de FibrilTool en rojo o verde.	9
3.4.	Ejemplo de FibrilTool en microtúbulos	10
3.5.	Etapas de individualización de filamentos de [1].	11
3.6.	Aplicación de filtros posterior a la segmentación.	12
3.7.	Ejemplos del método Restricciones Geométricas para Segmentar Estructuras Alargadas.	13
3.8.	Etapas de SOAX.	14
3.9.	Imagen de ejemplo de SOAX	15
3.10.	Imagen de ejemplo de SOAX	15
3.11.	Resultado de convergencia de curvas SOAC	16
3.12.	Identificación de intersecciones de curvas SOAC	17
3.13.	Extracción de caminos para formar P' en DeFiNe.	18
3.14.	Dificultad de individualización que enfretan los expertos al analizar manualmente una imagen de filamentos.	20
3.15.	Etapas de búsqueda de alimento y de comunicación entre hormigas.	25
4.1.	Diagrama de las etapas del algoritmo propuesto	28
4.2.	Camino de una hormiga b compuesto por 2 segmentos.	32
4.3.	Camino de una hormiga a compuesto por 4 segmentos.	33
4.4.	Funcionamiento de la propuesta de penalización de anti-feromonas con segmentos y recorrido inverso.	35
4.5.	Caminos de las hormigas a y b	36
4.6.	Curvatura de un camino de una hormiga a que contiene las aristas e_1, e_3, e_4 y e_7	38
4.7.	Criterio de magnitud y pronunciación del desplazamiento entre 2 segmentos contiguos.	39
4.8.	Ejemplo del algoritmo <i>closeness centrality</i> eligiendo los 3 nodos con menor distancia a los demás nodos.	41
5.1.	Filamentos sintéticos a evaluar con el algoritmo propuesto.	48

5.2.	Red de filamentos sintéticos a evaluar con el algoritmo propuesto.	48
5.3.	Muestra MT-A de <i>Arabidopsis Marchantia</i> con su individualización manual de microtúbulos.	51
5.4.	Individualización manual de microtúbulos para la muestra MT-B.	52
5.5.	Individualización manual de microtúbulos para la muestra MT-C.	53
5.6.	Muestra N1 con su individualización manual de una neurona de ratón.	54
5.7.	Muestra N2 con su individualización manual de una neurona de ratón.	55
5.8.	Muestra N3 con su individualización manual de una neurona de ratón.	56
6.1.	Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 5.1.	60
6.2.	Resultados de la individualización de filamentos para la Figura 5.2, obtenidas mediante DeFiNe y el algoritmo propuesto, con diferentes parámetros.	62
6.3.	Individualización de dos expertos ante la misma red de filamentos	64
6.4.	Filamentos propuestos y correctamente individualizados a partir la muestra MT-A en la Figura 5.3.	68
6.5.	Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.3b (Muestra MT-A2).	69
6.6.	Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.3c (Muestra MT-A3).	70
6.7.	Filamentos correctamente individualizados en la figura 5.4 correspondiente a la muestra MT-B.	71
6.8.	Comparación entre filamentos propuestos y correctamente individualizados para la muestra MT-C en la Figura 5.5, con los distintos métodos y parámetros.	72
6.9.	Individualización de filamentos para la Figura 6.9a (Muestra MT-D).	74
6.10.	Individualización de filamentos para la Figura 6.10a (Muestra MT-E).	75
6.11.	Filamentos propuestos e individualizados correctamente para las muestras N1, N2 y N3, por el algoritmo propuesto.	78
1.	Ejemplo de un filamento individualizado manualmente	89

Capítulo 1

Introducción

En la naturaleza, es posible encontrar de forma ubicua, estructuras alargadas (filamentos), las que conforman redes entre sí. La conformación de estas estructuras complejas y dinámicas se puede observar en ejemplos particulares como en una red de proteínas de una célula eucariota, así como en bacterias, ya que a pesar de pertenecer a distintas familias, ambas tienen estructuras formadas por filamentos. Características físicas de estas redes generan propiedades tales como la presencia o ausencia de ciclos, o la posibilidad de dividir o no cada filamento. Por su parte, el análisis de filamentos que conforman la red, puede indicar el estado de esta, respecto a su ambiente o de su interior, así como develar información relevante sobre la relación entre la estructura biológica y funciones fisiológicas.

Los métodos actuales para analizar las redes mencionadas se basan en el procesamiento directo de imágenes obtenidas a partir de microscopía (Figura 1.1a), pasando por etapas de segmentación (Figura 1.1b), para luego utilizar diversas técnicas como esqueletonización (Figura 1.1c), la transformada de Radón o *template matching*. Esto permite identificar el grafo que representa a la red (Figura 1.1d) o sirve de base para el uso de heurísticas que permiten identificar los filamentos de forma individual. La individualización de filamentos permite cuantificar las propiedades de la red tales como número de filamentos, largo, volumen, o curvatura. Estos métodos, basados en la observación mediante microscopía óptica tienen como cota máxima de resolución el límite de difracción, $\lambda/2$. Donde λ es la longitud de onda de la luz utilizada (o color). Este límite establece que 2 objetos cuya distancia sea inferior a $\lambda/2$ no pueden ser diferenciados, conllevando a que dos partes cercanas del grafo puedan ser observadas como una, dificultando su estudio. Lo anterior es relevante para asociar las propiedades de la red a los vértices del grafo extraído, dando pie a la caracterización del mismo (Figura 1.2a).

Existen investigaciones en la literatura que apuntan a las distintas etapas de este problema, las que tienen en común el uso de técnicas del área de procesamiento de imágenes de bajo nivel para el tratamiento inicial de la imagen de microscopía. La individualización de filamentos puede ser categorizada dependiendo si utiliza como enfoque primario herramientas de procesamiento de imágenes de bajo nivel, o si se aborda como un problema de optimización. En ambas categorías, las críticas más repetidas en los trabajos del área suelen ser la cantidad de parámetros y la dificultad en su ajuste, en las diversas herramientas



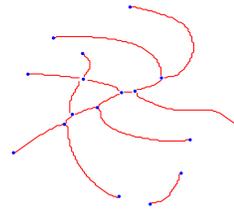
(a) Representación simplificada de una red con cruces y sobreposición de filamentos en una imagen de microscopía.



(b) Preprocesamiento de la imagen mediante segmentación para la extracción de la red.



(c) Esqueletonización representativa de la red sobre la imagen de 1.1b.



(d) Grafo que representa la esqueletonización de la red.

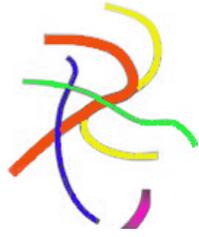
Figura 1.1: Procedimiento para obtener un grafo que representa la red, a partir del procesamiento de una imagen de microscopía, utilizando segmentación y esqueletonización. Fuente: Breuer and Nikoloski [2].

existentes. Un segundo problema común es que la obtención de información relacionada a la morfología y el comportamiento de las redes es más cualitativo que cuantitativo, como se indica en Asgharzadeh et al. [3] y en Qiu and LI [1], lo que supone un problema al trasladar el análisis a una gran cantidad de datos, dado que cada enfoque es demasiado específico a su correspondiente software.

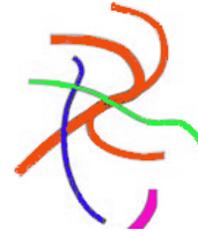
En resumen, el problema de identificar filamentos en imágenes de microscopía está limitado por la resolución, y los problemas de múltiples parámetros a ajustar, para los métodos basados en procesamiento de imágenes de bajo nivel, el costo computacional en los métodos basados en optimización, y falta de descriptores cuantitativos en ambas. La revisión bibliográfica da cuenta también de pocas herramientas disponibles. Todo lo anterior implica que parte del análisis deba ser manual, lo que para grandes cantidades de datos, hace los estudios más propensos a errores.

1.1. Hipótesis y Objetivos

Como se ha presentado, los métodos de individualización de filamentos que sólo usan herramientas de procesamiento de imágenes de bajo nivel, pueden representar la red a través de los filamentos que la conforman, sin embargo para realizar mediciones cuantitativas requieren ajustar múltiples parámetros. Por otra parte los métodos de optimización permiten reducir



(a) Visualización de un resultado posible de individualización, limitando los filamentos identificados a filamentos simples.



(b) Visualización de otro resultado posible de individualización, permitiendo filamentos más complejos, como es el caso de filamentos con ramificaciones.

Figura 1.2: Posibles resultados de la identificación de filamentos. Un planteamiento enfocado en un modelo de optimización permitirá asociar las aristas del grafo, como el de la figura 1.1d, a uno o más filamentos, entregando resultados como (a) o (b), expresado gráficamente mediante gradiente de colores. Fuente: Elaboración propia.

los parámetros, pero con un alto costo computacional, el que puede ser abordado con el uso de heurísticas o algoritmos de aproximación. También se debe tener en consideración en ambos métodos el número de características utilizadas al momento de realizar la individualización. Esto nos permite establecer la pregunta:

¿Es posible determinar correctamente la cantidad de filamentos, desarrollando un modelo de optimización que utilice una combinación de múltiples características como el largo, grosor, ángulo de bifurcación, curvatura o dirección, tanto en la asignación de pesos como en la selección del subconjunto de caminos?

Al comenzar el proceso de individualización de filamentos se desconoce el origen y fin de cada filamento, por lo que plantear un recorrido de grafos puede resultar computacionalmente costoso, lo que puede ser evitado al plantear un modelo de optimización que pueda hacer la asociación entre aristas que sean parte del mismo filamento.

Se debe agregar que existe información *a priori*, indicada por el tipo de estructura observada, la cual restringe los comportamientos posibles de la red. Por ejemplo, los filamentos de proteína actina no se pegan entre sí, formando estructuras sin ciclos, mientras que el retículo endoplasmático (organelo celular encargado principalmente de la síntesis de proteínas) si presenta ciclos. Estas condiciones pueden aportar límites más acotados a algunos de los criterios de cuantificación y es conocimiento disponible previo a la observación en el microscopio. Es importante destacar que alguna de la información *a priori* corresponde a reglas empíricas, que también deben ser consideradas durante el análisis.

En base a lo anterior, la hipótesis de esta investigación se basa en que a partir de un grafo con pesos, no dirigido, con o sin ciclos, que representa una red de filamentos, en conjunto con utilizar una combinación de características de los segmentos de filamentos como largo, grosor, ángulo de bifurcación o curvatura, sumado a la incorporación de información previa disponible sobre el tipo de célula (red de filamentos con o sin ciclos), es posible identificar a cada filamento en la red resolviendo un modelo de optimización.

Para validar esta hipótesis, se propone realizar el siguiente objetivo general y los siguientes objetivos específicos:

1.1.1. Objetivo General

Desarrollar un modelo de optimización para la individualización de filamentos a partir de un grafo representativo de la red de filamentos, evaluando la variación en el resultado con distintas combinaciones entre las propiedades de cada segmento de filamento como el grosor, largo, ángulo de bifurcación o dirección.

1.1.2. Objetivos Específicos

- Generar un modelo de optimización para la identificación de filamentos a partir de un grafo con pesos que representa la red de filamentos.
- Implementar un algoritmo que resuelva el modelo de optimización, entregando como salida la identificación de filamentos, considerando casos de solapamiento y/o cruce.
- Identificar la ponderación de propiedades que entregue mejores resultados para grafos que representen una neurona, una bacteria y una célula eucariota de planta.
- Evaluar técnicas en el estado del arte que realicen individualización de filamentos.

Como resultado de los objetivos planteados, se obtendrá una herramienta para el análisis automatizado de estructuras de filamentos, lo que permitiría mejorar el seguimiento de procesos dinámicos o la identificación de procesos patológicos de forma automática.

1.2. Estructura de esta tesis

El capítulo 3 resume el estado del arte de diversos aspectos relacionados a la individualización de filamentos. El capítulo 4 define un modelo de optimización que resuelve la identificación de filamentos, para el que se propone un algoritmo. Se presentan las métricas y mediciones en el capítulo 5. El capítulo 6 presenta los resultados obtenidos para imágenes sintéticas e imágenes reales de distintas células. Finalmente, en el capítulo 7 se presentan las conclusiones.

Capítulo 2

Material Suplementario

Capítulo 3

Antecedentes

La individualización de filamentos en una imagen es un problema derivado de la obtención de información de filamentos en base a una imagen. Este último ha ido variando a medida que la tecnología en microscopios ha mejorado, lo que a su vez ha permitido el desarrollo de nuevos métodos. Estos métodos pueden a su vez diferir entre sí dependiendo del foco de la información que buscan obtener a partir de la imagen, existiendo métodos que se enfocan en reconstruir filamentos discontinuos, a la identificación de segmentos de filamento para la construcción de las redes que estos conforman, o la individualización en sí. Independientemente del objetivo de cada método, es posible agruparlos en dos categorías: métodos basados en procesamiento de imágenes de bajo nivel y métodos basados en optimización. Los primeros realizan mayoritariamente operaciones de filtrado sobre la imagen, mientras que los segundos plantean parte de la obtención de la información mediante la resolución de problemas de asignación o de minimización de funciones.

Entre las herramientas disponibles para los métodos basados en optimización, se encuentra la metaheurística *Ant Colony Optimization* (ACO), que se basa en el comportamiento de las hormigas al buscar alimento. Esta metaheurística corresponde a la estrategia utilizada en esta investigación para individualizar filamentos.

3.1. Métodos basados en procesamiento de imágenes de bajo nivel

3.1.1. SIFNE

La investigación de Zhang et al. [4] identifica el problema de discontinuidad de filamentos en una imagen, pudiendo atribuirse esto a factores experimentales como la densidad del componente fluorescente, ruido u otras. En particular, el filamento analizado en esta investigación corresponde a microtúbulos. Para llevar a cabo la reconstrucción de un filamento en base a segmentos, establecen el uso de dos filtros, denominados *Filtro de Transformación Lineal* (LFT, Linear Filter Transformation) y *Filtro de Transformación en base a la Orientación* (OFT, Orientation Filter Transformation). Estos filtros se implementan en un software denominado *SIFNE*. El filtro LFT busca resaltar características lineales centrándose

en cada píxel y generando una serie de líneas con radio r , buscando la línea que contenga la mayor intensidad, según se observa en la Figura 3.1. El filtro OFT complementa lo anterior, mediante un criterio de similitud en la dirección proyectada de los segmentos, un segundo criterio respecto a la distancia entre los extremos finales de cada segmento y un tercer criterio de continuidad, que limita el ángulo entre el vector proyectado de un extremo y el vector que representa la distancia entre los extremos de cada segmento. Lo anterior se puede observar en la Figura 3.2.

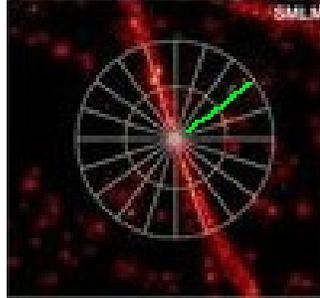


Figura 3.1: Búsqueda por líneas con radio r para diferentes ángulos, según filtro LFT. Fuente: [4]

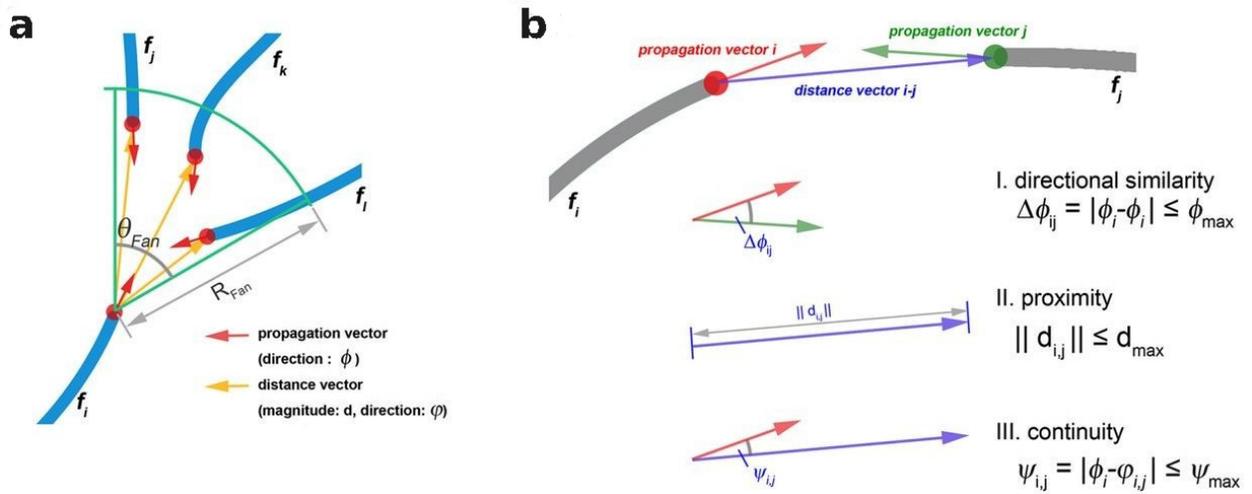


Figura 3.2: Criterios del filtro OFT: (a) Vectores de propagación (rojo) y distancia (amarillo) (b) Ejemplos de criterios de coincidencia de fragmentos de filamento I) Similitud de dirección, II) Proximidad, III) Continuidad. Fuente: Zhang et al. [4].

El fundamento de los criterios de OFT se basa en la asociación de las características geométricas como restricciones representativas del comportamiento mecánico de los microtúbulos. Uno de los objetivos de esta investigación es permitir el análisis automático, dada las dificultades que presenta el análisis manual de reconstrucción.

3.1.2. FibrilTool

El trabajo de Boudaoud et al. [5] conduce al desarrollo de *FibrilTool*, un plugin para el software *ImageJ*, utilizado en análisis científico de imágenes. FibrilTool calcula la orientación

principal y la anisotropía de estructuras alargadas dentro de una región de interés seleccionada manualmente. Para esto, utiliza el concepto de tensor nemático, extraído del comportamiento físico de los cristales líquidos. Específicamente, el tensor nemático es la matriz simétrica n de 2×2 construida a partir de un vector unitario t , ecuación 3.1, definido en base a la derivada de primer orden de la intensidad del pixel en x, y .

$$t = (t_x, t_y) = \left(\frac{\partial I}{\partial y}, -\frac{\partial I}{\partial x} \right) / \sqrt{\left(\frac{\partial I}{\partial x} \right)^2 + \left(\frac{\partial I}{\partial y} \right)^2} \quad (3.1)$$

Luego, a partir de la matriz n (ecuación 3.2) se obtiene su primer vector propio \vec{e}_1 que representa la orientación principal de los filamentos en el área de interés, mientras que la diferencia de los valores propios λ_1 y λ_2 ($\lambda_1 > \lambda_2$), denominada $q = \lambda_1 - \lambda_2$, define la anisotropía.

$$n = \begin{bmatrix} (t_x)^2 & t_x t_y \\ t_x t_y & (t_y)^2 \end{bmatrix} \quad (3.2)$$

Los autores de FibrilTool desarrollan este método para evitar el uso de derivadas de segundo orden, que presentan sensibilidad al ruido, necesitando de pasos previos en la limpieza de la imagen. Un ejemplo de la ejecución de FibrilTool sobre una imagen de microtúbulos de la planta *Arabidopsis Marchantia* se puede observar en la Figura 3.4.

A partir del ejemplo, el resultado de FibrilTool se indica mediante una línea de color rojo cuyo ángulo representa la orientación promedio de los filamentos en la región de interés, mientras que el largo es un indicador proporcional a la anisotropía de los filamentos. Además de la respuesta gráfica, FibrilTool genera una salida de datos en formato multi-columna, indicando la orientación promedio y la anisotropía (definida como q) en las columnas 6 y 7 respectivamente. Para el ejemplo, el ángulo tiene un valor de 52.78° , con un valor de anisotropía de 0.1493. El ángulo entregado por la herramienta siempre estará en el rango $[90, 90]$, mientras que la anisotropía puede variar entre 0 y 1, ambos inclusivos.

Debido a que el proceso para obtener un resultado a partir de FibrilTool considera una sola área de interés, los autores desarrollaron una versión denominada *FibrilTool Batch* (Louveaux and Boudaoud [6]), la que permite la creación de múltiples regiones de interés para ser calculadas.

3.1.3. IFS

Una investigación que propone la individualización de segmentos de filamentos es Qiu and LI [1], la que plantea un filtro de detección de características de 6 pasos más un algoritmo llamado *SBDA* que busca eliminar segmentos de filamentos menores a 2 píxeles (Figura 3.5), cuya implementación se denomina *IFS*. La individualización de segmentos de filamentos entrega información morfológica de un segmento con respecto a su entorno, señalando si se encuentra en estado de aislamiento, intersectado, bifurcado o en superposición con otro segmento de filamento.

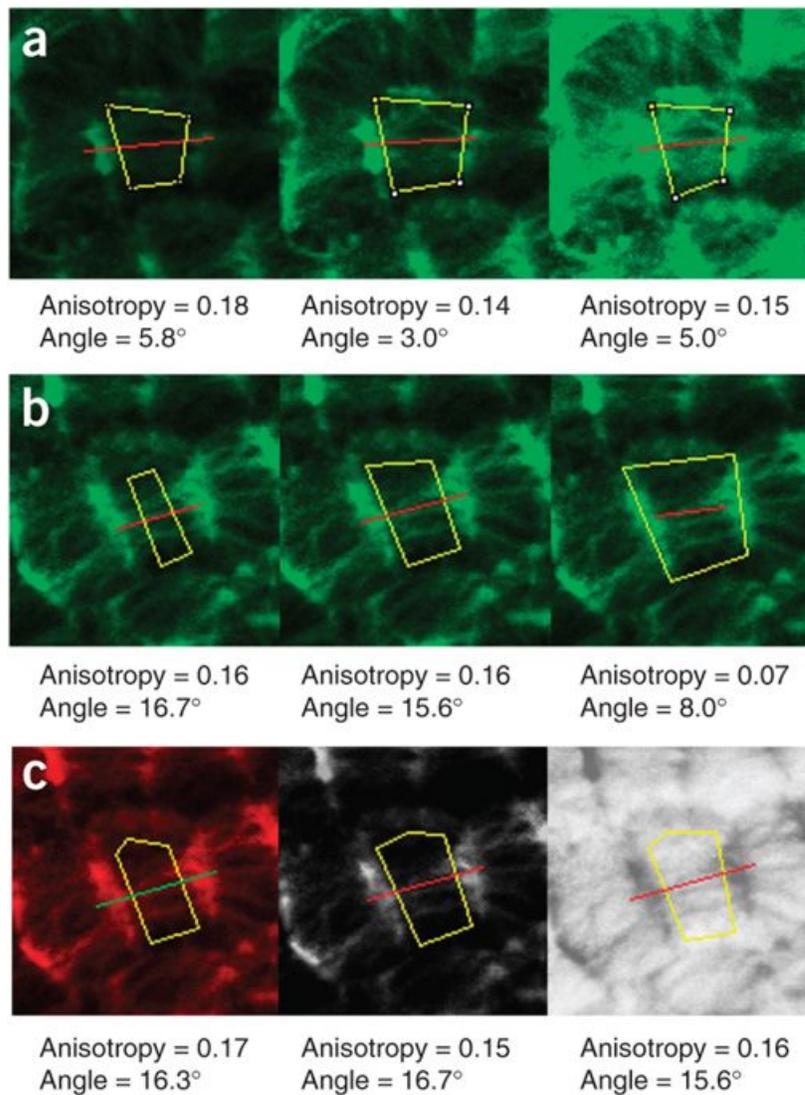


Figura 3.3: Área de interés seleccionada manualmente en amarillo, y resultado de FibrilTool en rojo o verde. Fuente: Boudaoud et al. [5].

Las 6 etapas del filtro consideradas en este método consisten en 4 etapas de análisis de la imagen y 2 de análisis topológico que se resumen en lo siguiente:

1. Depuración de la imagen: El primer filtro comienza por reducir el ruido y realzar el contraste de la imagen, aumentando la intensidad de los píxeles que estén sobre un umbral, al mismo tiempo que se descartan los píxeles que se encuentren debajo del mismo umbral.
2. Filtro de información estructural: El segundo paso consiste en filtrar información estructural, que se basa en los valores propios de la matriz Hessiana, buscando descartar objetos en la imagen que no correspondan a una figura tubular.
3. Eliminación de señales débiles: El tercer filtro ejecuta una limpieza de estructuras con una intensidad baja y/o aisladas, de las que puede concluirse que no corresponden a elementos en el plano focal de interés.
4. Generación de esqueleto: Finalmente dentro del análisis sobre la imagen, el filtro de

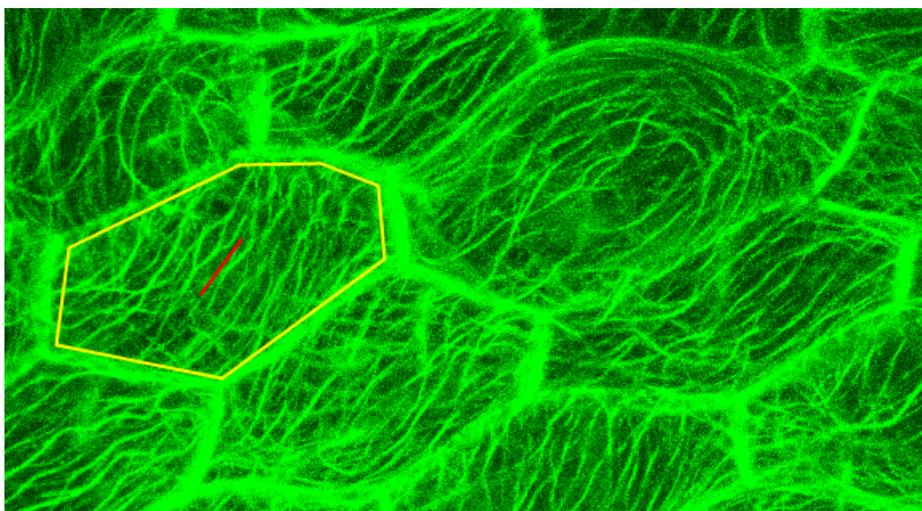


Figura 3.4: Uso de FibrilTool en una imagen de microtúbulos de la planta *Arabidopsis Marchantia*. La sección de interés corresponde al área dentro de la región delimitada en amarillo. La línea roja representa el resultado al ejecutar FibrilTool. El ángulo de orientación promedio para este ejemplo es 52.78° , mientras que la anisotropía reporta un valor de 0.1493. Fuente: Elaboración Propia.

esqueletonización realiza un adelgazamiento de la imagen, en el que cada estructura pasa a tener 1 píxel de ancho, facilitando el análisis topológico posterior.

5. Clasificación topológica y algoritmo SBDA: Consiste en el análisis a nivel de píxel y su vecindario de 8 píxeles alrededor para determinar si este corresponde a un punto aislado, al final de un fragmento de filamento, a un punto interior de un filamento, o a una unión de filamentos. Seguido de aquello, el algoritmo SBDA realiza otro análisis topológico a nivel de píxel que borra los segmentos menores a 3 píxeles, además de realizar los cálculos de distancia de cada segmento/fragmento.
6. Reconstrucción: Para finalizar el análisis topológico y el proceso de filtrado, se lleva a cabo una combinación de segmentos para construir los filamentos basándose en el parámetro W , llamado *ancho efectivo*, cuyo valor define si la unión de 2 o más segmentos se trata de una bifurcación, una intersección o un solapamiento.

Como resultados de [1], además de la información morfológica, se obtienen características geométricas como el largo de los segmentos de filamentos y la distribución de la orientación de los segmentos, así como el cambio de estos valores para variaciones en la relación entre la señal y el ruido de la imagen.

3.1.4. Marco de Análisis de Imágenes de Filamentos de Actina

La investigación de Alioscha-Perez et al. [7] presenta un marco para el análisis de imágenes con la finalidad de identificar filamentos de actina, mediante una combinación de filtros con un algoritmo de unión de segmentos. La propuesta inicial se basa en que una imagen puede ser separada en 3 componentes: el fondo o *background* de la imagen, los filamentos y el ruido. Para obtener la imagen que contiene los filamentos y descartar en mayor medida el ruido, los autores utilizan la librería MCALab en MATLAB. Luego, buscan intensificar los píxeles

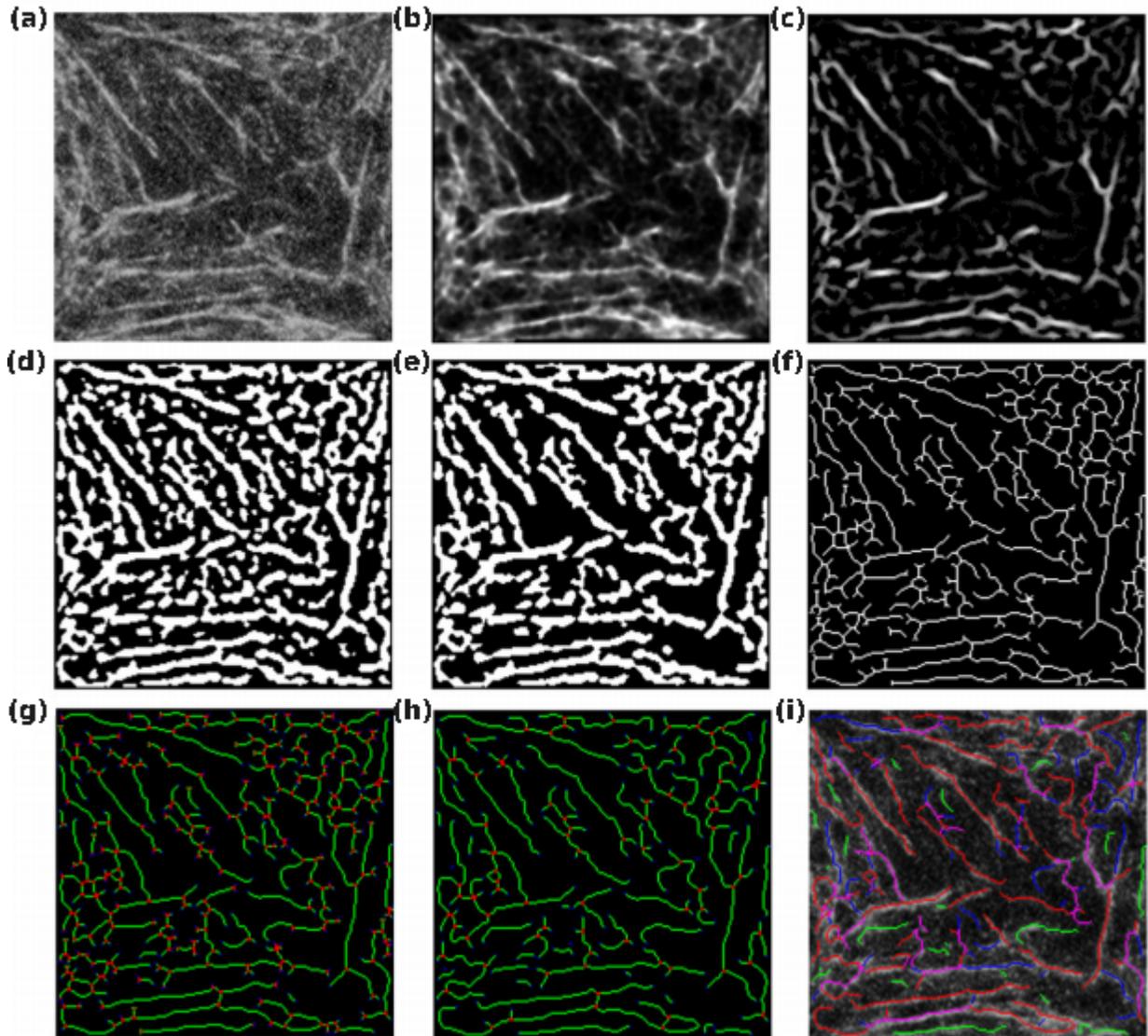


Figura 3.5: Etapas de individualización de filamentos de [1]. (a): Imagen de red de fibras de un osteoblasto. (b, c, d, e, f): Filtros de limpieza, tubularidad, segmentación, conectividad y esqueletonización. (g): Clasificación topológica de intersecciones a nivel de píxel. (h): Resultado de algoritmo SBDA. (i): Individualización de segmentos de filamentos según forma estructural: aislado (verde), solapado (morado) u otro (azul). Fuente: [1].

que corresponden a los filamentos, para lo cual aplican 3 filtros: un filtro Gaussiano, un filtro Laplaciano y un filtro Gaussiano direccional, obteniendo como resultado lo que se muestra en la Figura 3.6.

El paso siguiente es la aplicación de un detector de líneas multi-escala, que busca determinar la pertenencia de un píxel a una línea, que varía su grosor y ángulo dependiendo del tamaño/escala del vecindario elegido. Las líneas varían en su grosor $s \in [1, W]$, con W el grosor esperado de una fibra de actina, así como en su ángulo, que varía de forma discreta entre 0 y 180°. El detector de líneas asigna una puntuación que determina la probabilidad del píxel a pertenecer a una de las líneas. Lo anterior entrega como resultado candidatos de

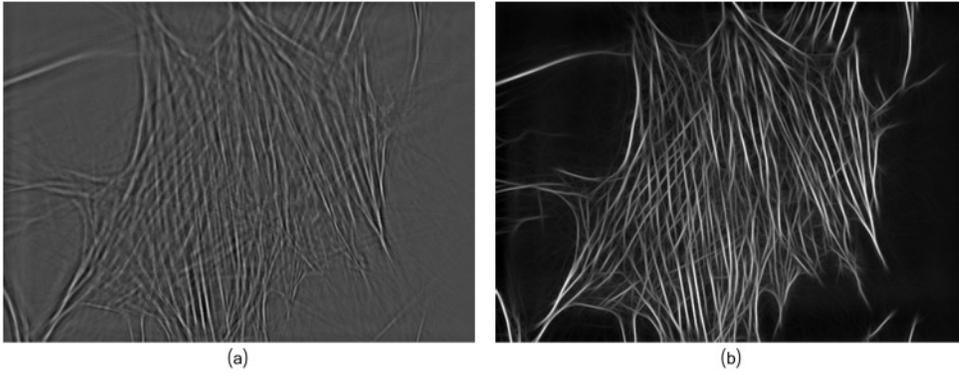


Figura 3.6: A partir de la segmentación obtenida con la librería MCALab (a), se aplican los filtros Gaussiano, Laplaciano y Gaussiano direccional para obtener una intensificación de los filamentos (b). Fuente: Alioscha-Perez et al. [7].

segmentos de líneas. Para elegir a un candidato como segmento de línea definitivo, la idea principal es recorrer los píxeles en secuencia y ajustar las líneas candidatas usando el método de mínimos cuadrados hasta que un umbral de error es sobrepasado. Por cada vez que se sobrepasa el umbral de error, se comienza la elección de un nuevo segmento de línea para los píxeles que siguen. Si el candidato de segmento de línea tiene un largo $l < L$, se descarta. L es un parámetro definido empíricamente.

Finalmente, para construir las fibras de actina, los segmentos de línea elegidos en el paso anterior son unidos entre sí mediante una ventana de largo L , en la que dos segmentos de línea se superpongan y no tengan una diferencia entre sus ángulos mayor al umbral T_θ .

Otras investigaciones similares que siguen un enfoque basado principalmente en procesamiento de imágenes de bajo nivel son Usov and Mezzenga [8], Lichtenstein et al. [9] y Asgharzadeh et al. [3]. En general, los trabajos desarrollados bajo este enfoque requieren del uso o sintonización de múltiples parámetros.

3.2. Métodos basados en optimización

3.2.1. Restricciones Geométricas para Segmentar Estructuras Alargadas

En la segunda categoría, Cerda et al. [10] plantea un algoritmo de 2 etapas para identificar y fusionar segmentos de filamentos. La primera etapa se enfoca en la detección de segmentos mediante un filtro de umbral automático y su posterior esqueletonización, permitiendo identificar los segmentos resultantes en base a los nodos en sus extremos. A partir de los segmentos de filamentos, la segunda etapa consiste en generar una matriz de costos para determinar que segmentos es posible fusionar. El funcionamiento del método propuesto por los autores se refleja en la Figura 3.7.

La construcción de la matriz de costos, de tamaño $n \times 2n$ con n como el número de nodos, se basa en la distancia euclidiana y la curvatura entre segmentos. Los autores plantean que es posible tratar el problema de fusión de segmentos de filamentos como un problema de

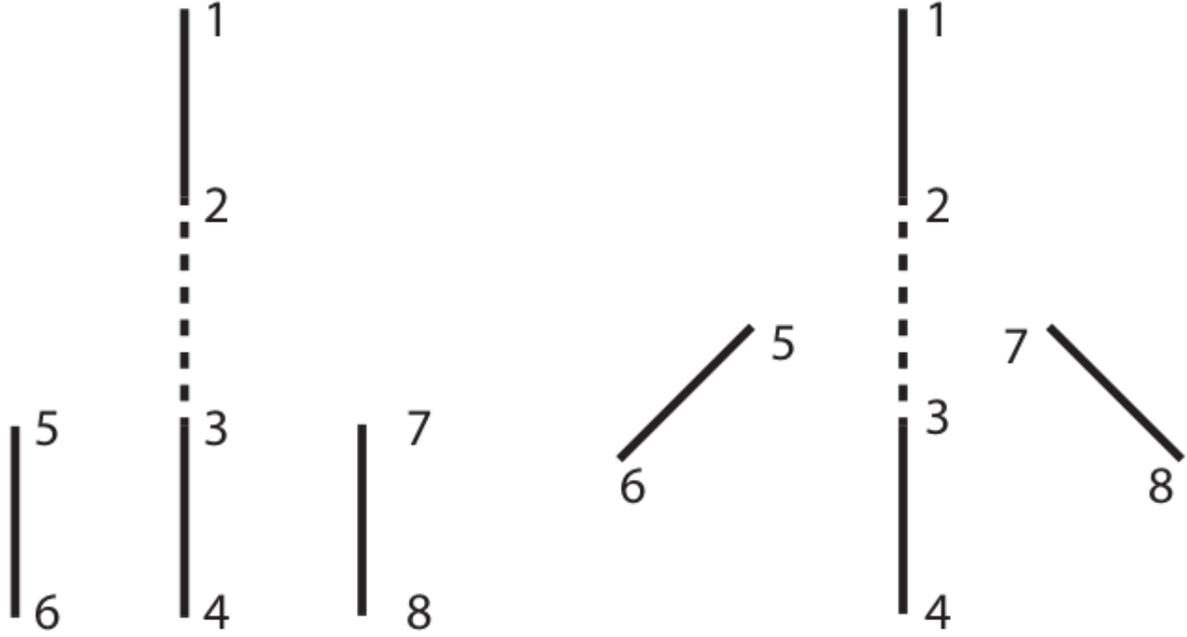


Figura 3.7: Ejemplos del método Restricciones Geométricas para Segmentar Estructuras Alargadas. La línea punteada refleja la fusión entre los segmentos de filamentos 1-2 y 3-4, representados por líneas continuas. En ambos ejemplos se fusionan los segmentos utilizando criterios de curvatura y distancia euclidiana. Fuente: Cerda et al. [10].

asignación, al que se le encuentra una solución óptima al resolver la matriz utilizando el algoritmo húngaro, con un costo computacional de $\mathcal{O}(n^3)$.

3.2.2. SOAX

Por su parte, la investigación de Xu et al. [11] llamada SOAX, emplea curvas paramétricas de contorno abierto (SOAC, *Stretching Open Active Contour*) en conjunto con una función de minimización para obtener un grupo acotado de redes de filamentos, entre las que el usuario puede elegir una, con la finalidad de realizar un análisis posterior. Las curvas paramétricas de contorno abierto son reguladas por los parámetros τ y K_{str} . El parámetro τ fija el umbral de intensidad que permite discriminar los puntos en la imagen donde existen máximos locales, a partir de los cuales puede comenzar el recorrido de una curva SOAC. Por su parte, el parámetro, K_{str} es el factor que regula la elongación/evolución de cada curva paramétrica de contorno abierto (SOAC). Una vez concluida la inicialización y evolución de curvas SOAC, se identifican las uniones/intersecciones generadas entre diversas curvas SOAC, agrupándose según su cercanía. Lo anterior se puede observar en la Figura 3.8.

La función de minimización que los autores denominan como *F-Function*, $F = -L_{total} + c \cdot L_{<t}$, se encuentra definida por otros 2 parámetros: el factor c ($c > 1$) que regula la penalización de las SOACs con bajo *Signal to Noise Ratio (SNR)*, y el umbral t , que define el valor mínimo de SNR necesario para no descartar una SOAC. El umbral t se ve reflejado

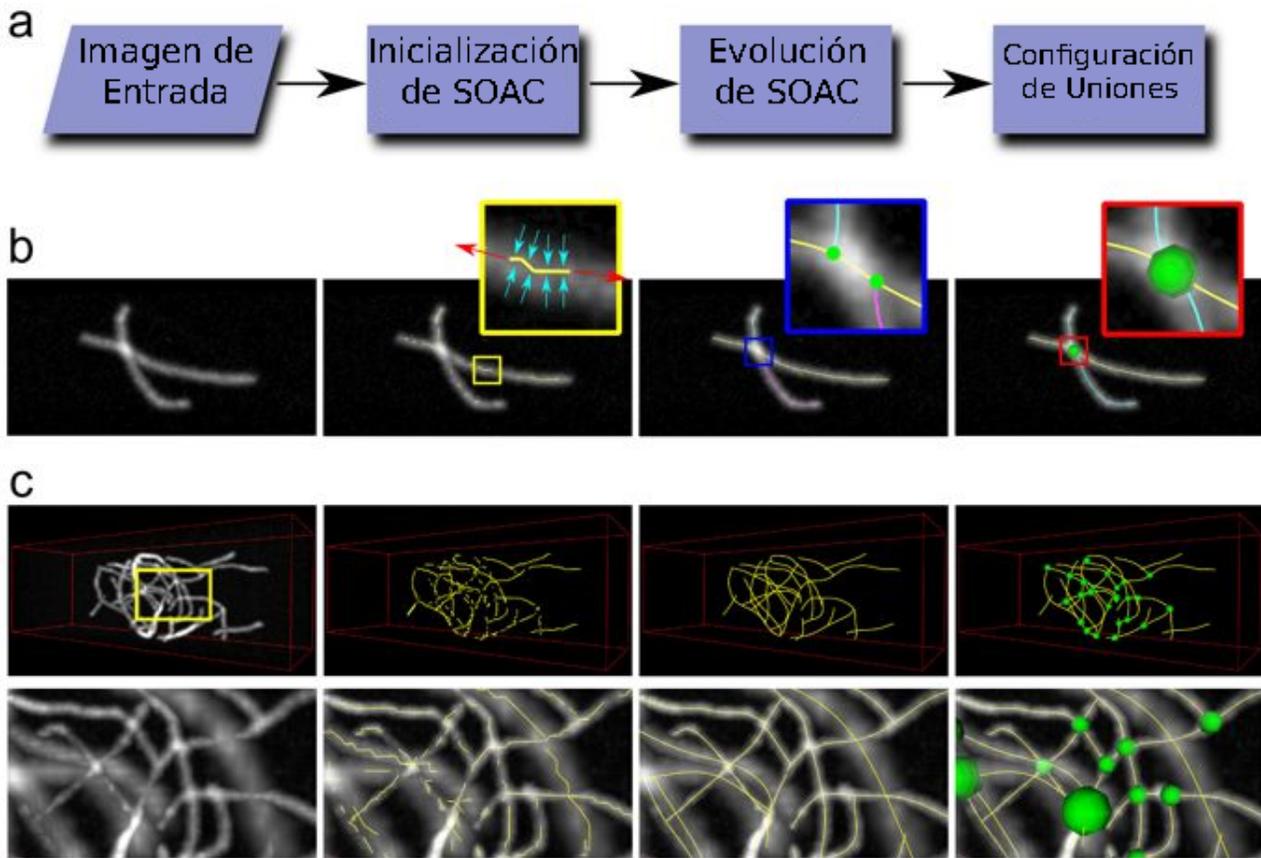


Figura 3.8: Etapas de SOAX: A partir de los puntos de alta intensidad se inicializan curvas paramétricas de contorno abierto (SOAC), las que crecen, generando intersecciones entre ellas. Finalmente, se agrupan las intersecciones más cercanas. Fuente: Xu et al. [11].

en $L_{<t}$ dentro de la *F-Function*, como el largo de SOACs en una región de la imagen con SNR por debajo de t , y que serán penalizados dado que son calificados como de baja certeza. L_{total} representa el largo total de las SOACs en el resultado final.

Un ejemplo del funcionamiento de SOAX en base a una imagen de microtúbulos en la planta *Arabidopsis Marchantia* (Figura 3.10) puede observarse en la Figura 3.11. En base a la imagen de ejemplo, SOAX obtiene múltiples características de la red de filamentos, a partir de las curvas paramétricas SOAC identificadas, las que se visualizan de color morado. Estas características son la rotación esférica, la orientación, la densidad de puntos, curvatura y largo de estas. A su vez, SOAX también identifica las intersecciones de las curvas SOAC, lo cual corresponde a los puntos verdes en la Figura 3.12.

La obtención de las curvas SOAC requiere de la ejecución del proceso de inicialización de curvas, identificado con el símbolo $+$, el que entrega un resultado preliminar, que corresponde a 90 curvas para la imagen de ejemplo. Luego, para reducir el número de curvas obtenidas es necesario ejecutar el proceso de convergencia de curvas, mediante el símbolo \triangleright , lo cual transforma las 90 curvas previas en solo 21. Estos 2 pasos son los que permiten obtener las curvas SOAC, siendo lo que se observa en la Figura 3.11, sin embargo, no existe una identificación automática de las intersecciones o junciones entre estas. Esto último, observado



Figura 3.9: Imagen recortada con microtúbulos de *Arabidopsis Marchantia* para el ejemplo de SOAX. Para ser aceptada como entrada valida en SOAX, la imagen tuvo que ser convertida de 32 a 16 bit. Fuente: Llanos [12].

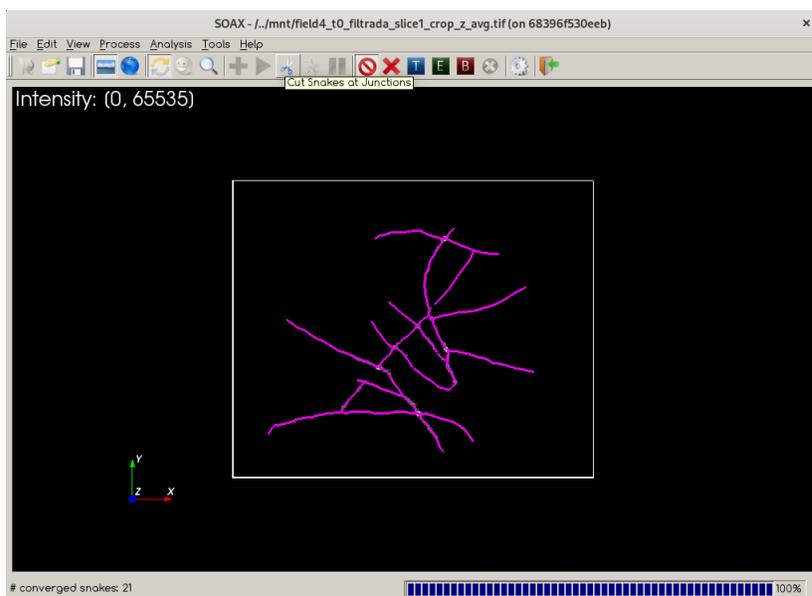


Figura 3.10: Imagen recortada con microtúbulos de *Arabidopsis Marchantia* para el ejemplo de SOAX. Fuente: Elaboración propia.

en la Figura 3.12, implica un paso adicional en el que se cortan de curvas en puntos de intersección, según un criterio que apunta a evitar que las curvas SOAC resultantes luego de la convergencia no finalicen en una intersección o que no tengan bifurcaciones marcadas.

Dado que el enfoque de la investigación de SOAX se encuentra en obtener un análisis cuantitativo de la red de filamentos, las curvas SOAC pueden coincidir con un filamento o simplemente corresponder a un segmento de filamento, sin existir la posibilidad de diferenciar ambos casos. En cuanto a los parámetros, SOAX presenta 26 campos para poder modificar valores, con 2 opciones adicionales del tipo booleano.

Los mismos autores de SOAX crearon una extensión, denominada TSOAX (Xu et al. [13]), la cual incorpora la posibilidad de analizar y extraer redes de filamentos a partir de un conjunto de imágenes que reflejen la evolución de los filamentos en el tiempo. Los pasos utilizados para la construcción de SOAX y TSOAX pueden encontrarse en los archivos del tipo *Dockerfile* disponibles en el repositorio de esta tesis para su reproducción (Pizarro [14]).

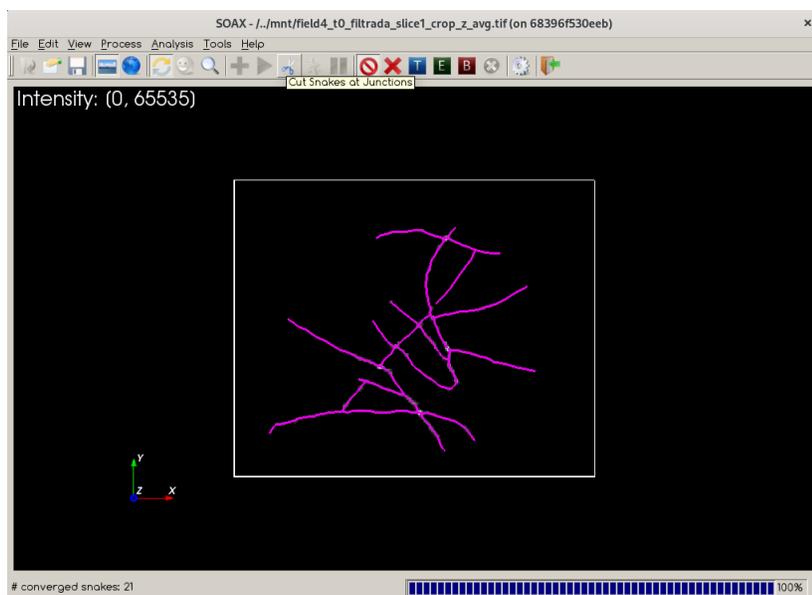


Figura 3.11: Resultado de SOAX al procesar la imagen recortada con microtúbulos de *Arabidopsis Marchantia*, obteniendo 21 curvas paramétricas tras la convergencia. Fuente: Elaboración Propia.

3.2.3. DeFiNe

En esta categoría también se incluye Breuer and Nikoloski [2], que en base a un grafo no dirigido con pesos que representa la red de filamentos, realiza la individualización de estos. El uso de un grafo como dato de entrada permite representar a cada filamento mediante un conjunto de aristas adyacentes en el grafo. Esto permite que la búsqueda e individualización de filamentos, con un segmento de filamento representado por una o más aristas del grafo, sumado a las restricciones que plantea el autor, sea tratado como un problema de *Set Cover* (Caprara et al. [15]). La definición de *Set Cover* es:

Dado un conjunto de elementos, denominado universo, y n conjuntos cuya unión comprende el universo, el *Set Cover Problem* consiste en identificar el menor número de conjuntos cuya unión aún contiene todos los elementos del universo.

El programa DeFine desarrollado en Breuer and Nikoloski [2] describe el problema de individualización de filamentos como un problema de búsqueda de conjuntos de aristas en un grafo. En esta investigación, un filamento es representado por un conjunto de aristas, denominado como un camino. Los autores de esta investigación se basan inicialmente en un problema del tipo *Path Cover* (Ntafos and Hakimi [16]), extendiendo la definición de *Path Cover* mediante la asignación de un peso a cada segmento/arista del grafo, para calcular la *aspereza* o diferencia de homogeneidad en un camino. El cálculo del peso a lo largo de un camino es lo que permite individualizar filamentos, siendo el peso un reflejo del grosor o intensidad de una arista. Este problema particular es llamado *Filament Covering Problem* (FCP) y los autores demuestran que es NP-Hard, por lo que proponen un algoritmo de aproximación mediante *Set Cover*, cuyo objetivo es que cada arista pertenezca al menos a un camino (conjunto de aristas).

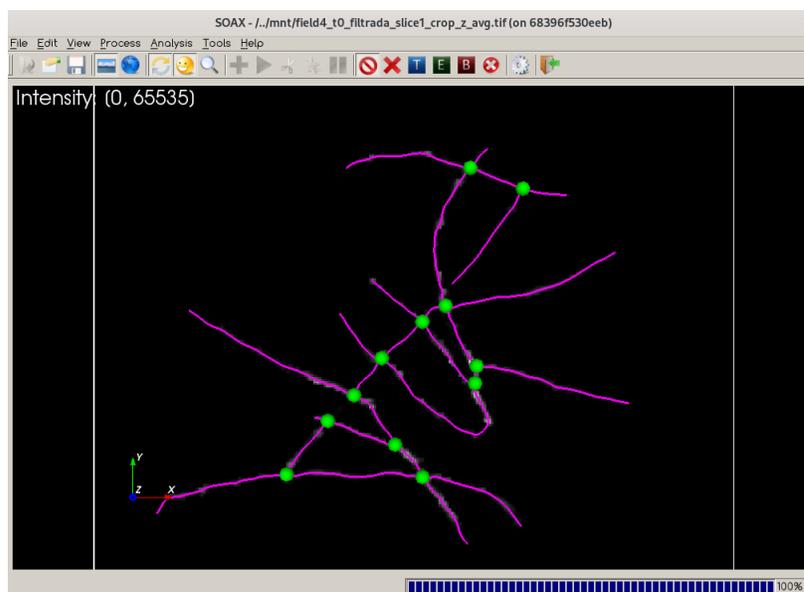


Figura 3.12: Resultado de SOAX al recortar las intersecciones de las curvas SOAC, tras aplicar la opción de corte de curvas paramétricas, y la posterior selección de la opción de agrupar curvas paramétricas, lo que posibilita observar las intersecciones en color verde. Fuente: Elaboración propia.

La adaptación a lo definido por *FCP* es:

Sea el universo U conformado por las aristas del grafo, y un conjunto S , conformado por conjuntos de aristas, cada uno con costo c_s , $s \in S$:

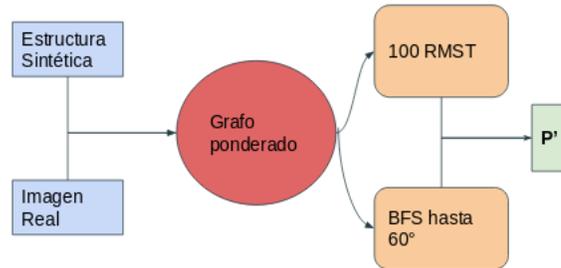
Encontrar un subconjunto $S_{set} \subseteq S$ con costo mínimo (o promedio, dependiendo de la forma en que se calcula la aspereza) tal que cada elemento en U este cubierto al menos una vez.

Los autores de DeFiNe demuestran que el FCP es NP-Hard en grafos, debido a la relación entre el número de nodos y el conjunto total de caminos posibles. Esta relación implica que al aumentar el número de nodos, el conjunto total de caminos en un grafo, definido como P , crece de forma exponencial. Para esta situación, los autores de DeFiNe prueban que el FCP en árboles es soluble en tiempo polinomial, basándose en Lin et al. [17]. La complejidad computacional para resolver el FCP en árboles es de $\mathcal{O}(n^4)$ para caminos que no comparten nodos o aristas, es decir, no se superponen. Para permitir caminos que se superponen se agrega el parámetro k que define el número máximo de superposiciones de caminos en una arista, quedando la complejidad computacional igual a $\mathcal{O}(n^{2k+2})$.

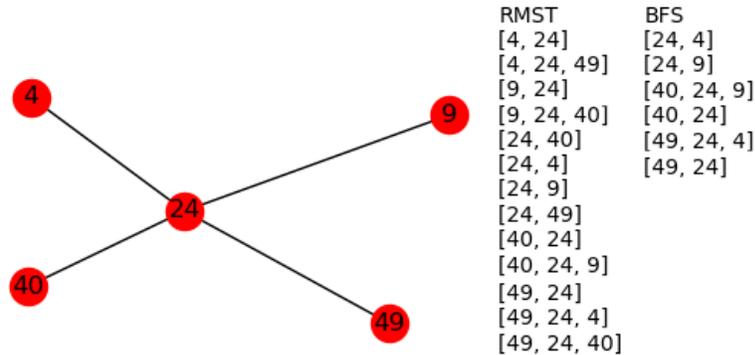
DeFine propone 2 heurísticas para construir árboles de los cuales se puedan extraer un subconjunto $P' \in P$ representativo de caminos simples: recorrer el grafo por su anchura (*Breadth-First Search*), deteniéndose al momento en que el ángulo de deflexión entre aristas adyacentes supere los 60° , o generar caminos a partir de 100 árboles de expansión mínima aleatoria (*RMST*), aportando cada uno con $N(N-1)/2$ caminos no triviales y sin dirección.

El subconjunto P' constituye los datos de entrada del problema de *Set Cover*, el que

se resuelve a través de un algoritmo de aproximación lineal fraccional binario (Breuer and Nikoloski [2]), en el que al subconjunto P' se le aplica la función objetivo, para encontrar los miembros de $p \in P'$ que mejor minimicen la diferencia de homogeneidad en sus caminos. La restricción de lo anterior es que cada arista del grafo pertenezca a lo menos a un camino. El flujo de decisiones para generar el subconjunto de caminos puede observarse en la Figura 3.13a.



(a) Entrada de datos y elección de subconjunto de caminos P' en DeFine.



(b) Subconjunto P' para el grafo de 5 nodos a la izquierda, utilizando la opción de 100 *RMST* o heurística de *BFS* que corta el camino al encontrar un ángulo de deflexión mayor a 60° entre aristas adyacentes.

Figura 3.13: (a) A partir de un grafo ponderado proveniente de una estructura sintética o de una imagen real, se elige P' entre los $N(N - 1)/2$ caminos no triviales y sin dirección que aporta cada uno de los 100 árboles de expansión mínima aleatoria (*RMST*), o los caminos resultantes de la heurística de búsqueda por anchura (*BFS*) con interrupción al dar con una arista que tenga un ángulo superior a 60° . (b) Subconjunto P' de P , para el grafo de 5 nodos de ejemplo a la izquierda. Fuente: Breuer and Nikoloski [2].

Cabe destacar que el *FCP* sólo utiliza 2 características independientes para describir los segmentos de los filamentos, siendo el ángulo de deflexión entre aristas usado en la etapa de selección de subconjuntos de caminos sólo si es seleccionada la heurística *BFS*, y el grosor o intensidad, empleado para describir el peso de las aristas. En el caso de la heurística de *RMST*, se asignan pesos aleatorios uniformemente distribuidos a las aristas, por lo que no hay uso de características asociadas a los filamentos.

En el caso de los métodos basados en optimización, la mayor crítica es su costo compu-

tacional, el que aumenta a medida que se complejiza, limitando en parte aquel enfoque. Se debe agregar que los parámetros utilizados por estas técnicas (ángulos o distancias máximas entre filamentos) son complejas de obtener de los expertos directamente. Sin embargo, una de sus ventajas es que automatizan la recuperación de información incluyendo una mayor cantidad de propiedades a cada arista.

En resumen, el problema de identificar filamentos en imágenes de microscopía está limitado por la resolución, y los problemas de múltiples parámetros a ajustar, para los métodos basados en procesamiento de imágenes de bajo nivel, el costo computacional en los métodos basados en optimización, y falta de descriptores cuantitativos en ambas. La revisión bibliográfica da cuenta también de pocas herramientas disponibles. Todo lo anterior implica que parte del análisis deba ser manual, lo que para grandes cantidades de datos, hace los estudios más propensos a errores.

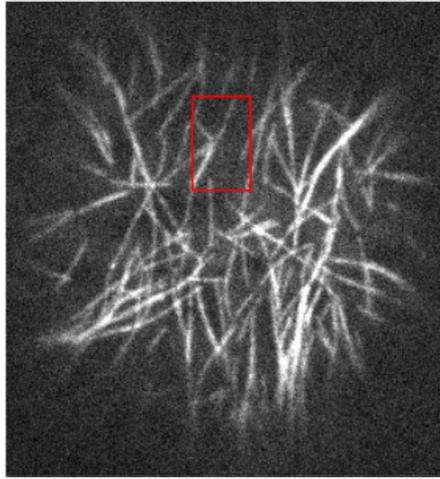
3.3. Generación de un grafo desde una imagen

El uso de grafos para la individualización de filamentos implica la necesidad de obtener el grafo a partir de una imagen para luego realizar su análisis. Algunas de las dificultades involucradas en la obtención de un grafo se relacionan al ruido y la resolución de la imagen. Un ejemplo de aquello se observa en la Figura 3.14a.

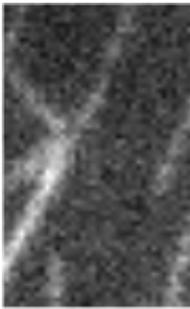
Mientras que el ruido ha sido estudiado en la literatura, el problema de resolución depende principalmente de la capacidad del microscopio que se utilice. La aproximación al límite máximo de resolución, denominado $\lambda/2$, determina el tamaño mínimo que 2 objetos que se encuentren juntos pueden tener para no observarse como un único elemento. Lo anterior sucede para algunos tipos de filamentos como los microtúbulos que pueden medir tan solo 25 nanómetros, lo que se encuentra por debajo de $\lambda/2$ para diversos microscopios.

Un problema derivado de la resolución de una imagen en la que se observan filamentos yace en que 2 expertos pueden disentir al realizar una individualización manual de filamentos, como se indica en las figuras 3.14c y 3.14d. Esta discrepancia implica que para algunos tipos de células donde se observen filamentos no es posible conocer a priori el origen y el final de un filamento. Adicionalmente, se presenta una dificultad en la representación de un filamento en un grafo, ya que esta se basa en un conjunto de aristas adyacentes, lo que lleva a tener un universo de hasta $n!$ posibles combinaciones en el espacio de soluciones. Cada conjunto de aristas adyacentes que representa a un filamento se denomina camino o recorrido.

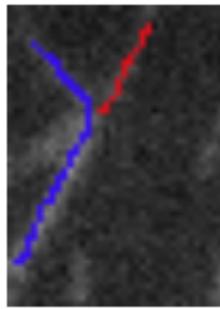
La extracción de un grafo a partir de una imagen, consiste en extraer un grafo $G = (V, E)$ de una imagen tal que G sea un grafo simple, no dirigido, ponderado, conectado o desconectado, con o sin ciclos. Esto implica que exista a lo más 1 arista por cada par de nodos adyacentes, prohibiéndose la existencia de nodos conectados consigo mismos. Es importante evitar que G sea un grafo completo, dado que al tener una arista por cada par de nodos G llega a tener $\frac{n(n-1)}{2}$ aristas, con n como el número de nodos. Se definen los vértices/nodos del grafo G como $V(G)$ y las aristas de G como $E(G)$. Que el grafo G sea ponderado implica que para las aristas del grafo ($E(G)$), existen características asociadas que se expresan como características geométricas, topológicas, espaciales y/u otras.



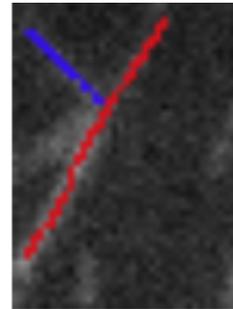
(a) Microtúbulos en planta *Marchantia*.
Fuente: Llanos [12].



(b) Sección resaltada en rojo de (a)



(c) Opción 1 de microtúbulos en (b)



(d) Opción 2 de microtúbulos en (b)

Figura 3.14: Dificultad de individualización que enfrentan los expertos al analizar manualmente una imagen de filamentos, en particular, microtúbulos. Fuente: Elaboración Propia.

La importancia del procedimiento de extracción o generación de un grafo que representa una red de filamentos a partir de una imagen radica en que define la cantidad de información disponible para llevar a cabo la individualización de filamentos. A partir de una imagen es posible obtener una cantidad de características de distinta índole, lo que permite en etapas posteriores clasificar de diversas formas nodos, aristas, de forma aislada o en conjuntos, efectivamente disminuyendo el espacio de búsqueda. Con las herramientas actuales disponibles en la literatura, es posible realizar la extracción de una red de filamentos con algún nivel de información como con la herramienta SOAX desarrollada por Xu et al. [11]. Sin embargo, las transformación de aquella red a un grafo, así como la incorporación de las características y/o propiedades hacia el grafo son un procedimiento no automatizado, por lo que el esfuerzo que el experto debe realizar es cercano a individualizar los filamentos de manera manual.

A partir de las investigaciones en la literatura, es posible agrupar los métodos para extraer la información que permite la construcción de un grafo a partir de una imagen, como lo son los nodos y las aristas, en dos conjuntos: los que se basan en eskeletonización (Lavado [18]) y los que no.

3.3.1. Extracción de un Grafo mediante Esqueletonización

Los métodos basados en esqueletonización consisten primariamente en la reducción de los píxeles pertenecientes al plano de interés o *foreground* en una imagen binaria, hasta formar una representación del objeto en la imagen de 1 píxel de ancho. El proceso debe mantener la conectividad del objeto adelgazado y a su vez, reducir la dimensión del objeto en la imagen para facilitar su análisis (Saha et al. [19]). Un análisis de los vecindarios de los píxeles del esqueleto construido es una de las formas más sencillas en que se puede distinguir si un píxel representa un nodo o si es parte de un arista. Una librería que realiza tal análisis es *skan* (Nunez-Iglesias et al. [20]), entregando estadísticas del grafo extraído como largo promedio de una rama del esqueleto (equivalente a una arista del grafo), tipo de rama, curvatura de una rama, entre otras mediciones. Sin embargo, el formato de salida del grafo para esta herramienta corresponde a *Compressed sparse row* o CSR, lo que causa que un análisis de mayor profundidad o el paso del grafo a una herramienta de individualización de filamentos necesite de una librería adicional.

Otra herramienta que realiza un análisis similar para obtener un grafo a partir de un esqueleto es *sknw*, parte del framework *ImagePy* (Wang et al. [21]). La diferencia propuesta por *sknw* radica en que se integra con la librería *NetworkX* (Hagberg et al. [22]), utilizando la estructura de datos para grafos que esta última posee para elegir entre múltiples formatos de salida. Aquello otorga flexibilidad en la integración de herramientas que utilizan como base el grafo para realizar análisis posteriores, como es el caso de la individualización de filamentos.

Independiente de la herramienta usada para obtener la información topológica de la estructura observada, el procedimiento de esqueletonización puede sufrir de pérdida de información, dado que mediante el o los pasos de adelgazamiento, puede perderse la relación entre los píxeles que conforman el esqueleto obtenido y los que son sus vecinos previo al adelgazamiento. Lo anterior puede dificultar el cálculo de características geométricas como el ancho, así como puede impedir la obtención de información mediante métodos como *image moments* (Flusser et al. [23], Chaumette [24]).

Una forma de obtener la información de vecindario mencionada es siguiendo la idea de agrupación de píxeles similares o cercanos utilizada en superpíxeles (Achanti et al. [25]). Los algoritmos de generación de superpíxeles juntan conjuntos de píxeles para obtener nuevas unidades atómicas sobre las cuales se realiza el análisis, reduciendo la complejidad de este. El propósito de basarse en la idea de superpíxeles y no en el concepto de forma completa se fundamenta en que no se busca generar una segmentación acabada, sino que poder asociar los superpíxeles con los nodos y aristas que resultan al transformar el esqueleto a un grafo. Para aquello, es sólo necesario realizar una agrupación gruesa utilizando criterios de tamaño máximo de un superpíxel, dado que en imágenes binarias o en escala de grises el uso de criterios de similitud según el valor de cada píxel puede llevar a obtener un número muy bajo de superpíxeles, ocasionando que múltiples nodos y/o aristas se refieran al mismo superpíxel. El detalle de la implementación de la agrupación de píxeles y su asociación con los nodos se encuentra en la sección A.1.

3.3.2. Obtención de Información Adicional

Para dar uso a la información recuperada de acuerdo a lo expresado en la sección 3.3.1, se analizaron diversos filtros útiles para describir estructuras alargadas, como *Gabor*, *Anisotropic Diffusion* y Frangi para *Vesselness*. El filtro de Gabor es un filtro sensible a la orientación, facilitando la extracción de características así como el análisis de texturas. Por su parte, *Anisotropic Diffusion* es un filtro que apunta a reducir el ruido en una imagen, y al mismo tiempo resalta los bordes de los objetos. El filtro Frangi para *Vesselness* (Frangi et al. [26], Fu et al. [27]), cuantifica cuan alargada es una estructura (*vesselness value*), en base a los valores y vectores propios de la matriz Hessiana (ecuación 3.3) posterior a la aplicación de uno o varios filtros Gaussianos para suavizar una imagen. Este filtro es utilizado en la detección de estructuras alargadas como arterias y venas, pudiendo replicarse parcialmente mediante el análisis de píxeles con *image moments* (Flusser et al. [23]). A diferencia del filtro de Gabor y de *Anisotropic Diffusion*, el filtro de Frangi no requiere de configurar parámetros para su uso, lo que lo hace más simple y lleva a su elección en este trabajo por sobre los filtros.

$$H = \begin{bmatrix} H_{xx} & H_{xy} \\ H_{xy} & H_{yy} \end{bmatrix} \quad (3.3)$$

Una respuesta de *vesselness value* que denota una estructura alargada se obtiene si los 2 valores propios, λ_1 y λ_2 ($|\lambda_2| \geq |\lambda_1|$) satisfacen $|\lambda_1| \approx 0$ y $|\lambda_2| \gg |\lambda_1|$. Los valores propios se obtienen mediante la ecuación 3.4.

$$\lambda_{1,2} = \frac{(H_{xx} + H_{yy}) \pm \sqrt{(H_{xx} - H_{yy})^2 + 4 \cdot H_{xy}^2}}{2} \quad (3.4)$$

Otra forma de obtener los valores propios de la ecuación 3.4 es utilizando los *central image moments* o momentos centrales, que son un tipo de *image moments*. Los momentos centrales se derivan del tipo más básico de *image moments* que son los *raw image moments*. Se define un *raw image moments* de orden $p + q$ para una imagen en la ecuación 3.5a, donde $f(x, y)$ corresponde a la intensidad de la imagen en un punto (x,y). El *raw image moments* M_{00} refleja la “masa” de la imagen, correspondiendo al área o volumen si se trata de una imagen binaria.

Para el cálculo de los momentos centrales se agregan los componentes del centroide, \bar{x} e \bar{y} , basados en los *raw image moments*, como indican las ecuaciones 3.5b y 3.5c.

$$M_{pq} = \sum_x \sum_y x^p \cdot y^q \cdot f(x, y) \quad (3.5a)$$

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}} \quad (3.5b)$$

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \cdot f(x, y) \quad (3.5c)$$

Así, es posible construir una matriz de covarianza, equivalente a la matriz hessiana en la ecuación 3.3, utilizando los momentos centrales de segundo orden, μ_{20} , μ_{02} y μ_{11} divididos por el momento central de orden cero μ_{00} (ecuaciones 3.6a, 3.6b y 3.6c), obteniendo los valores propios mediante la ecuación 3.6e.

$$\mu'_{20} = \frac{\mu_{20}}{\mu_{00}} = \frac{M_{20}}{M_{00}} - \bar{x}^2 \quad (3.6a)$$

$$\mu'_{02} = \frac{\mu_{02}}{\mu_{00}} = \frac{M_{02}}{M_{00}} - \bar{y}^2 \quad (3.6b)$$

$$\mu'_{11} = \frac{\mu_{11}}{\mu_{00}} = \frac{M_{11}}{M_{00}} - \bar{x} \cdot \bar{y} \quad (3.6c)$$

$$cov[f(x, y)] = \begin{bmatrix} \mu'_{20} & \mu'_{11} \\ \mu'_{11} & \mu'_{02} \end{bmatrix} \quad (3.6d)$$

$$\lambda_{1,2} = \frac{(\mu'_{20} + \mu'_{02}) \pm \sqrt{(\mu'_{20} - \mu'_{02})^2 + 4 \cdot \mu'^2_{11}}}{2} \quad (3.6e)$$

Con los valores propios, es posible calcular características de una estructura alargada como su excentricidad o su eje principal de inercia. Estas medidas pueden ayudar a mejorar la clasificación de segmentos del grafo durante la identificación de filamentos.

Una manera adicional de generar información que facilite la discriminación de secciones del grafo es a través del cálculo de los ángulos entre las aristas del grafo. Esto se relaciona al criterio de rectitud que tienen los filamentos, que varía dependiendo de la célula a la que pertenezca. Este comportamiento de los filamentos permite delimitar el ángulo máximo que 2 aristas adyacentes pueden tener para ser considerados parte del mismo filamento. Cualquier valor por sobre este ángulo máximo permitiría descartar de forma absoluta esa combinación de aristas para un mismo filamento.

Finalmente, para el caso de los nodos, el análisis del grado de cada uno permite identificar la existencia de ciclos (ver Wilson [28]) en un filamento. La propiedad de un filamento de poder tener o no tener un ciclo es información disponible a priori que depende del tipo de célula observada, permitiendo limitar posibles asociaciones entre nodos. En el caso particular de no permitir ciclos, un filamento no podría pasar más de una vez por cada nodo que lo conforma.

3.4. Metaheurística ACO

Uno de los problemas principales de la individualización de filamentos recae en desconocer el comienzo y el final de los filamentos. Lo anterior resulta similar al comportamiento de las hormigas al explorar caminos en búsqueda de alimento, donde se desconoce el lugar en el que se encuentra la comida. Por esto resulta natural utilizar la metaheurística ACO para efectuar

la exploración y elección de filamentos. La metaheurística *Ant Colony Optimization* (ACO) se inspira en el comportamiento de una colonia de hormigas en la búsqueda del trayecto más corto a una fuente de alimento, comunicándose entre ellas mediante feromonas. El recorrido de una hormiga se define como un camino s y representa una solución en la búsqueda del camino más corto.

Las hormigas realizan una exploración aleatoria alrededor de su hormiguero en busca de alimento. Una vez que lo encuentran, marcan el camino de retorno depositando una cantidad de feromonas, la que varía dependiendo de la calidad del camino. Para caminos de buena calidad, que son los de menor distancia entre un hormiguero y una fuente de alimento, las feromonas sirven de guía para otras hormigas. Así, las hormigas se comunican de forma indirecta, convergiendo en los caminos que tienen una mayor cantidad de feromonas. Este comportamiento se define como el modelo de feromonas y se observa en la Figura 3.15.

En este modelo, se define un camino como una solución s que consiste en un conjunto de componentes de solución c_i , por lo que una concatenación de componentes de solución forma el camino que recorre una hormiga. Cada componente c_i tiene asociado un valor de feromona τ_i , la cual influye en la elección que realiza una hormiga en relación a el o los componentes de solución disponibles para avanzar durante la construcción de un camino. La metaheurística ACO permite encontrar una única solución o un conjunto de soluciones.

El conjunto de caminos que recorren las hormigas, para ir del hormiguero hacia la fuente de alimento y de retorno, puede ser interpretado como un grafo que representa esta red de caminos. En esta representación, la búsqueda de caminos corresponde a encontrar conjuntos de nodos o aristas adyacentes, y una componente de solución c_i corresponde a una arista, siendo el recorrido desde una arista inicial hasta una arista final. La construcción de una solución utiliza hormigas artificiales, también llamadas agentes, que se desplazan a través del grafo mediante aristas adyacentes. La elección de aristas en cada paso de la construcción utiliza la información de las feromonas de cada una de las aristas candidatas, formando la solución incrementalmente. La metaheurística ACO, indicada en el Algoritmo 1, consiste en 4 etapas, donde las 3 últimas no tienen un orden específico.

Algoritmo 1: Algoritmo metaheurística ACO

```

Ajuste de Parámetros & inicialización de feromonas;
mientras Criterio de finalización no se cumple hacer
    | Construcción_de_solución_de_cada_hormiga();
    | Método_de_búsqueda_no_local(); //Paso opcional
    | Actualización_de_feromonas();
fin

```

3.4.1. Solución de un modelo COP mediante la metaheurística ACO

Un problema de optimización con restricciones, (COP, Constrained Optimization Problem) puede ser representado como $P = (S, \Omega, F)$, donde S es el espacio de soluciones, Ω son las restricciones, y F es la función objetivo. Si está definido por un conjunto discreto de variables $X = 1 \dots n$, con valores $v_i^j \in D_i = \{v_i^1 \dots v_i^{|D_i|}\}$. Se define como una variable *instanciada* la asignación a X_i de un valor $v_i^j \in D_i$. Una solución candidata $s \in S$ es una

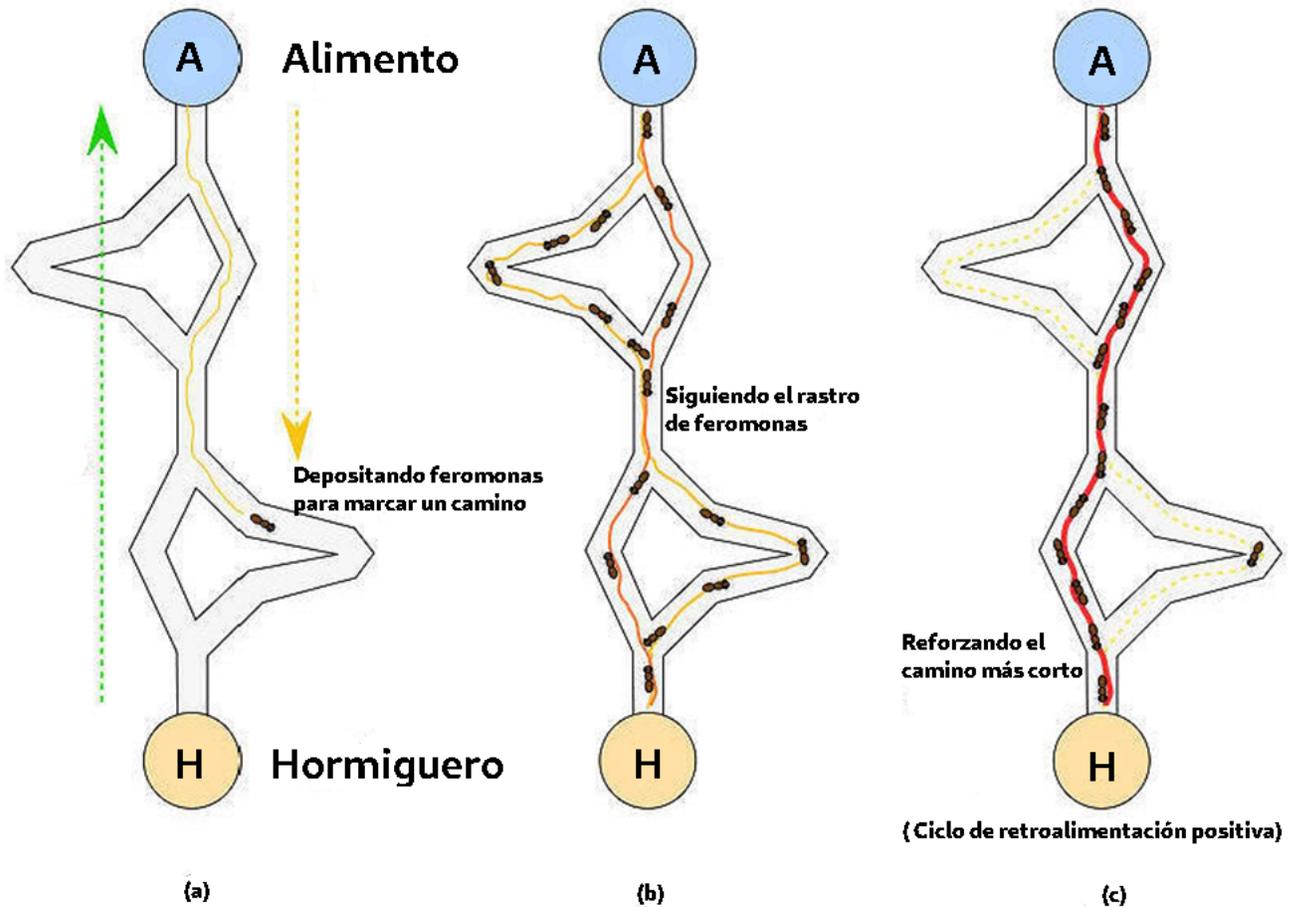


Figura 3.15: Etapas de búsqueda de alimento y de comunicación entre hormigas. (a) Flecha verde refleja la dirección de la búsqueda aleatoria, mientras que la flecha amarilla indica como una hormiga va depositando feromonas. (b) Otras hormigas siguen rastros de feromonas existente, aportando con sus propias feromonas. (c) Convergencia de las hormigas sobre el camino más corto en base a la cantidad de feromonas depositada por múltiples hormigas mediante un ciclo de retroalimentación positiva. Fuente: Liu et al. [29].

solución factible si satisface las restricciones del conjunto Ω . La función objetivo $F : S \rightarrow \mathbb{R}_0^+$, es la función de evaluación que asigna una puntuación a las soluciones candidatas. Al mismo tiempo, se define s^* como una solución óptima y S^* como el conjunto que engloba todas las soluciones óptimas, dado que pueden existir múltiples soluciones s^* , y se relacionan mediante $s^* \in S^* \subseteq S$ (ver Socha and Dorigo [30]).

Es posible adaptar la definición del modelo COP al modelo de feromonas de la metaheurística ACO, mediante la asociación entre la definición de variable *instanciada* del modelo COP y lo que representa un componente de solución o arista en un recorrido de una hormiga. Específicamente, si se extiende la definición de componente de solución c_i para que esta pueda tener valores v_i^j que varíen dependiendo de un dominio D_i , el componente de solución pasa a ser c_{ij} . Así se obtiene la equivalencia entre la asignación $X_i = v_i^j$ que representa una variable *instanciada* del modelo COP, y la selección de un componente de solución c_{ij} , para una solución o camino s en ACO (Socha and Dorigo [30]).

Luego, la definición de una solución s puede representarse como un conjunto de componentes de solución $c_{ij} \in C, i = 1 \dots n, j = 1 \dots |D_i|$. A su vez, la feromona asociada a una componente de solución se transforma de τ_i a τ_{ij} . El Algoritmo 1 se modifica para representar la equivalencia entre el modelo COP y la metaheurística ACO, añadiéndose las definiciones de los datos y el resultado esperado, quedando como lo indica el algoritmo 2.

Algoritmo 2: Algoritmo de un modelo COP adaptado a una metaheurística ACO

Datos: Variables $X_1 \dots X_n$, dominios $D_1 \dots D_n$, Restricciones $\in \Omega$

Resultado: conjunto $s' \subseteq S \neq \emptyset$, si existen soluciones factibles

Ajuste de Parámetros & inicialización de feromonas ;

mientras *Criterio de finalización no se cumple* **hacer**

 | Construcción_de_solución_de_cada_hormiga();

 | Método_de_búsqueda_no_local(); //Paso opcional

 | Actualización_de_feromonas();

fin

Capítulo 4

Algoritmo Propuesto

En base a lo recopilado en los capítulos previos de esta investigación, es posible destacar los siguientes aspectos del problema a resolver:

- Se desconoce a priori el número de filamentos a buscar.
- Generalmente se busca individualizar más de un filamento por imagen, lo que conlleva a elegir los mejores filamentos entre las soluciones que se encuentren.
- El uso de un grafo para representar la red de filamentos puede implicar que las combinaciones de soluciones crezcan de manera exponencial.

Lo anterior permite clasificar el problema de identificar filamentos a partir de un grafo como un problema de optimización con restricciones (Blum et al. [31]).

Se ha establecido en la sección 3.2 que una red de filamentos puede ser representada mediante un grafo, de forma similar a lo que sucede con la representación del conjunto de caminos que las hormigas recorren en búsqueda de alimento. En particular, un filamento en un grafo corresponde a un conjunto de aristas adyacentes, con lo que se puede asociar la individualización de un filamento a la elección de una o más aristas o componentes de solución, tal como se lleva a cabo en la metaheurística ACO. A su vez, durante la elección de las aristas adyacentes que permiten individualizar un filamento, se puede utilizar una o más características que las aristas o componentes de solución poseen, para dirimir entre las aristas a elegir durante este proceso. Este comportamiento es similar al que las hormigas realizan utilizando las feromonas. Lo anterior constituye el fundamento para utilizar la metaheurística ACO en el proceso de construcción y evaluación de soluciones que permiten individualizar filamentos.

La búsqueda de conjuntos de aristas adyacentes para individualizar uno o más filamentos implica examinar un espacio de soluciones que no es posible de recorrer en tiempo polinomial, dado que sin restricciones las combinaciones crecen exponencialmente (ver Buchin et al. [32], Biswas et al. [33]). La metaheurística ACO permite en sus distintas etapas (Construcción de solución, búsqueda no local, actualización de feromonas) incorporar información que puede acotar el espacio de búsqueda. En el caso de la individualización de filamentos, las diversas características asociadas a las aristas, así como las características que definen el

comportamiento global de un filamento permiten realizar esta tarea.

Un diagrama resumido de los pasos del algoritmo propuesto se presenta en la Figura 4.1. A partir de una imagen, se obtiene un grafo que representa a la red de filamentos mediante alguna herramienta externa, como *sknw*. Opcionalmente, información adicional obtenida a partir de la imagen o del grafo puede ser añadida al algoritmo. Lo anterior, en conjunto con los parámetros de entrada constituyen los datos con los que se puede inicializar y ejecutar la metaheurística ACO, obteniendo una individualización de filamentos. El detalle de cada etapa de la metaheurística ACO se presenta en las siguientes secciones de este capítulo.

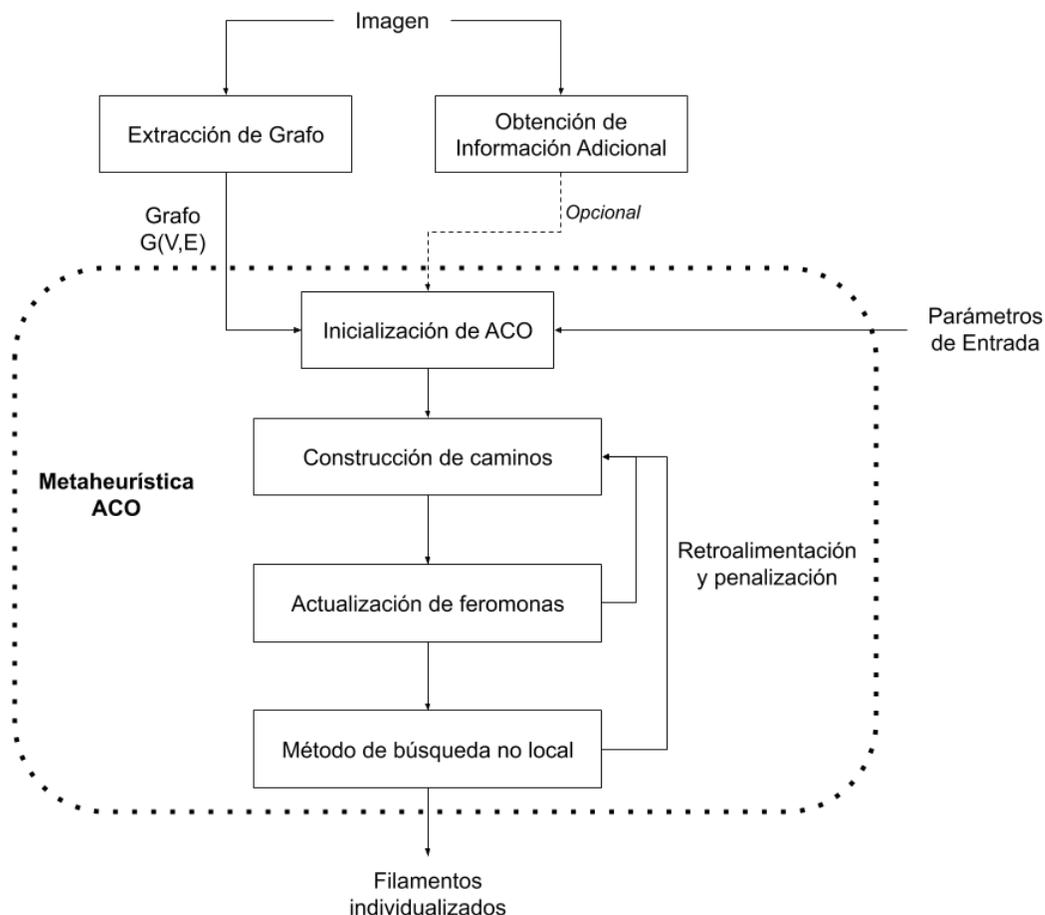


Figura 4.1: Diagrama simplificado de las etapas del algoritmo propuesto. Fuente: Elaboración Propia.

4.1. Inicialización de la metaheurística ACO

En el paso de inicialización de ACO se deben definir los valores de los parámetros relacionados a las heurísticas y de las feromonas utilizadas. Una de las características más relevantes para discriminar aristas en el proceso de construcción corresponde al ángulo que forman 2 aristas adyacentes. El primer umbral que permite definir si ambas aristas pertenecen al mismo filamento se define como θ . Así, si el ángulo se encuentra en el rango $[0, \theta]$ se puede dirimir que ambas aristas pertenecen al mismo filamento. Un segundo umbral se

define como Max_Angle , que indica el límite por sobre el cual se puede inferir que estas 2 aristas no pertenecen al mismo filamento. Estos parámetros dependen de la estructura observada, asignando a θ valores de 30° para microtúbulos de planta y de 45° para neuronas. Max_Angle depende de θ al definirse como el máximo entre 2.5 veces θ y 90° . El factor 2.5 es un valor obtenido experimentalmente que no ha sido sometido a sintonización.

Otros parámetros a utilizar corresponden a $Max_Axial_Displacement$ y a Max_Score . El primero hace referencia a un parámetro de tolerancia que influye en 2 aspectos: al evaluar la curvatura general de una solución, y al evaluar la bifurcación entre 2 grupos de aristas al interior de la solución. La diferencia de la evaluación de bifurcación con respecto a la medición de ángulos entre aristas radica en que este criterio también considera la magnitud de estos grupos de aristas. Los valores de $Max_Axial_Displacement$ son de 1.5 para el caso de los microtúbulos de planta o de 2.5 para las neuronas. Por su parte, Max_Score define el puntaje máximo a obtener por un camino de buena calidad. Su valor durante este trabajo corresponde a 2, y es utilizado por la heurística miope (ecuación 4.2), que se define más adelante en la sección 4.2. A su vez, los parámetros α y β también utilizados en la heurística miope, que regulan la ponderación entre esta heurística y las feromonas, son fijados en 1.

En el caso de las feromonas, se define un valor inicial de $\tau_{ij} = 1$, pero utilizando un concepto inverso de feromona con respecto al que se presentó previamente. El cambio consiste en el uso de *anti-feromonas* o SAP (*Subtractive Anti-Pheromone*), definido por Montgomery and Randall [34], que se explica en la sección 4.3 y consiste en disminuir el valor de la feromona asociada a su respectiva arista. El uso de SAP introduce el parámetro γ , el que se define en 0.5 en base a lo encontrado en la literatura.

En cuanto al criterio de finalización, este consiste en generar nuevas hormigas artificiales hasta que todas las aristas sean parte de al menos una solución de buena calidad, o que el número de hormigas generadas sea superior a 4 veces la cantidad de aristas. En lo que sigue de este trabajo se utiliza el término hormigas en referencia a las hormigas artificiales o agentes.

El grafo que representa la red de caminos que pueden recorrer las hormigas se puede obtener mediante las herramientas descritas en la sección 3.3.1, siendo *sknw* la seleccionada en esta investigación. Esta herramienta recibe como dato de entrada la imagen de microscopía binarizada, realizando internamente el proceso de esqueletonización y posterior transformación a una estructura de grafo.

4.2. Método de construcción de caminos de cada hormiga

Al comenzar un recorrido, cada hormiga es asignada una arista de acuerdo a la heurística de asignación inicial, la que corresponde al primer elemento del camino o recorrido parcial s^P . La heurística de asignación inicial analiza hasta 3 situaciones, dependiendo del tipo de célula:

1. La arista a asignar debe tener al menos uno de sus nodos con grado 1, indicando que es el inicio o final de una parte del grafo.
2. De no haber aristas disponibles con esas características, se realiza una asignación inicial

de una arista que cumpla con 2 criterios:

- Tener uno de sus nodos con grado 2 o superior.
 - El ángulo que forma la arista candidata a elegir, junto a otra arista a la que pertenece el nodo, debe pertenecer al rango $]\theta, Max_Angle]$.
3. De no existir aristas con algún nodo que cumpla con los dos criterios previos, es posible asignar una arista aleatoria. Esta arista no debe pertenecer al una solución o camino evaluado como de buena calidad. La calidad de un camino se presenta más adelante en esta sección.

Una vez asignada la primera arista según la heurística previamente descrita, cada hormiga debe avanzar mediante la elección de nuevas aristas para añadirlas a su recorrido. Este procedimiento corresponde a la probabilidad de elegir una arista o componente c_{ij} a partir de un conjunto de aristas vecinas $N(s^P)$. Las aristas que pertenecen a $N(s^P)$ son solo las aristas adyacentes a la última arista agregada a la solución parcial s^P que son factibles de agregar a la solución parcial. La factibilidad de añadir una arista de $N(s^P)$ a s^P corresponde a la probabilidad expresada en la ecuación 4.1. Esta probabilidad depende del valor de la feromona τ_{ij} asociada a la arista c_{ij} , así como del valor de la heurística miope η_{ij} (ecuación 4.2) que evalúa el ángulo formando entre la arista $c_{ij} \in N(s^P)$ y la última arista agregada a s^P , que se define como s_n^P .

En particular, la heurística miope η_{ij} entrega una puntuación que disminuye a medida que el ángulo mencionado aumenta, siendo el rango $[0, \theta]$ el que otorga la puntuación máxima de Max_Score . Si el ángulo es mayor a Max_Angle la puntuación corresponde a 0. Para el rango intermedio $]\theta, Max_Angle]$, la puntuación depende de la diferencia entre el ángulo formando entre la arista $c_{ij} \in N(s^P)$ y s_n^P . Mientras más aumenta esta diferencia, menor es la puntuación asignada.

$$P(c_{ij}|s^P) = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{c_{ij} \in N(s^P)} \tau_{ij}^\alpha \eta_{ij}^\beta}, \forall c_{ij} \in N(s^P). \quad (4.1)$$

La evaluación de ángulos entre aristas adyacentes que pertenezcan al rango intermedio $]\theta, Max_Angle]$ en la heurística miope, facilita la exploración de soluciones o caminos que de forma miope no califican como soluciones candidatas factibles de representar un filamento, pero que tampoco pueden ser descartados completamente. En el caso que un camino tenga uno o más pares de aristas adyacentes cuyo ángulo pertenezca a este rango, se definirá el camino como de *calidad intermedia*.

$$\eta_{ij} = \begin{cases} Max_Score & \text{si } \angle(c_{ij}, s_n^P) \in [0, \theta], \\ Max_Score \cdot \left(1 - \frac{|\angle(c_{ij}, s_n^P) - \frac{\theta}{2}|}{180}\right) & \text{si } \angle(c_{ij}, s_n^P) \in]\theta, Max_Angle], \\ 0 & \text{en otro caso.} \end{cases} \quad (4.2)$$

Con la finalidad de cuantificar la calidad de la solución construida por una hormiga, en cada selección de arista realizada, se suma el resultado de la heurística miope η_{ij} a la calidad del camino que la hormiga lleva hasta ese punto. Cada hormiga comienza con una calidad 0, que al finalizar su recorrido es dividida por el número de aristas menos 1, para normalizar el valor entre las distintas soluciones. Se establece que para ser consideradas como una solución de *buena calidad*, la hormiga debe tener una calidad mayor o igual a $Max_Score/2$. Es importante destacar que la calidad de una solución puede variar en cada etapa de la metaheurística ACO, pudiendo descartarse una solución en base a los criterios de cada método. El desechar una solución implica ajustar su calidad a 0, lo que la clasifica como de *mala calidad*. Las hormigas que en cada elección de arista hayan elegido componentes c_{ij} que formaban un ángulo en el rango $[0, \theta]$ con s_n^P , son las que obtienen el máximo puntaje de Max_Score al normalizarse, y se definen como de *calidad óptima*. Los caminos clasificados como de *calidad intermedia* son inicialmente parte del grupo de soluciones de buena calidad.

4.3. Método de actualización de feromonas

Una vez que la hormiga termina un recorrido, se evalúa su calidad para actualizar las feromonas τ_{ij} asociadas a los componentes de solución que lo conforman. Así, en una metaheurística ACO tradicional, se aumenta el valor de τ_{ij} en cada c_{ij} que es parte de un camino de buena calidad. Adicionalmente, los valores de las feromonas sufren decaimiento en el tiempo, dado por el parámetro ρ , que busca evitar la convergencia que las feromonas pueden causar en caminos de buena calidad obtenidos al inicio de las iteraciones.

En la individualización de filamentos, se utilizan *anti-feromonas*, con el propósito de indicar a las hormigas de futuras iteraciones los caminos que no son de buena calidad. La diferencia radica en que las anti-feromonas buscan acotar el espacio de soluciones S sin necesariamente buscar la convergencia sobre un único camino. Este cambio implica no utilizar el parámetro ρ , ya que no existe decaimiento en este uso de anti-feromonas, y reemplazarlo por el parámetro γ como factor de reducción/penalización (Montgomery and Randall [34]). Al ser la penalización un factor que se aplica sobre el valor de τ_{ij} , la anti-feromona puede disminuir constantemente sin llegar a valer exactamente cero, por lo que se define un límite de 2 penalizaciones como máximo para cada anti-feromona. Al alcanzar este límite, se reduce el valor de τ_{ij} a 0, haciendo imposible la elección de la arista c_{ij} respectiva.

Además de cambiar el uso de feromonas por el de anti-feromonas, se introduce una segunda modificación, relativa a las aristas a las que aplica la penalización de las anti-feromonas. Un uso tradicional de anti-feromonas implica la disminución del valor de τ_{ij} para su respectiva arista c_{ij} , en caso de que se trate de un camino de mala calidad. Esto puede llevar a la pérdida de exploración de soluciones en la individualización de filamentos, ya que una arista con un τ_{ij} penalizado por pertenecer a uno o más caminos de mala calidad, puede bloquear otros caminos, dado que el valor de τ_{ij} no guarda relación con el conjunto de aristas que causa la penalización. Así, las soluciones que pudiesen pasar por esa arista se ven limitadas.

En este trabajo se propone que la penalización de τ_{ij} guarde relación con el conjunto de aristas en el camino s que preceden a la arista c_{ij} . Se define a aquel conjunto de aristas como un segmento de un camino recorrido por la hormiga a , expresado como seg_n^a . Los segmentos se generan durante la construcción de un recorrido, existiendo siempre al menos un segmento en

cada camino. Dado que el ángulo entre 2 aristas adyacentes es uno de los criterios utilizados durante la construcción de un camino, el fin de un segmento puede suceder en 2 casos: si la última arista en el camino corresponde al final del recorrido, o si la arista s_n^P (la última arista agregada a s^P) forma un ángulo en el rango $]\theta, Max_Angle]$ con la arista $c_{ij} \in N(s^P)$ a agregar a la solución parcial.

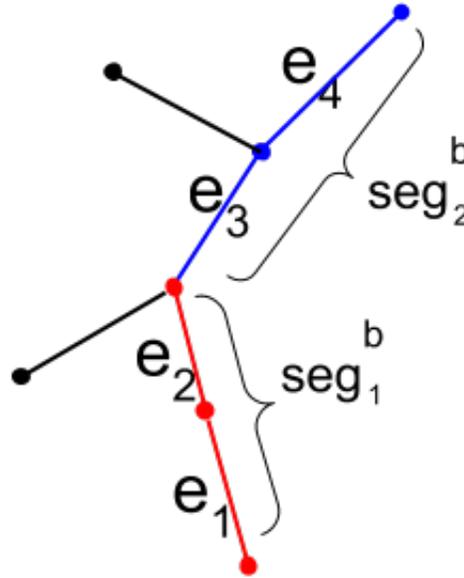


Figura 4.2: Camino de una hormiga b compuesto por 2 segmentos. El ángulo entre las aristas e_2 y e_3 pertenece al rango $]\theta, Max_Angle]$, finalizando el segmento seg_1^b e iniciando el segmento seg_2^b . Fuente: Elaboración Propia.

Así, cada segmento está formado por una o más aristas, donde cada arista del segmento forma un ángulo en el rango $[0, \theta]$ con sus vecinos. Un ejemplo de esto se puede observar en la Figura 4.2, en el que si se penaliza la arista e_3 , esta penalización será asociada al segmento seg_1^b , disminuyendo la posibilidad de que futuras hormigas que recorran este segmento escojan e_3 , sin perjuicio de otros caminos que pasen por e_3 pero no por el segmento seg_1^b .

Una ventaja de utilizar segmentos de camino es evitar volver a evaluar la relación entre todas las aristas que conforman el camino, ya que las aristas que conforman un segmento cumplen con un criterio para ser consideradas como parte del mismo filamento. Así, la evaluación de la calidad del camino se realiza solo en el caso que 2 aristas vecinas formen un ángulo de rango intermedio ($]\theta, Max_Angle]$), enfocándose en las soluciones de calidad intermedia que permiten explorar el espacio de soluciones.

Se debe señalar que esta propuesta de penalización realiza la evaluación de la calidad del camino en orden inverso al utilizado durante la construcción del recorrido, imitando el retorno que realiza una hormiga desde el alimento hacia su hormiguero. Esto, sumando al uso de segmentos, permite comenzar la revisión en el último par de aristas cuyo ángulo pertenece al rango intermedio, es decir, el nodo donde termina el penúltimo segmento y comienza el último. A modo de ejemplo, en la Figura 4.2, esto sería el nodo común de las aristas e_2 y e_3 , que separa los segmentos seg_1^b y seg_2^b . En este ejemplo, en caso de penalizar la arista e_3 , será solo disminuyendo su posibilidad de ser elegida para caminos que contengan al segmento seg_1^b en su solución.

Por último en este aspecto, y a diferencia del uso tradicional de feromonas o anti-feromonas, donde se penalizan todas las aristas pertenecientes a una solución, el recorrido inverso y el uso de segmentos permite que sea suficiente la penalización de una arista en la solución para clasificar la solución s como de mala calidad y desecharla. Esto permite que a partir de una solución de mala calidad se remueva el último segmento, permitiendo a otra hormiga recorrer un camino similar con un distinto final, favoreciendo la exploración. Para visualizar esto, es posible utilizar la Figura 4.3, en donde una hormiga pudiese construir un camino incorporando las aristas e_1 , e_2 , e_4 , e_5 y e_6 . Al desechar el último segmento, conformado solo por la arista e_6 , permite que otra hormiga recorra desde e_1 hasta e_5 , pudiendo agregar e_7 a su camino o terminando en e_5 .

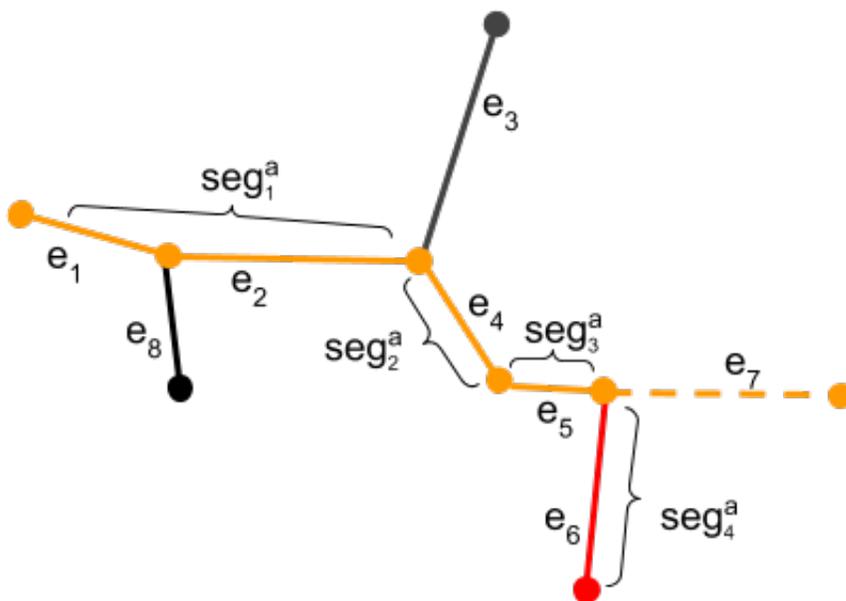
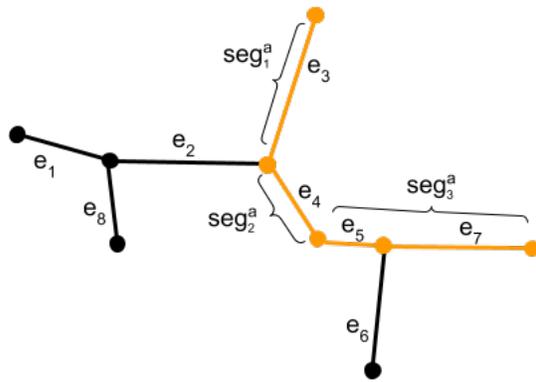


Figura 4.3: Camino de una hormiga a compuesto por 4 segmentos. El comienzo en orden inverso a la construcción favorece la exploración, reemplazando el último segmento por otro o simplemente concluyendo el recorrido. Fuente: Elaboración Propia.

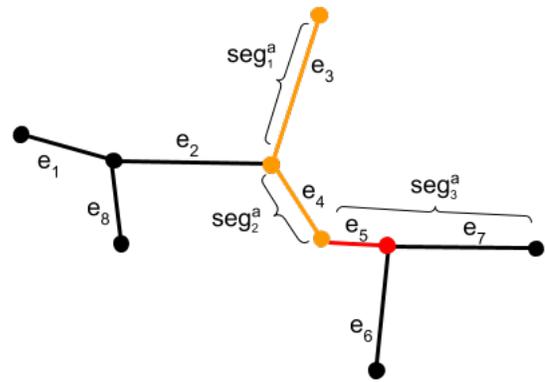
Lo anterior permite definir que la penalización de τ_{ij} este asociada a la arista c_{ij} y al segmento que precede a c_{ij} en el recorrido de una hormiga a . El segmento que precede a c_{ij} en un camino s se define como seg_{prev}^a . Adicionalmente, se tiene que entre distintos caminos pueden existir segmentos equivalentes, lo que sucede en el caso de que contengan las mismas aristas. Esto permite que se refieran al mismo valor τ_{ij} a pesar de los nombres relativos al camino al que pertenecen. El funcionamiento del método de penalización propuesto se visualiza en la Figura 4.4.

Un caso particular en el que esta propuesta puede sufrir del mismo problema de bloqueo de soluciones, es al momento que un segmento formado por solo una arista se encuentre entre 2 segmentos. Este situación, reflejada en la Figura 4.5, implica que la penalización que involucre un segmento de una sola arista, como el segmento seg_2^a , puede generar un bloqueo para los recorridos de otras hormigas. Este tipo de segmentos puede generarse a partir de aristas que forman un ángulo que pertenece al rango $[\theta, Max_Angle]$ con cada uno de sus vecinos, como es el caso de la arista e_4 en el ejemplo.

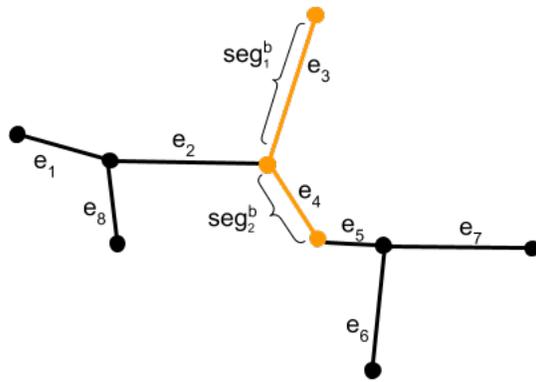
Para subsanar aquel caso, a los segmentos de solo una arista se le añaden las aristas del segmento que lo precede con el objetivo de evitar que el análisis sea solo entre 2 aristas. Así, el segmento seg_2^a pasa a ser formado por las aristas e_3 y e_4 , evitando el bloqueo de la hormiga b .



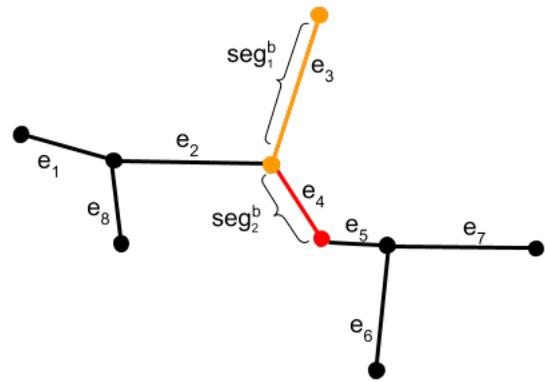
(a) Camino de una hormiga a que contiene las aristas e_3 , e_4 , e_5 y e_7 .



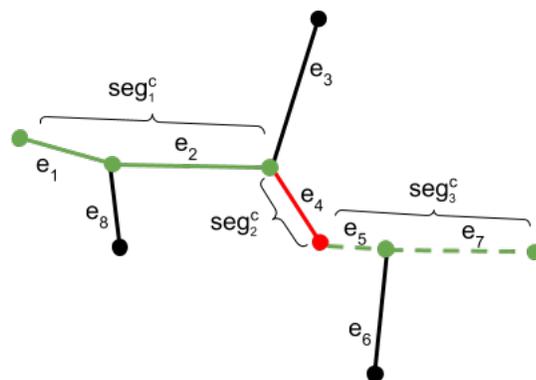
(b) Penalización del camino de la hormiga a en la arista e_5 con respecto al segmento seg_2^a conformado solamente por la arista e_4 .



(c) Camino de una hormiga b que contiene las aristas e_3 , e_4 . El camino b no puede agregar la arista e_5 debido a que se encuentra penalizada para el segmento formado por la arista e_4 .

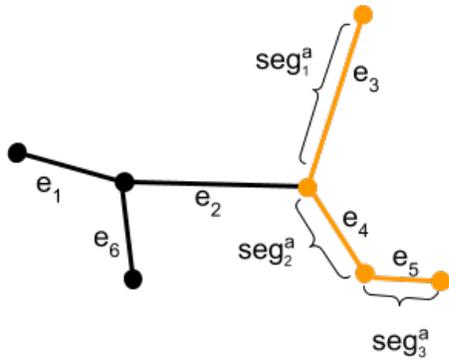


(d) Penalización del camino de la hormiga b en la arista e_4 con respecto al segmento seg_1^b conformado solamente por la arista e_3 .

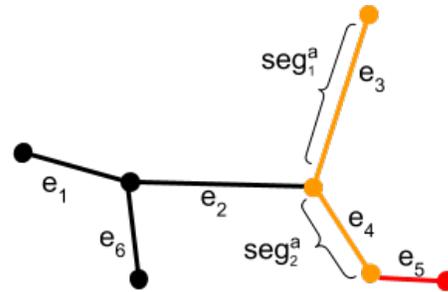


(e) El camino de la hormiga c no puede pasar de la arista e_4 en el caso que la penalización no tenga relación con el segmento que precede a esa arista.

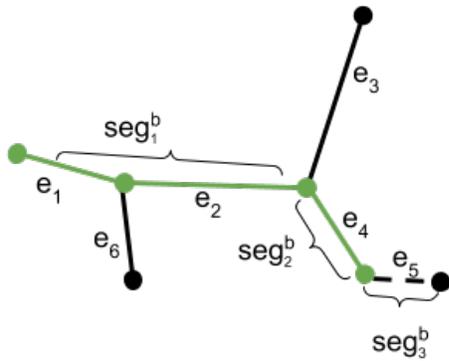
Figura 4.4: Funcionamiento de la propuesta de penalización de anti-feromonas con segmentos y recorrido inverso. Por simplicidad se supone que la penalización de τ_{ij} solo una vez es suficiente para eliminar la posibilidad de elegir una arista c_{ij} . En este ejemplo, los caminos a y b se desechan por evaluarse como de mala calidad. El camino es penalizando en e_4 respecto al segmento que lo precede. El camino c puede seleccionar la arista e_4 ya que esta arista no se encuentra penalizada con respecto al segmento seg_1^c , pudiendo luego agregar el segmento seg_3^c . Fuente: Elaboración Propia.



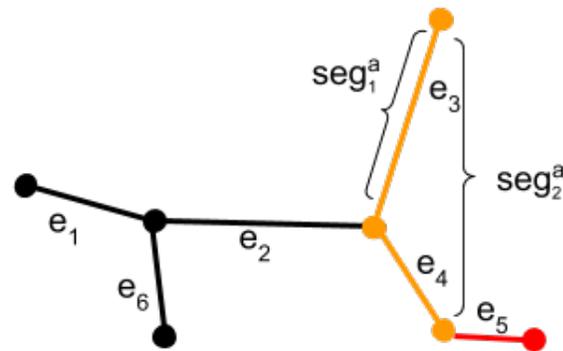
(a) Camino de una hormiga a que contiene las aristas e_3 , e_4 y e_5 .



(b) Penalización del camino de la hormiga a en la arista e_5 con respecto al segmento seg_2^a conformado solamente por la arista e_4 .



(c) Camino de una hormiga b que contiene las aristas e_1 , e_2 , y e_4 . El camino b no puede agregar la arista e_5 debido a que se encuentra penalizada para el segmento conformado por la arista e_4 , que corresponde al seg_2^a .



(d) Solución propuesta para los segmentos de solo 1 arista no causen bloqueos como el que afecta a la hormiga b , mediante la extensión del segmento, añadiendo el segmento anterior.

Figura 4.5: Caminos de las hormigas a y b . Para evitar que la penalización sobre e_5 por el segmento que lo precede, seg_2^a , bloquee el camino b , se modifica el inicio del segmento asociado a la penalización, incorporando las aristas del segmento previo, seg_1^a . Así se mantiene la lógica de asociar la penalización no solo a la arista sino que también al segmento que la precede. Estos casos suceden para aristas como e_4 , que forman ángulos en el rango $[\theta, Max_Angle]$ con todas sus aristas vecinas, quedando solas en un segmento de largo 1. Fuente: Elaboración Propia.

4.4. Criterios para la actualización de anti-feromonas

La anti-feromona se aplica sobre hormigas que han finalizado su recorrido, cuya calidad normalizada se encuentre entre $[Max_Score/2, Max_Score[$, las que tienden a ser caminos de calidad intermedia. Aquello implica que al menos un par de las aristas del recorrido forman un ángulo que pertenece al rango intermedio $]\theta, Max_Angle]$, necesitando un análisis adicional para determinar si corresponde a una solución de buena calidad. Este análisis consta de dos criterios que se aplican sobre todas las células cuyos filamentos se deseen individualizar, mientras que existe un tercer criterio adicional aplicable solo en el caso de las neuronas. Los dos criterios generales, o criterios comunes, consisten en evaluar la curvatura del recorrido, como también la magnitud del desplazamiento entre la proyección de un segmento con respecto a un segmento adyacente. El criterio adicional relativo solamente a las neuronas busca validar que la finalización del recorrido no sea en una arista con uno de sus nodos con grado 1. Es decir, esta arista final no debe cumplir con el primer criterio de la heurística de asignación de aristas iniciales, descrita en la sección 4.2.

4.4.1. Curvatura de una solución

La curvatura del recorrido de una hormiga a se obtiene al calcular el ángulo entre la proyección de un vector con respecto a otro. Los vectores son formados por el nodo inicial del camino, v_1^a , el centro de masa del recorrido de la misma hormiga, mc^a , y el nodo final, v_n^a . Así, se definen $\vec{p} = v_1^a - mc^a$ y $\vec{q} = mc^a - v_n^a$.

El ángulo entre la proyección de \vec{p} y \vec{q} no debe superar el valor que resulta al multiplicar el ángulo θ por el factor $Max_Axial_Displacement$. Este factor permite flexibilizar la tolerancia de la curvatura en base a θ . Si el recorrido de la hormiga tiene un ángulo igual o mayor al umbral, implica que la solución encontrada es demasiado curva para representar un filamento, por lo que es penalizada y descartada. Un ejemplo se puede observar en la Figura 4.6, con el camino formado por las aristas e_3, e_4, e_5 y e_7 , que presenta una curvatura muy marcada, por lo que es penalizado y descartado. La actualización del valor de τ_{ij} mediante el criterio de curvatura se refleja en la ecuación 4.3.

$$\tau_{ij} \leftarrow \begin{cases} \tau_{ij} \cdot \gamma & \text{si } \angle(\text{proy}(\vec{p}), \vec{q}) \geq \theta \cdot \text{Max_Axial_Displacement}, \\ 0 & \text{si } \tau_{ij} \leq 0,25, \\ \tau_{ij} & \text{en otro caso.} \end{cases} \quad (4.3)$$

4.4.2. Magnitud de desplazamiento entre segmentos

El análisis respecto a la magnitud del desplazamiento entre la proyección de un segmento en relación a sus segmentos vecinos se fundamenta en la rigidez que algunos tipos de filamentos como los microtúbulos y los filamentos de actina poseen (ver Stam et al. [35], Civalek and Demir [36]). El criterio de máximo desplazamiento entre segmentos contiguos consiste en limitar la magnitud del movimiento que representa un segmento en relación al ángulo que

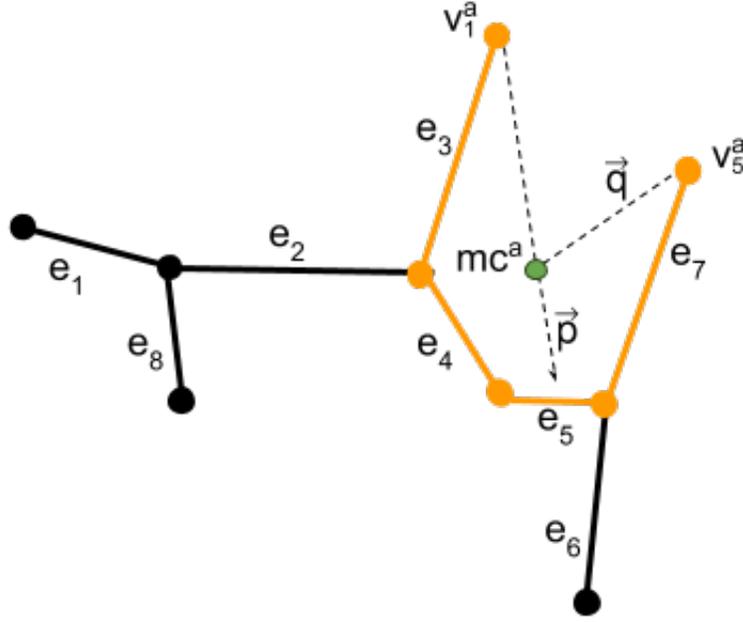


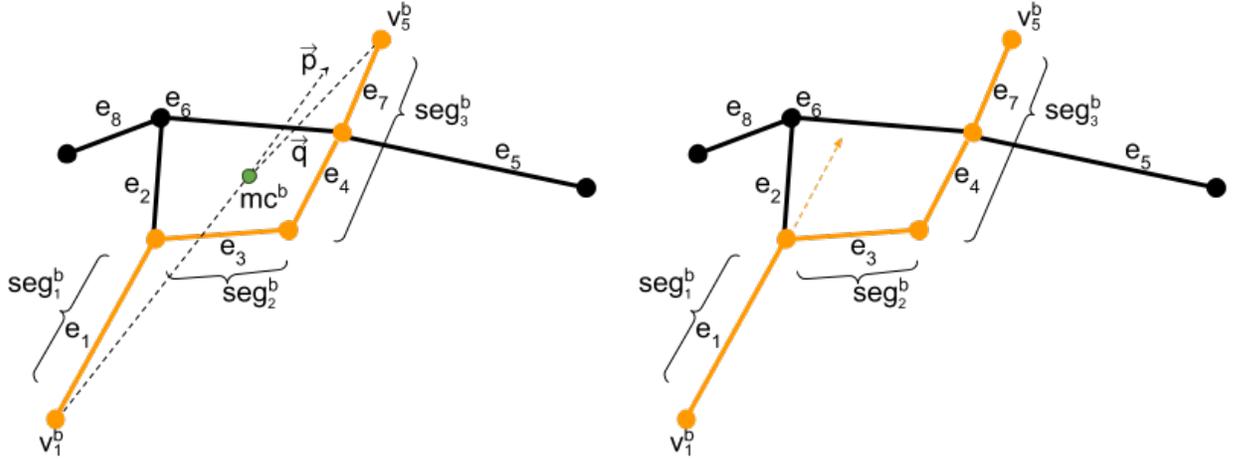
Figura 4.6: Curvatura de un camino de una hormiga a que contiene las aristas e_1 , e_3 , e_4 y e_7 . \vec{p} es la proyección del vector formado por v_1^a y mc^a . \vec{q} es el vector formado por mc^a y v_5^a . Para cumplir con el criterio de curvatura, el ángulo entre la proyección de \vec{p} y \vec{q} debe ser menor a $\theta \times Max_Axial_Displacement$ para no descartar la solución.

forma con sus segmentos adyacentes. Esto permite descartar soluciones que respeten el criterio de curvatura, pero que presentan movimientos pronunciados que sobrepasan la rigidez esperada de un filamento.

A partir de cada par de segmentos vecinos en una solución, se selecciona el segmento de mayor longitud, el que servirá para determinar cual es el movimiento proyectado con respecto al movimiento reflejado por el segmento de menor longitud. Se define el segmento más largo entre ambos segmentos comparados en un camino a como seg_{iMax}^a , mientras que el segmento menor es seg_{jMin}^a . El ángulo entre seg_{jMin}^a y la proyección de seg_{iMax}^a se define como $\angle seg_{pMaxMin}^a$ por simplicidad. En el caso de que el ángulo $\angle seg_{pMaxMin}^a$ este en el rango $[0, \theta]$ se entiende que se respeta el criterio de máximo desplazamiento.

Si el ángulo es mayor a θ , se espera que la magnitud del segmento menor seg_{jMin}^a multiplicada por el seno de $\angle seg_{pMaxMin}^a$ sea menor al umbral que delimita al desplazamiento máximo entre segmentos, definido como el múltiplo de seg_{iMax}^a por el 10% de $Max_Axial_Displacement$. Lo anterior se ilustra en la Figura 4.7, y se describe la actualización de τ_{ij} en la ecuación 4.4.

$$\tau_{ij} \leftarrow \begin{cases} \tau_{ij} \cdot \gamma & \text{si } \sin(\angle seg_{pMaxMin}^a) > seg_{iMax}^a \cdot 0,1 \cdot Max_Axial_Displacement \\ & \wedge \angle seg_{pMaxMin}^a > \theta, \\ 0 & \text{si } \tau_{ij} \leq 0,25, \\ \tau_{ij} & \text{en otro caso.} \end{cases} \quad (4.4)$$



(a) Solución b cumple con el criterio de curvatura, pero la proporción de la magnitud entre los segmentos seg_2^b y seg_1^b con respecto a su ángulo puede indicar que no es una buena solución.

(b) Análisis del desplazamiento del segmento seg_2^b con respecto a la proyección del seg_1^b denota un giro brusco y pronunciado que podría no respetar la rigidez de algunos tipos de filamentos.

Figura 4.7: Criterio de magnitud y pronunciación del desplazamiento entre 2 segmentos contiguos. (a) El criterio de curvatura puede no ser suficiente por si mismo para descartar todas las soluciones de mala calidad. (b) La información a priori respecto a la rigidez de un filamento permite descartar soluciones que presenten cambios demasiado pronunciados entre los segmentos que la conforman. Fuente: Elaboración Propia.

4.4.3. Criterio específico para neuronas

Si la célula observada es una neurona, corresponde aplicar el criterio adicional específico a esta célula. Este criterio adicional se fundamenta en que el comportamiento esperado de los filamentos en una neurona permite caracterizar los lugares desde los cuales nuevos filamentos pueden generarse. Así, es necesario validar que los filamentos de una neurona parten del soma o centro de la misma, y que filamentos posteriores que no comiencen del soma solo pueden nacer a partir de otros filamentos. Un camino que termine en un nodo final de grado de 1 no respeta aquel comportamiento, debido a que en un grafo que representa una red de filamentos de una neurona, los nodos ubicados en el soma o en la intersección entre filamentos presentan un grado mayor a 1. La penalización al valor de τ_{ij} para este caso particular se refleja en la ecuación 4.5.

$$\tau_{ij} \leftarrow \begin{cases} \tau_{ij} \cdot \gamma & \text{si } \deg(v_n^a) = 1, \\ 0 & \text{si } \tau_{ij} \leq 0,25, \\ \tau_{ij} & \text{en otro caso.} \end{cases} \quad (4.5)$$

4.5. Método de búsqueda no local

Una vez que la hormiga termina un recorrido, es posible agregar un método de retroalimentación sobre la calidad del camino construido, basado en lógicas globales/centralizadas que escapan de la búsqueda local que realiza cada hormiga. El o los métodos en esta etapa permite extender la lógica con respecto a una metaheurística ACO, y son opcionales, por lo que no siempre son utilizados.

Para la individualización de filamentos, la evaluación global corresponde a eliminar soluciones candidatas que no aporten información nueva, como puede suceder si una solución se encuentra contenida dentro de otra. Específicamente, si una solución s^a se conforma por uno o más segmentos y todos estos existen de forma equivalente en otra solución s^b , que a su vez puede tener segmentos que no posean equivalencia en s^a , se cumple que s^b contiene a s^a . Al ser una solución contenida en otra, s^a es redundante, pudiendo descartarse al disminuir su calidad a 0. En el caso que s^a y s^b sean idénticas, se conserva la solución que se haya construido primero.

Mediante la comparación de segmentos equivalentes, esta revisión centralizada busca evitar recorridos con un bajo número de aristas, especialmente los caminos de 1 arista. Un potencial problema asociado a este método se encuentra en el caso de que la solución descartada, s^a en el ejemplo, sea la que representa de mejor forma a un filamento, y que la solución restante por ende sea una representación no exacta.

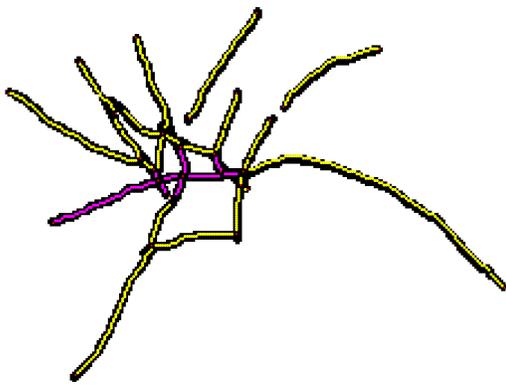
4.6. Extracción de información para individualizar filamentos

Previo al uso de la metaheurística ACO para individualizar filamentos, es necesario recabar la mayor cantidad de información posible a partir de la imagen, así como del grafo extraído. Además se debe considerar la incorporación de información específica de los filamentos en la célula observada, como el o los comportamientos esperados de estos, las que pueden conocerse previamente a la ejecución del algoritmo propuesto. A medida que se obtiene más información, es posible incorporar en el algoritmo diversas heurísticas que permitan acotar el espacio de búsqueda.

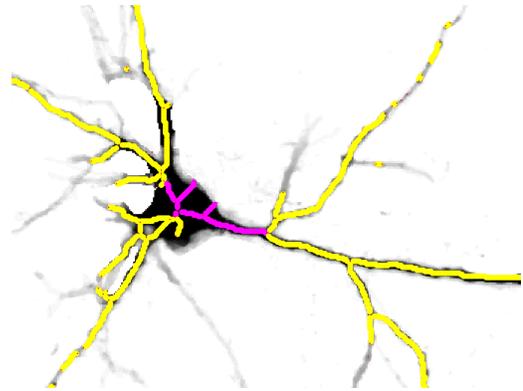
Como se observa en el capítulo 3, diversas investigaciones obtienen información geométrica o topológica, por lo general a partir de una imagen de filamentos. Para extender la obtención de información topológica con respecto a otros métodos, es posible utilizar los algoritmos para grafos existentes en la librería *sknw*, que son las implementaciones de algoritmos conocidos para grafos. Así, es posible utilizar un algoritmo de centralidad como *closeness centrality*, el que calcula la distancia de cada nodo con respecto a lo demás, permitiendo identificar los que se encuentran a la menor distancia de todos.

La relevancia de esta información adicional es directa en el caso de las neuronas, siendo utilizada en la sección 4.4.3, mientras que puede requerir de análisis adicional por parte de expertos para otros filamentos. La aplicación del algoritmo *closeness centrality* refleja la posibilidad de expandir y/o incrementar el uso de información topológica en el algoritmo propuesto. Un ejemplo de *closeness centrality* para filamentos en neuronas y microtúbulos se

observa en la Figura 4.8, en la que se eligieron los 3 nodos con menor distancia a los demás. Luego, el color magenta refleja las aristas que poseen al menos uno de estos nodos, mientras que las aristas en amarillo indican las que no poseen ninguno.



(a) Uso del algoritmo *closeness centrality* en un grupo de microtúbulos



(b) Uso del algoritmo *closeness centrality* para reflejar nodos y aristas centrales, las que pueden corresponder a la sección del soma en una neurona.

Figura 4.8: Ejemplo del algoritmo *closeness centrality* eligiendo los 3 nodos con menor distancia a los demás nodos. El color magenta indica que aristas poseen al menos uno de estos 3 nodos. En el caso de una neurona, permite asociar una o más aristas a una sección específica como el soma. Fuente: Elaboración Propia.

Además de la extracción de información previo al uso de la metaheurística ACO, es posible realizar de manera similar, la obtención de información durante la individualización de filamentos. Un ejemplo de esto es la información espacial con respecto a los nodos y aristas que pueden ser elegidas como parte de un camino. La información espacial facilita la comparación y el análisis entre caminos para identificar casos de cruce, intersección o solapamiento. En particular, en la sección 4.5 se utiliza la consulta espacial *within* disponible en la extensión espacial *Postgis* para la base de datos *PostgreSQL*, para comparar la información distinta que entrega un camino con respecto a los demás, considerando que puede existir repetición de caminos entre distintas hormigas. Este análisis resulta simple a través de una consulta espacial, especialmente a medida que aumentan los caminos a comparar.

El uso de información geométrica, topológica y espacial, repartidos en los métodos de la metaheurística ACO, se enfocan en ampliar las opciones de análisis que puede realizar un experto, así como aumentar la precisión de la individualización de filamentos. Nuevas características que modifiquen la exploración del espacio de soluciones pueden ser añadidas en el método de construcción de soluciones, ampliando el criterio que solamente utiliza el ángulo para determinar o descartar con certeza que 2 aristas pertenezcan o no al mismo filamento. A su vez, en este mismo método es posible incorporar nuevas condiciones que delimiten con mayor detalle el inicio o final de un camino. Por su parte, agregar restricciones adicionales que penalicen caminos finalizados puede realizarse tanto en el método de actualización de feromonas, en caso de que se trate de restricciones que no requieran comparación con otras hormigas, o en el método de búsqueda no local para comparaciones entre 2 o más hormigas.

4.7. Ponderación de Características

El algoritmo propuesto desarrollado para la individualización de filamentos considera el uso de características espaciales, topológicas y geométricas, cuyo uso puede variar dependiendo de la célula representada por el grafo extraído, así como la etapa de la metaheurística ACO en la cual se aplican. La etapa de construcción de soluciones emplea el ángulo entre aristas vecinas para definir la posibilidad de elección que cada hormiga tiene al ir avanzando. En esta etapa también se aplica la heurística de asignación inicial, en la que influye el grado de los nodos para determinar la arista inicial de cada camino. Por su parte, en el método de actualización de feromonas se usan las características geométricas como la curvatura del filamento y la diferencia en la magnitud entre segmentos contiguos, agregándose el uso de información topológica obtenida con el algoritmo de centralidad en grafos *closeness centrality*, en el caso de las neuronas. En cuanto al método de búsqueda no local, se utiliza información espacial para descartar soluciones que presenten información repetida con respecto a otras soluciones.

Dado que las diversas características utilizadas en el algoritmo propuesto se encuentran en distintos métodos de la metaheurística ACO, no es posible utilizar un único parámetro para obtener una ponderación directa. En cambio, se propone un criterio para su ponderación basado en otorgar $\frac{1}{3}$ del peso a cada etapa del modelo, distribuyendo aquel valor internamente dependiendo de si la característica se utiliza por sí sola o en combinación con otra, así como considerando si el uso es repetitivo durante la ejecución de la etapa o solo se realiza una vez por etapa. Lo anterior permite establecer la siguiente ponderación:

- Construcción de soluciones: $\frac{100}{n}$ % el grado de los nodos, y el remanente para el ángulo entre aristas, con n como el número de aristas.
- Método de búsqueda no local: 100 % posición de la arista.
- Actualización de feromonas: 50 % curvatura, 50 % diferencia de magnitud entre segmentos. En el caso específico de las neuronas, corresponde a $33.\bar{3}$ % curvatura, $33.\bar{3}$ % diferencia de magnitud entre segmentos y $33.\bar{3}$ % información topológica de centralidad.

Durante la construcción de soluciones, el grado de los nodos solo se utiliza en la heurística de asignación de la primera arista, por lo que para caminos de una sola arista corresponde a la totalidad de la ponderación de la etapa. A medida que crece el número de aristas, su peso va disminuyendo gradualmente al irse traspasando al criterio de ángulo entre aristas. En cuanto a la diferencia de magnitud entre segmentos, esta se conforma en partes iguales del ángulo entre segmentos y del uso de la magnitud de los mismos.

4.8. Complejidad computacional del algoritmo propuesto

Para calcular la complejidad computacional del algoritmo propuesto es necesario establecer previamente las condiciones del peor caso, lo que permitirá establecer la cota superior de la complejidad computacional. Estas condiciones consisten en suponer que el grafo que representa la red de filamentos es completamente conectado, así como que el algoritmo propuesto permita la opción de ciclos en los filamentos y superposición entre los mismos.

La asunción de que el grafo sea completamente conectado permite establecer que el número

de aristas corresponde a $\frac{n(n-1)}{2}$, con n como el número de nodos. Otras características del grafo corresponden a que es simétrico y no dirigido.

Basándose en Li [37] que establece la complejidad computacional para un grafo dirigido y simétrico, utilizado durante la resolución del problema del *Vendedor Viajero* mediante el uso de la metaheurística ACO, se realiza el cálculo de la complejidad computacional de las etapas del algoritmo propuesto, cuyo resultado se resume en la tabla 4.1. Los parámetros utilizados en la tabla 4.1 corresponden a s para indicar el número de hormigas, n para el número de nodos y Max_Iter para el número de iteraciones. Un tercer parámetro, Max_Iter , se agrega al cálculo de la complejidad computacional para reflejar el número máximo de iteraciones del algoritmo propuesto.

Etapa de ACO	Complejidad Computacional
Construcción de una solución	$\mathcal{O}(n^2s)$
Actualización de feromonas	$\mathcal{O}(n^2s)$
Búsqueda no local	$\mathcal{O}(\log(s))$

Tabla 4.1: Complejidad computacional de las etapas del algoritmo propuesto. n representa el número de nodos mientras que s representa el número de hormigas utilizadas.

En la etapa de construcción de una solución, sección 4.2, consiste en asignar una arista inicial, mediante la heurística de asignación inicial, para luego realizar el recorrido del grafo. Al suponer el peor caso, la heurística de asignación inicial debe ejecutar las 3 etapas que considera, implicando que se deba recorrer todo el conjunto de aristas, lo cual es de complejidad $\mathcal{O}(n^2)$.

Por su parte, el peor caso al ir recorriendo el grafo para añadir nuevas aristas a una hormiga s consiste en agregar las n^2 aristas disponibles al recorrido. Dado que cada hormiga podría realizar este procedimiento, se llega a la complejidad computacional de $\mathcal{O}(n^2s)$. Como $\mathcal{O}(n^2s)$ es mayor que la complejidad computacional de la heurística de asignación inicial, es esta complejidad la que define a esta etapa.

En cuanto a la etapa de actualización de feromonas, el peor caso se enfoca en el criterio respeto a la magnitud del desplazamiento entre segmentos (sección 4.4.2), dado que los otros criterios corresponden a un cálculo de costo constante $\nu(1)$. En particular, la situación en que todos los segmentos sean de solo una arista llevaría a tener que evaluar las n^2 aristas de cada hormiga, conllevando a una complejidad computacional de $\mathcal{O}(n^2s)$.

En la etapa de búsqueda no local (sección 4.5), el algoritmo propuesto realiza la búsqueda de información duplicada a través de una base de datos con extensión espacial, la que utiliza árboles para la indexación y búsqueda ante consultas. Esto permite establecer la complejidad computacional de esta etapa como $\mathcal{O}(\log(s))$, debido a que la comparación se realiza entre los recorridos de las s hormigas.

Así, es posible establecer que la complejidad computacional del algoritmo propuesto es $\mathcal{O}(Max_Iter(n^2s + \log(s)))$.

En resumen, las condiciones del problema de identificación de filamentos dan pie a establecer su representación mediante un problema de optimización de restricciones (COP), estando el modelo para la resolución del COP basado en la metaheurística ACO para su resolución. La implementación del algoritmo propuesto que resuelve el COP basado en la metaheurística ACO se encuentra escrito en C++.

Capítulo 5

Metodología de evaluación

Al basar la individualización de filamentos en un método que usa un grafo, el resultado obtenido es un conjunto de caminos, donde cada camino es a su vez un conjunto de aristas. El proceso de individualizar filamentos genera una partición o *clustering* del grafo, donde las aristas corresponden al *data set*, con cada camino representando a un cluster perteneciente al clustering. La partición propuesta por un método basado en grafos debe ser comparada con respecto a la partición generada por un experto, a la que nos referiremos como *ground truth*. Uno de los requisitos de esta comparación, es que ambas particiones se refieran al mismo *data set*, lo que implica que ambas deben realizarse en base al mismo grafo. Además se debe considerar que el clustering realizado por el experto puede tener una cantidad igual o distinta de clusters con respecto al clustering producido por el método a evaluar.

5.1. Métricas y Medidas

La comparación de particiones en donde una corresponde al *ground truth* implica que las mediciones a utilizar son del tipo de criterio externo (Manning et al. [38]). En este tipo de criterio hay índices basados en comparación mediante conteo de pares, como el índice *Rand* (\mathcal{R}), el índice *Jaccard* (\mathcal{J}), y la medida \mathcal{F} también conocida como *F1-Score*. Por otra parte, también existen métricas del área de teoría de la información como *Variation of Information* o VI para el mismo propósito. Estos métodos serán la base de comparación de particiones en esta investigación.

5.1.1. Índices Rand y Jaccard

La mayoría de los criterios para comparar particiones mediante conteo de pares suele fundamentarse en el uso de la matriz de confusión, también llamada matriz de asociación o tabla de contingencia (Meilă [39]). Esta tabla considera 4 casos en los que puede estar un par de elementos del *data set*, que para el caso de la individualización de filamentos son pares de aristas, en las particiones C y C' :

- N_{11} : Número de pares que están en el mismo cluster en C y C'
- N_{00} : Número de pares que están en distintos clusters en C y C'

- N_{10} : Número de pares que están en el mismo cluster en C pero no en C'
- N_{01} : Número de pares que están en el mismo cluster en C' pero no en C

Se tiene que $N_{11} + N_{00} + N_{10} + N_{01} = \frac{n(n-1)}{2}$, con n como el número de aristas.

La definición de los casos utilizados para formar la tabla de contingencia posibilita asociar cada caso con una correspondiente evaluación de clasificación como se indica en la tabla 5.1.

Casos para un par de aristas	Clasificación
N_{11}	Verdadero Positivo (<i>True Positive</i> o TP)
N_{00}	Verdadero Negativo (<i>True Negative</i> o TN)
N_{10}	Falso Positivo (<i>False Positive</i> o FP)
N_{01}	Falso Negativo (<i>False Negative</i> o FN)

Tabla 5.1: Equivalencia entre casos para un par de aristas con su respectiva clasificación

La ecuación 5.1 expresa que el índice Rand puede ser escrito en función de evaluaciones de clasificación, lo que lo hace coincidir con la definición de la medida de exactitud (*Accuracy*). \mathcal{R} adquiere el valor 0 para particiones totalmente diferentes, subiendo hasta llegar a 1, lo que indica que las particiones son idénticas. Algunas críticas al índice Rand señalan la alta sensibilidad que tiene al valor de TN, el que tiende a ser mucho mayor que el resto (Ben-Hur et al. [40]), la existencia de una alta dependencia al número de clusters (Wagner and Wagner [41]), y que los valores de \mathcal{R} tienden a concentrarse en un intervalo cerca de 1, puntualmente en el rango [0.5, 1] (Meilă [39], Vinh et al. [42]).

$$\mathcal{R}(C, C') = \frac{N_{11} + N_{00}}{n(n-1)/2} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.1)$$

Por su parte, el índice Jaccard es similar al índice Rand con la excepción que ignora la clasificación TN, como se observa en la ecuación 5.2. Su valor también oscila entre 0 y 1 para particiones totalmente distintas y particiones idénticas respectivamente. Una de las crítica es que puede entregar resultados no confiables para data sets muy pequeños.

$$\mathcal{J}(C, C') = \frac{N_{11}}{N_{11} + N_{01} + N_{10}} = \frac{TP}{TP + FP + FN} \quad (5.2)$$

La ventaja que presentan \mathcal{R} y \mathcal{J} sobre VI radica en que estos índices si pueden considerar casos de superposición.

5.1.2. Variation of Information

La métrica VI definida en Meilă [39] se fundamenta en la información asociada a la entropía de las particiones a comparar, así como en la información mutua que comparten. El rango de valores de VI comienza en 0 para 2 particiones iguales, llegando a $\log n$ para particiones absolutamente distintas, con n como el número de elementos del data set. En la ecuación 5.3 el término $H(C, C')$ hace referencia a la información que la partición C pierde al pasar a

C' . De forma similar, $H(C', C)$ indica la información que se gana al pasar de C a C' . Así, la función H mide la incertidumbre.

$$VI(C, C') = H(C, C') + H(C', C) \quad (5.3)$$

Uno de los problemas de VI es que esta definida de forma clara sólo para particiones que no se superponen entre sí (ver Breuer and Nikoloski [2]).

5.1.3. Mediciones adicionales

Mediante la relación entre los casos en que un par de aristas puede ser asignada (Tabla 5.1) y la clasificación equivalente es posible evaluar el resultado de una partición con respecto al *ground truth* con mediciones del área de recuperación de información como *Precision*, *Recall*, las que a su vez son la base para calcular *F-Measure*. La forma de obtener *Precision* y *Recall* a través de la clasificación equivalente a los casos en que un par de aristas puede ser asignada se refleja en las ecuaciones 5.4 y 5.5. Estos cálculos entregan mayor información sobre el comportamiento de la partición propuesta por el algoritmo a evaluar.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5.4)$$

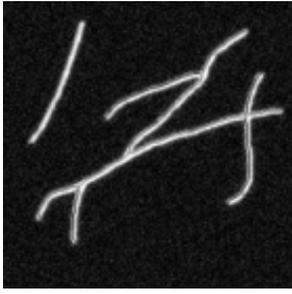
$$\text{Recall} = \frac{TP}{TP + FN} \quad (5.5)$$

Al mismo tiempo, en base a la definición de *Precision*, una evaluación directa a realizar es el calculo del número de filamentos correctos respecto al total de filamentos propuestos. A su vez. otra forma de evaluar *Recall* es mediante el calculo del número de filamentos correctos en relación a los filamentos individualizados manualmente por un experto. Otras medidas que se consideran son el porcentaje de cobertura de aristas, los tiempos de cómputo y la existencia de soluciones no consideradas por otros algoritmos. El porcentaje de cobertura corresponde a la cantidad de aristas contenidas al menos una vez en alguno de los filamentos propuestos.

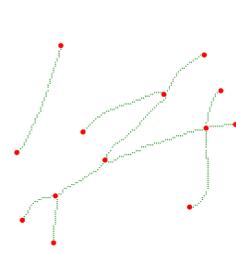
5.2. Imágenes Sintéticas

El uso de imágenes sintéticas se enfoca en la validación del algoritmo propuesto presentado en el capítulo 4, para casos simples que consideren filamentos que se superponen y/o intersectan. Ejemplos de las imágenes sintéticas utilizadas se observan en las Figuras 5.1 y 5.2, cada una acompañada de su grafo representativo de la red de filamentos y de su individualización manual. La Figura 5.1a corresponde a una imagen sintética obtenida de Qiu and LI [1], mientras que la Figura 5.2a es un subconjunto de la Figura 1b en Breuer and Nikoloski [2]. Las imágenes sintéticas se han obtenido de investigaciones en el estado del arte.

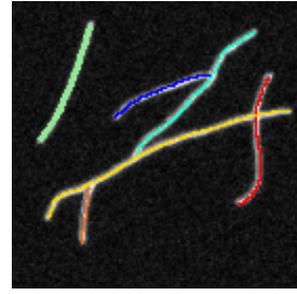
Para la evaluación, cada imagen sintética se ejecuta en DeFiNe con los parámetros base `BFS - Overlapping - Pairwise - Total`, ya que fueron los de mejor resultado en Breuer and Nikoloski [2], configurando el ángulo de umbral en 30° y 60°. Por otra parte, el algoritmo



(a) Red de filamentos sintética disponible en Qiu and LI [1].



(b) Grafo extraído de (a) utilizando la herramienta *sknw*.

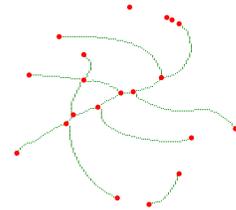


(c) Individualización manual de (a).

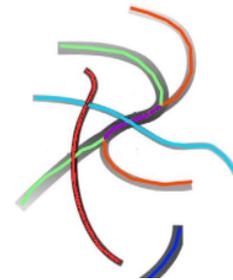
Figura 5.1: Filamentos sintéticos. a) Los colores originales hacen referencia a la segmentación de filamentos en vez de la individualización, por lo que se utiliza la imagen en escala de grises. b) Grafo de 11 aristas extraído. c) Individualización manual utilizando codificación por colores.



(a) Red artificial de filamentos extraída de una sección de la Figura 1b en Breuer and Nikoloski [2].



(b) Grafo extraído de (a) utilizando *sknw*.



(c) Individualización manual de filamentos.

Figura 5.2: Red sintética de filamentos a) Sección de la Figura 1b en Breuer and Nikoloski [2], correspondiente a una red sintética de filamentos que abarca casos de superposición y cruce. b) Grafo de 17 aristas. Se puede observar que el final de la arista superior derecha existe una discontinuidad que generando una arista y un nodo adicional. Esto es atribuible a la herramienta utilizada para extraer el grafo. c) Codificación por colores para individualizar filamentos. La sección de color púrpura claro en el centro de la red de filamentos corresponde a un caso de superposición entre dos filamentos.

propuesto cuenta con parámetros predefinidos dependiendo del tipo de célula, contando con las opciones de microtúbulo de planta, microtúbulo animal, neurona, retículo y sintético. Además cuenta con la opción de aplicar la heurística de asignación completa o sólo el primer paso, como sucede con las neuronas.

Cada una de estas opciones cuenta con un valor predefinido del umbral θ , del factor *Max_Axial_Displacement*, de si aplica la heurística de asignación inicial completa y de otros factores descritos en el capítulo 4, los que pueden modificarse. También es posible agregar nuevos tipos de célula con sus respectivos valores predefinidos. El objetivo de esto es facilitar el uso del algoritmo, asociando parámetros a tipos de célula para activar o desactivar diversas heurísticas existentes, o que sean añadidas en el futuro, buscando catalogar el com-

portamiento esperado de las células cuyos filamentos se busca individualizar. La configuración predefinida por tipo de célula se indica en la Tabla 5.2.

<div style="text-align: center;"> <div style="display: flex; justify-content: space-between;"> Parámetro Célula </div> </div>	θ	<i>Max Axial Displacement</i>	Permite Superposición	Aplica Heurística de Asignación Completa
Neurona (AP N)	45°	2	Si	No
MT Planta (AP MT-P)	30°	1.5	Si	Si
MT Animal (AP MT-A)	60°	2.5	Si	Si
Sintético (AP S)	45°	1.5	Si	No

Tabla 5.2: Configuración de parámetros predefinidos por tipo de célula, para el algoritmo propuesto (AP). MT es la abreviación de microtúbulo. Todos los parámetros indicados son ajustables por el usuario.

En las mediciones del algoritmo propuesto, se muestra el promedio de las 10 iteraciones, en las que se utiliza una semilla diferente en cada iteración. El uso de una semilla permite replicar los resultados de una iteración, dado que esta influye en el comportamiento de los números aleatorios generados para determinar la elección de aristas durante el proceso de construcción de una solución, definido en la sección 4.2, particularmente en la ecuación 4.1. Luego, al modificar la semilla en cada una de las 10 iteraciones, es posible medir la capacidad exploratoria del algoritmo.

5.3. Imágenes Reales

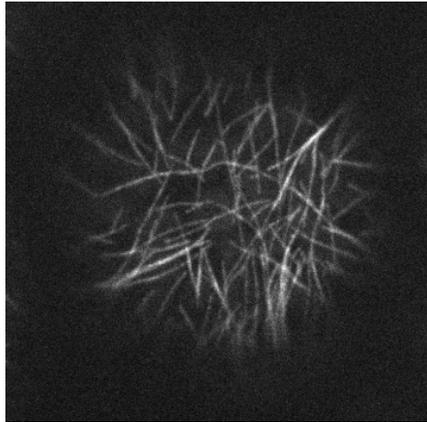
El procedimiento para individualizar filamentos en imágenes de microscopía comienza de forma similar para las células observadas en esta investigación, consistiendo en el análisis del *stack* o conjunto de imágenes capturadas durante una observación, las que pueden variar en el tiempo y en el eje Z. El criterio principal utilizado por los expertos se basa en determinar si existe continuidad de un posible filamento entre dos o más imágenes del *stack*.

Al observar un *stack*, el experto anota los filamentos de la célula mediante la selección del área de interés o ROI sobre una imagen que proyecta la unión de las imágenes del *stack* bajo algún criterio asociado a la intensidad de los píxeles en el eje Z. Las Figuras 5.3b, 5.4b y 5.5b son ejemplos de las anotaciones de ROIs realizadas para microtúbulos de la planta *Arabidopsis Marchantia*. Una descripción adicional de la individualización manual realizada por un experto puede ser encontrada en la sección A.2.

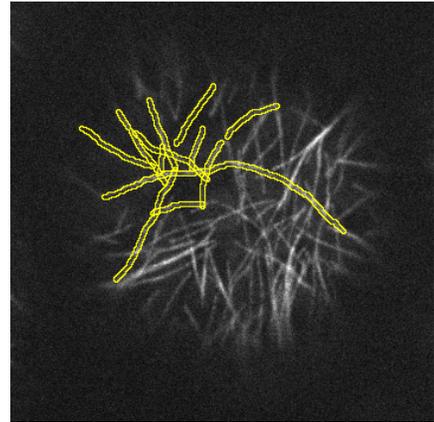
A diferencia de las imágenes de neuronas de ratón obtenidas de Ampuero et al. [43], en las que se procesa la imagen completa, la elección de un subconjunto de microtúbulos desde la imagen se debe a la complejidad de la individualización manual de estos por parte del experto. Se diferenciarán las imágenes de microtúbulos de las Figuras 5.3, 5.4 y 5.5 bajo la denominación de muestra MT-A, MT-B y MT-C respectivamente. Por su parte, las imágenes de neuronas de ratón en las Figuras 5.6, 5.7 y 5.8 se denominan N1, N2 y N3 respectivamente. Esta misma nomenclatura se utilizará para referirse a muestras adicionales a evaluar.

Una vez obtenida la individualización manual de filamentos, independientemente del tipo de célula, se construye una segmentación de las áreas de interés al pintar todos los píxeles fuera de estas zonas del mismo color que los píxeles en segundo plano o *background*, para luego generar una codificación de colores de los filamentos identificados por el experto que se encuentran en el primer plano. La codificación de colores permite una comparación visual con los resultados de DeFiNe o del algoritmo propuesto. Por su parte, el grafo se obtiene mediante la extracción de un esqueleto desde la imagen segmentada al utilizar la herramienta *sknw*, parte del software *ImagePy* (Wang et al. [21]). A partir del grafo se calcula el *closeness centrality* (Freeman [44]), que mide la lejanía promedio de cada nodo con respecto al resto, otorgando un valor más alto a los nodos que se encuentran más cerca de todos los demás. La selección de los 3 nodos de mayor valor proporciona información de la ubicación del **soma** en el caso de las neuronas.

Posteriormente, el grafo es escrito a un archivo en los formatos GML y JSON, los que son utilizados como archivos de entrada en DeFiNe y en el algoritmo propuesto respectivamente, para la individualización automática de filamentos. Los resultados de cada imagen analizada en el algoritmo propuesto se obtienen con el mismo procedimiento utilizado para las imágenes sintéticas.



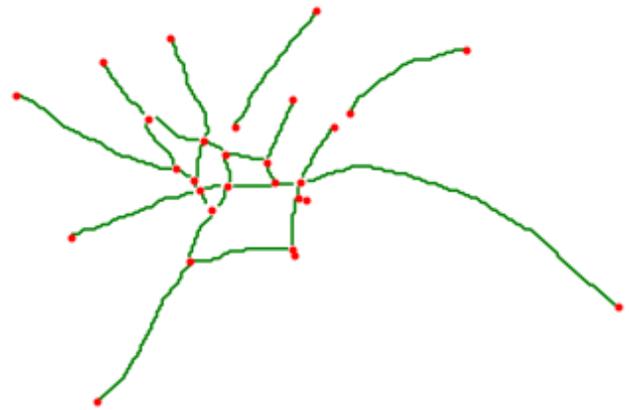
(a) Imagen de Microtúbulos.



(b) Filamentos manualmente individualizados por un experto.

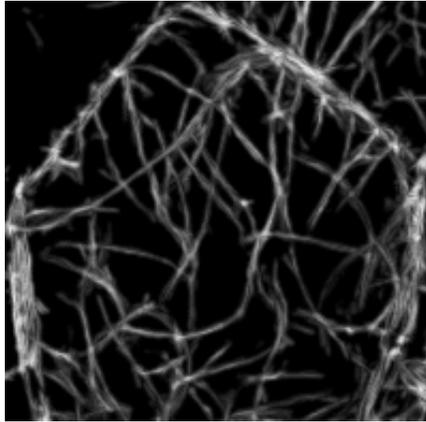


(c) Codificación por color de la individualización manual.

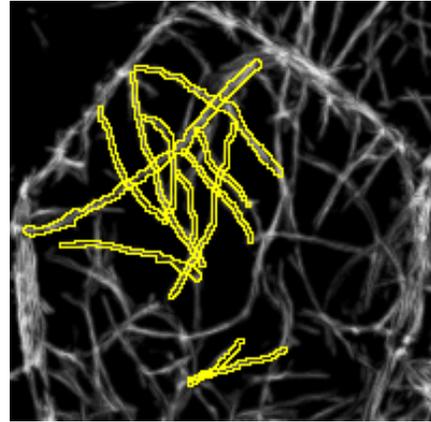


(d) Grafo que representa la red de microtúbulos manualmente identificada.

Figura 5.3: Individualización manual de microtúbulos. a) Imagen 6 de un stack de 301 cuadros de *Arabidopsis Marchantia*. b) Individualización manual c) Representación mediante colores de los diferentes microtúbulos. d) Grafo de 29 aristas extraído a partir de la individualización manual, que es el dato de entrada para los algoritmos comparados.



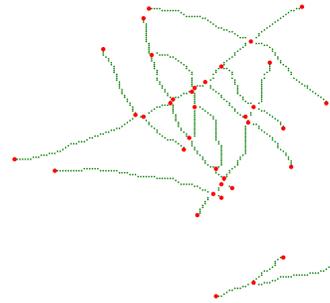
(a) Imagen de Microtúbulos.



(b) Filamentos manualmente individualizados por un experto.

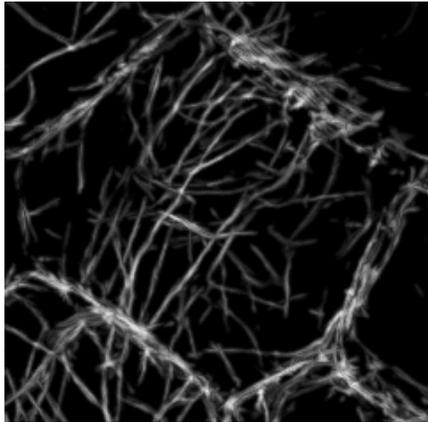


(c) Codificación por color de la individualización manual.

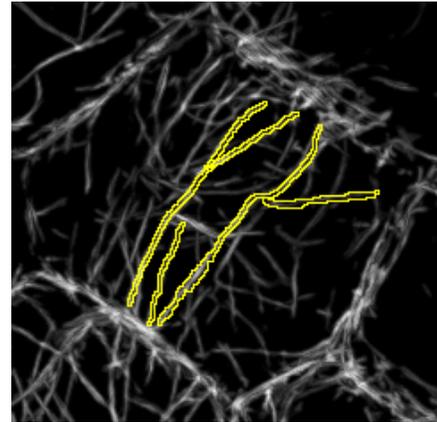


(d) Grafo extraído a partir de la individualización manual que es el dato de entrada para los algoritmos comparados.

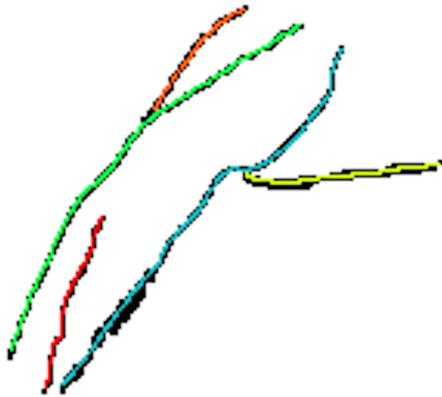
Figura 5.4: Individualización manual de microtúbulos. a) Imagen de *Arabidopsis Marchantia* obtenida al realizar la proyección del eje Z bajo el criterio de promedio de píxeles. b) Áreas de interés de la individualización manual c) Representación mediante colores de los diferentes microtúbulos. d) Grafo de 40 aristas utilizado en la individualización automática de filamentos.



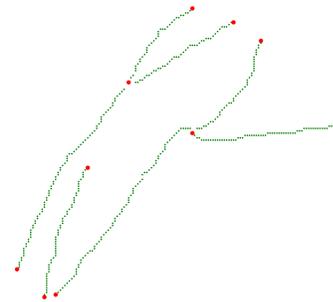
(a) Imagen de Microtúbulos.



(b) Filamentos manualmente individualizados por un experto.

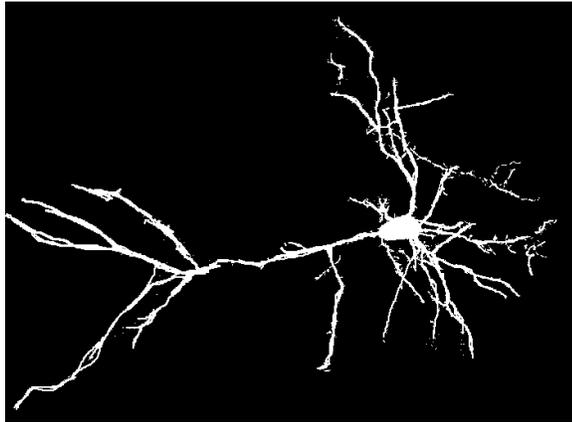


(c) Codificación por color de la individualización manual.

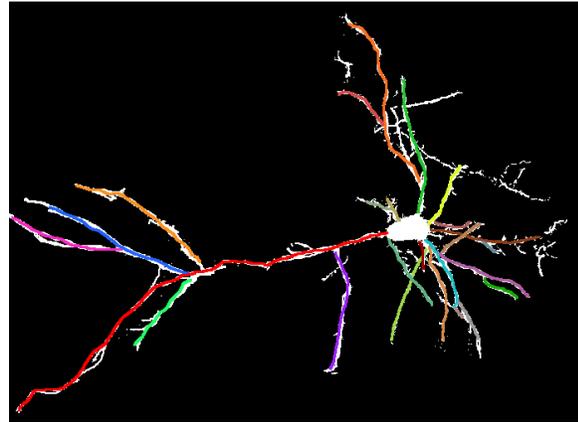


(d) Grafo base para la individualización automatizada de filamentos.

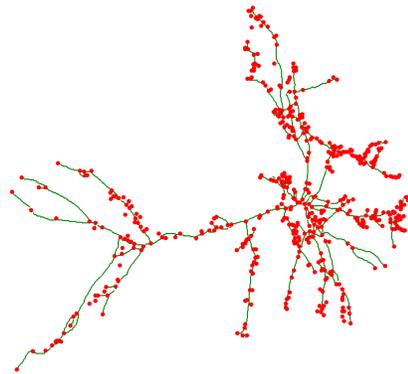
Figura 5.5: Individualización manual de microtúbulos. a) Imagen de *Arabidopsis Marchantia* obtenida al realizar la proyección del eje Z bajo el criterio de promedio de píxeles. b) Áreas de interés de la individualización manual c) Representación mediante colores de los diferentes microtúbulos. d) Grafo de 7 aristas extraído a partir de la individualización manual.



(a) Imagen de una neurona.

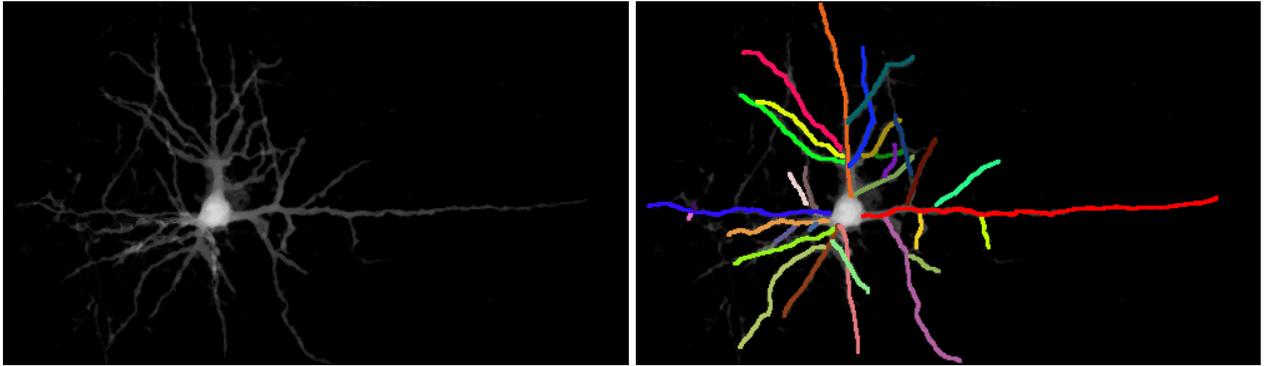


(b) Filamentos manualmente individualizados por un experto.



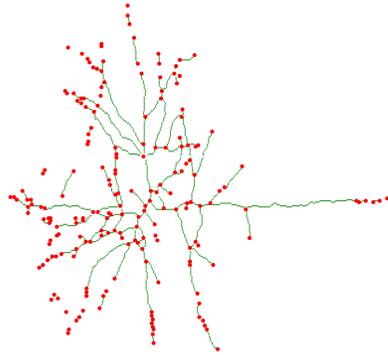
(c) Grafo base para la individualización automatizada de filamentos.

Figura 5.6: Individualización manual de una neurona. a) Imagen de Ampuero et al. [43] obtenida al realizar la proyección del eje Z bajo el criterio de valor máximo de píxeles. b) Representación mediante colores de los filamentos de la neurona, representando el axón y las dendritas. Se ignoran las estructuras alargadas que corresponden a ruido o son artefactos. d) Grafo de X aristas extraído a partir de la individualización manual.



(a) Imagen de una neurona.

(b) Filamentos manualmente individualizados por un experto.

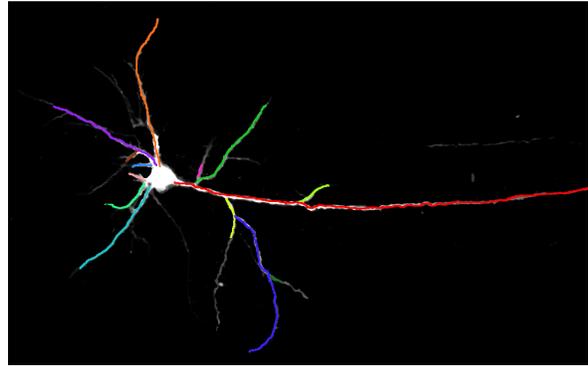


(c) Grafo base para la individualización automatizada de filamentos.

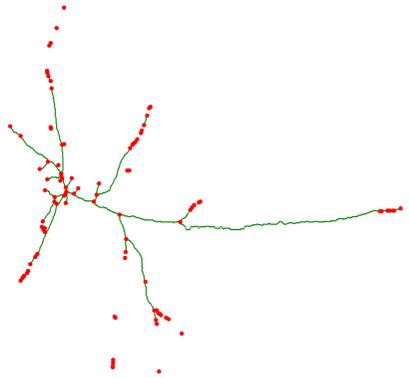
Figura 5.7: Individualización manual de una neurona. a) Imagen de Ampuero et al. [43] obtenida al realizar la proyección del eje Z bajo el criterio de valor máximo de píxeles. b) Representación mediante colores de los filamentos de la neurona, representando el axón y las dendritas. d) Grafo de X aristas extraído a partir de la individualización manual.



(a) Imagen de una neurona.



(b) Filamentos manualmente individualizados por un experto.



(c) Grafo base para la individualización automatizada de filamentos.

Figura 5.8: Individualización manual de una neurona. a) Imagen de Ampuero et al. [43] obtenida al realizar la proyección del eje Z bajo el criterio de valor máximo de píxeles. b) Representación mediante colores de los filamentos de la neurona, representando el axón y las dendritas. d) Grafo de X aristas extraído a partir de la individualización manual.

Capítulo 6

Resultados

Como se indica en la sección 3.3, la obtención de un grafo a partir de una imagen es un proceso complejo. Las imágenes analizadas en este trabajo son aquellas en las que fue posible realizar el procedimiento de esqueletonización, obteniendo un resultado que mantenía la topología de la estructura original. Lo anterior no es un resultado que pueda garantizarse para toda imagen, debido a que existen múltiples definiciones e implementaciones para realizar una esqueletonización (Lavado [18]). Así, para los métodos disponibles en las herramientas utilizadas durante esta investigación, existieron imágenes cuya información topológica no pudo ser obtenida.

Las mediciones del algoritmo propuesto, indicado como AP en las tablas de resultados, pueden utilizar los parámetros predefinidos para distintos tipos de célula, o pueden ser alterados manualmente. Cada evaluación del algoritmo propuesto consiste en el promedio de 5 iteraciones, en las que varía la semilla utilizada, pero se mantienen los parámetros elegidos. El detalle de cada iteración, en conjunto con el código y otros elementos necesarios para replicar estos resultados se encuentra en el repositorio git de esta tesis. Parte de este detalle se incluye también en el anexo 7. Las ejecuciones de DeFiNe y el algoritmo propuesto fueron realizadas en un computador con un procesador Intel i5-7200U de 4 núcleos, 8GB de RAM y un disco de estado sólido, bajo el sistema operativo Fedora 31.

6.1. Extracción de un grafo desde una imagen

La obtención de un grafo y la asociación de propiedades a sus nodos o aristas constituye el paso previo a la individualización automática de filamentos. De acuerdo a lo descrito en la sección 3.3.1, existen diversas formas para realizar este procedimiento. Para el caso de DeFiNe, sus autores indican que se preprocesa la imagen en escala de grises para resaltar las estructuras alargadas que debiesen ser filamentos, mediante un filtro de *vesselness* y un umbral de mediana adaptativa. El paso siguiente consiste en realizar la esqueletonización de la imagen, para luego construir un grafo en el que se asigna un peso a las aristas. Para obtener el peso de las aristas, se le aplica un filtro gaussiano a la imagen original en escala de grises, para luego promediar la intensidad a lo largo de cada arista.

Una réplica exacta de los pasos de aquel procedimiento no puede ser realizada, debido a que no se señala el tamaño del filtro gaussiano. Otro aspecto que dificulta una comparación directa es que DeFiNe no considera la curvatura de las aristas, sino que solo las representa como líneas rectas mediante la conexión punto a punto entre los nodos. Adicionalmente, no se encontraron las imágenes originales utilizadas en aquella investigación. Lo anterior implica que solo se pudo realizar una comparación precisa entre DeFiNe y el algoritmo propuesto, utilizando el único ejemplo que provee DeFiNe, que corresponde al grafo ponderado con el que se realiza la individualización de filamentos de la Figura 1b de aquella investigación, a partir de la cual se extrae una sección, representada en la Figura 5.2.

En los otros casos donde se utiliza DeFiNe para obtener una individualización de filamentos, se extrae el grafo necesario utilizando la herramienta *sknw*, la que pondera cada arista con su respectivo largo, el cual sí considera la curvatura. Este enfoque es compatible con lo que se presenta en Breuer and Nikoloski [2], ya que se indica que el método utilizado por DeFiNe sirve para cualquier red ponderada extraída a partir de una imagen.

6.2. Imágenes Sintéticas

Para la individualización de los filamentos sintéticos en la Figura 5.1 se utilizan los parámetros predefinidos en la opción de microtúbulos de planta (MtP), ya que la figura refleja similitudes con este tipo de filamentos. Los resultados se indican en la Tabla 6.1, observándose que DeFiNe logra 4 de 6 individualizaciones al utilizar 30° , mientras que con 60° obtiene 3 de 6 filamentos correctos. Este último resultado es el mismo que se obtiene al promediar todas las iteraciones del algoritmo propuesto con los parámetros de microtúbulo de planta,

Debido a que el comportamiento de los filamentos sintéticos es más simple que el de los filamentos de microtúbulos de planta, se realizaron pruebas utilizando como base los parámetros para filamentos sintéticos, modificando el ángulo a 20° y aplicando la heurística de asignación inicial parcialmente. La personalización de los parámetros se define como la configuración S1. El ajuste de parámetros se basa en que los filamentos sintéticos no requieren de explorar diversos puntos de partida como si sucede en una imagen real de filamentos. Estos parámetros manualmente asignados permiten encontrar 4 de los 6 filamentos individualizados por un experto, obteniendo también los mejores resultados para el cálculo de *Precision* y *Recall* que utilizan la clasificación de la Tabla 5.1 como base. La comparación entre los filamentos propuestos por cada método y los correctamente identificados se observa en la Figura 6.1. En la misma tabla se indican los resultados de la mejor iteración del algoritmo propuesto al utilizar la configuración MtP y S1.

En este caso de filamentos sintéticos, el filamento azul observable en la Figura 5.1c, corresponde a una sola arista del grafo (Figura 5.1b), por lo que también en todas las pruebas con DeFiNe y con el algoritmo propuesto se obtiene para ese filamento un calce no exacto. Este calce no exacto no se considera una respuesta correcta, pero representa una situación que puede suceder en imágenes reales, en las que puede representar el nacimiento de un filamento a partir de otro existente, como sucede en neuronas o microtúbulos. Luego, la sobre-representación del filamento correcto en la figura (filamento azul) sucede al agregar una arista del filamento vecino, de color celeste.

Ambas ejecuciones de DeFiNe, así como todas las ejecuciones del algoritmo propuesto logran asignar cada arista a al menos un camino que representa un filamento. Este comportamiento es obligatorio para el método de DeFiNe, ya que es una restricción en el *Set Cover* que DeFiNe usa, pero no lo es para el algoritmo propuesto.

Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.72	-	0.47	-	0.57	-	4/6	-	4/6	-	2.3
DeFiNe 60°	1	0.63	-	0.58	-	0.60	-	3/5	-	3/6	-	2.3
AP MtP	10	0.51	0.57	0.32	0.57	0.39	0.57	3/6.2	3/5	3/6	3/6	0.3
AP S1	10	0.68	0.87	0.57	1	0.62	0.93	4/5.8	4/5	4/6	4/6	0.2

Tabla 6.1: Resultado de la individualización de filamentos de la Figura 5.1, para las distintas configuraciones de DeFiNe y el algoritmo propuesto. El número de filamentos correctos es 6 en un grafo de 11 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el máximo obtenido tras el número de iteraciones, mientras que las sin asterisco representan el promedio.

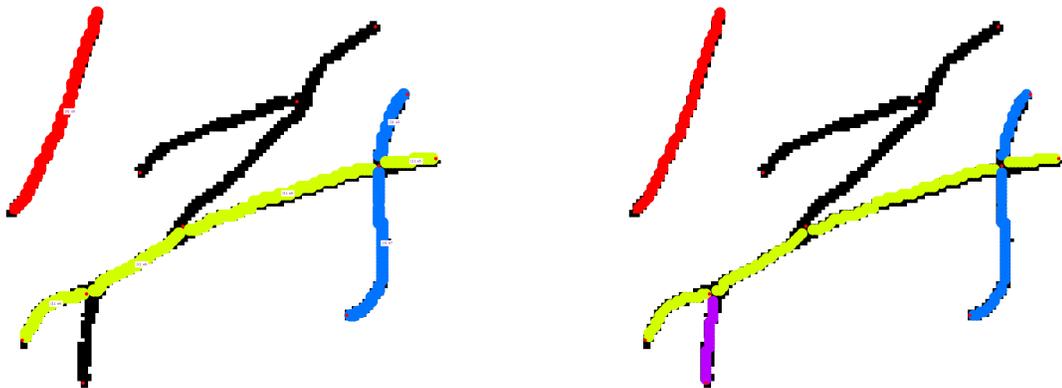
Como se menciona en la sección 6.1, la Figura 5.2 corresponde a una parte de la Figura 1b en Breuer and Nikoloski [2], siendo la única imagen para la que se dispone de un grafo ponderado según el criterio de los autores de DeFiNe. Así, se asume que aquel es el resultado óptimo, en el que se individualizan correctamente los 5 filamentos. Debido a una limitación de la herramienta *sknw*, el grafo de la Figura 5.2, utilizado en la evaluación del algoritmo propuesto y de DeFiNe con la ponderación utilizando el largo, contiene una arista adicional, por lo que los resultados consideran 6 filamentos como *ground truth*.

En la evaluación del algoritmo propuesto utilizando los parámetros definidos para microtúbulos de plantas, se obtienen 3 individualizaciones correctas de los 5 filamentos que corresponden al *ground truth*. En esta evaluación con la configuración MtP, solo 1 filamento individualizado se encuentra dentro del grupo de filamentos que se intersectan y sobrepone entre sí. Es posible atribuir esto a la curvatura de algunos de los filamentos sintéticos, la que es mayor a la que se espera de un microtúbulos de planta. Para complementar la evaluación, se realiza una segunda evaluación utilizando una configuración basada en el tipo de célula sintética, modificando el factor *Max_Axial_Displacement* a 2, asociado a la curvatura que presenta la figura, y no utilizando la heurística de asignación inicial completa, dado que no existen casos complejos de nacimiento de filamentos a partir de otros como en filamentos reales. Esta configuración se define como S2.

Así, en la evaluación que emplea la configuración S2 se encuentra un filamento adicional, y a su vez, es capaz de encontrar el filamento naranja de la individualización manual del experto, indicado en la Figura 5.2c, pero es descartado por no respetar el criterio de la magnitud de desplazamiento entre segmentos. El comportamiento del filamento naranja en el *ground truth* realiza un cambio de sentido en sus curvaturas, además de que estas son bastante pronunciadas, lo que no es un comportamiento esperado en los filamentos estudiados durante esta investigación. Los resultados se presentan en la Tabla 6.2.



(a) Representación del resultado de DeFiNe con ángulo de 30° , con 4 filamentos correctamente individualizados, identificados con colores. (b) Representación del resultado de DeFiNe con ángulo de 60° , con 3 filamentos correctamente individualizados, identificados con colores.

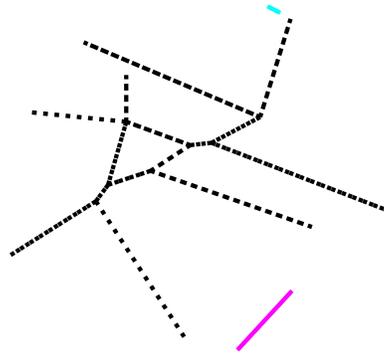
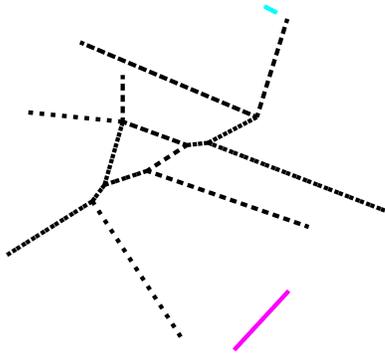


(c) Filamentos individualizados con colores a partir de la mejor iteración del algoritmo propuesto con la configuración MtP. (d) Filamentos individualizados con colores a partir de la mejor iteración del algoritmo propuesto con la configuración S1.

Figura 6.1: Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 5.1. El algoritmo propuesto individualiza correctamente 3 filamentos con la configuración para microtúbulos de planta, mejorando a 4 filamentos con una configuración más simple.

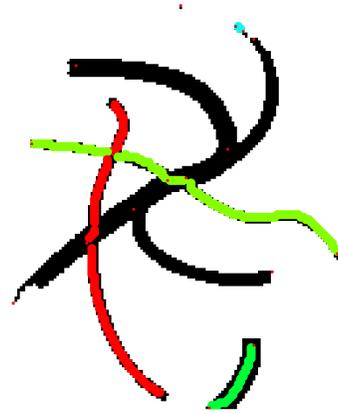
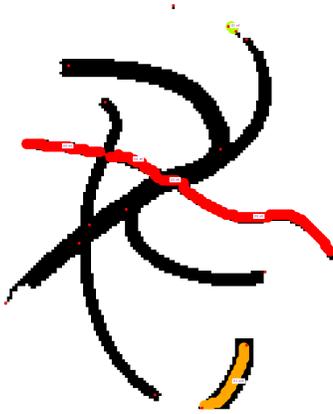
Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.33	-	0.18	-	0.23	-	2/11	-	2/5	-	2.8
DeFiNe 60°	1	0.23	-	0.25	-	0.24	-	2/7	-	2/5	-	3.6
AP MtP	10	0.34	0.35	0.27	0.28	0.3	0.31	3/9.2	3/9	3/6	3/6	0.3
AP S2	10	0.32	0.33	0.25	0.31	0.28	0.32	4/8.6	4/9	4/6	4/6	0.3

Tabla 6.2: Resultado de la individualización de filamentos para la Figura 5.2. El número de filamentos definidos manualmente por un experto es 5, en un grafo de 17 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el máximo obtenido tras el número de iteraciones, mientras que las sin asterisco representan el promedio. La diferencia del número de filamentos en el *ground truth* se debe a que la herramienta utilizada para la obtención del grafo corta un filamento, generando un filamento adicional.



(a) Representación del resultado obtenido a través de DeFiNe con 30° , logrando 2 filamentos correctamente individualizados, identificados con colores.

(b) Representación del resultado obtenido a través de DeFiNe con 60° , logrando 2 filamentos correctamente individualizados, identificados con colores.



(c) 3 filamentos correctamente individualizados por el algoritmo propuesto con la configuración MtP.

(d) 4 filamentos correctamente individualizados por el algoritmo propuesto con la configuración S2.

Figura 6.2: Resultados de la individualización de filamentos para la Figura 5.2, obtenidas mediante DeFiNe y el algoritmo propuesto, con diferentes parámetros. El algoritmo propuesto encuentra más filamentos que los que se encuentran en las pruebas con DeFiNe. Sin embargo se aleja del comportamiento óptimo indicado en Breuer and Nikoloski [2]. Esta diferencia puede atribuirse a la excesiva curvatura de los filamentos sintéticos.

6.3. Imágenes Reales

En esta sección se muestra un análisis con respecto a la individualización de filamentos realizada por dos expertos sobre la misma imagen, así como los resultados de la individualización de filamentos para 2 tipos de células, los microtúbulos existentes en la planta *Arabidopsis Marchantia*, y las dendritas en neuronas de ratón.

6.3.1. Individualización de diferentes expertos sobre la misma red de filamentos

En relación a lo indicado en la sección 3.3 con respecto a la posibilidad de que dos expertos difieran al individualizar filamentos, debido al problema asociado a la resolución de la imagen con la red de filamentos, se le solicitó a un segundo experto la realización de la individualización manual de filamentos sobre la imagen 5.3a (muestra MT-A).

La diferencia entre las individualizaciones realizadas por los expertos coincide con lo descrito en la sección 3.3. En este caso, el primer experto realiza la individualización de 12 filamentos, mientras que el segundo experto identifica 19. En base a lo anterior se tienen casos de filamentos encontrados solo por un experto, filamentos encontrados por ambos, así como filamentos con un calce parcial entre lo que individualiza cada experto. Lo anterior se resume en la Figura 6.3.

Entre todas las pruebas realizadas al algoritmo propuesto, solo la muestra MT-A es evaluada por 2 expertos, con el riesgo asociado de que exista un sesgo del algoritmo propuesto a la forma en que un experto individualiza filamentos. Sin embargo, el uso de dos tipos de células, así como la posibilidad de ajustar los parámetros de ejecución del algoritmo propuesto son mecanismos que permiten acotar un posible sesgo.

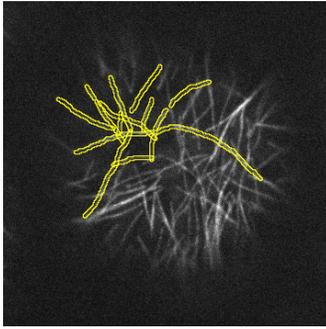
Tanto la evaluación de la muestra MT-A por parte de un segundo experto, como la combinación de las individualizaciones de filamentos realizada por cada experto permiten añadir 2 nuevas muestra a evaluar por el algoritmo propuesto. La segunda evaluación de la muestra MT-A, que contiene 19 filamentos se denomina MT-A2, mientras que la combinación de las individualizaciones que consta de 23 filamentos se denomina MT-A3. La muestra MT-A2 corresponde a la Figura 6.3b, y por su parte la muestra MT-A3 corresponde a la Figura 6.3c.

6.3.2. Microtúbulos de planta

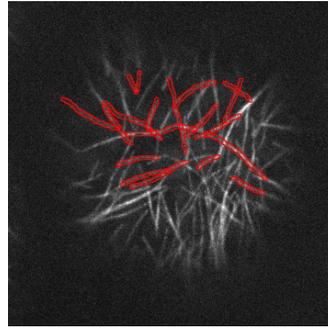
Muestra MT-A

La primera evaluación corresponde a la muestra MT-A, indicada en la figura 5.3, la que corresponde a una individualización manual realizada por un experto a partir de una red de microtúbulos en la planta *Arabidopsis Marchantia*. La evaluación llevada a cabo mediante el algoritmo propuesto utiliza los parámetros predefinidos para microtúbulos de planta (configuración MtP), y sus resultados, en conjunto con los resultados de DeFiNe se presentan en la Tabla 6.3.

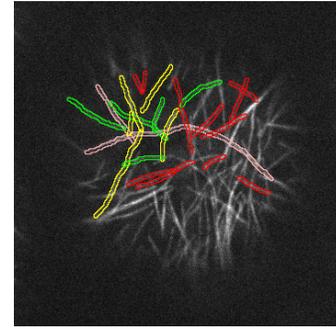
De los 12 filamentos identificados por un experto, DeFiNe con un ángulo de 30° individualiza 8 correctamente, mientras que con un ángulo de 60° logra 3 individualizaciones



(a) Muestra MT-A: 12 filamentos manualmente individualizados por un experto.



(b) Muestra MT-A2: 19 filamentos manualmente individualizados por un experto diferente.



(c) Muestra MT-A3: Comparación entre las individualizaciones de los expertos.

Figura 6.3: Individualización de filamentos de dos expertos ante la misma red de filamentos. a) y b) indican los filamentos identificados por cada experto, utilizando el color amarillo para los filamentos individualizados por el primer experto, mientras estos se señalan en rojo para el segundo experto. c) Comparación entre los filamentos indicados por cada experto. Los 5 filamentos en amarillo corresponden a filamentos encontrados solo por el primer experto, mientras que en rojo se indican los 11 filamentos encontrados solamente por el segundo experto. Los 5 filamentos individualizados de la misma forma por ambos expertos se destacan en verde, mientras que los 2 filamentos en color rosado indican un calce parcial entre lo que ambos expertos individualizan en aquel sector de la imagen.

correctas. Por su parte, el resultado del algoritmo propuesto con los parámetros predefinidos en la configuración MtP, indica que 4 de las 5 iteraciones individualizan al menos 7 filamentos correctamente, teniendo un caso donde se logran 8 aciertos. Lo anterior se refleja en la Figura 6.4. Se debe destacar que en 3 de las 4 iteraciones que no logran encontrar el filamento adicional, existe una sobre-representación de aquel filamento ya que se le asigna una arista adicional, lo que lleva a no considerarlo como una respuesta correcta.

Otros aspecto de estos resultados es que todas las evaluaciones logran una cobertura total de las aristas en sus propuestas de individualización. Por su parte, las iteraciones con la configuración MtP en el algoritmo propuesto muestran estabilidad en el resultado para los parámetros predefinidos, logrando obtener la mayoría de los filamentos que obtiene DeFiNe en sus 2 evaluaciones.

Muestra MT-A2

La evaluación de la muestra MT-A2 (Figura 6.3b) que corresponde al análisis realizado por un segundo experto sobre los filamentos en la Figura 5.3, tiene como resultado la individualización de 9 filamentos por parte de DeFiNe con su configuración DeFiNe 60°, mientras que se identifican 11.9 filamentos en promedio mediante el algoritmo propuesto, con la configuración para microtúbulos de planta o MtP.

En cuanto a los filamentos propuestos por cada método y su respectiva configuración, el algoritmo propuesto identifica un total de 21 filamentos en todas sus iteraciones, mientras que DeFiNe indica 25 o 19 filamentos totales, dependiendo de la configuración usada, como

Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.65	-	0.45	-	0.53	-	8/16	-	8/12	-	4.1
DeFiNe 60°	1	0.28	-	0.21	-	0.24	-	3/12	-	3/12	-	3.6
AP MtP	10	0.44	0.45	0.34	0.35	0.38	0.4	7.2/12.4	8/13	7.2/12	8/12	0.3

Tabla 6.3: Resultados de individualización de filamentos para la muestra MT-A en la Figura 5.3. El número de filamentos en el *ground truth* es 12, en un grafo de 29 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio.

se indica en la tabla 6.4. El tiempo promedio de las ejecuciones del algoritmo propuesto es menor al de DeFiNe con cualquier configuración.

Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.37	-	0.25	-	0.3	-	8/25	-	8/19	-	2.0
DeFiNe 60°	1	0.29	-	0.3	-	0.3	-	9/19	-	9/19	-	4.7
AP MtP	10	0.46	0.46	0.41	0.44	0.43	0.45	11.9/21	12/21	11.9/19	12/19	1.6

Tabla 6.4: Resultados de la individualización de filamentos para la muestra MT-A2 (Figura 6.3b). 19 son los filamentos definidos por un experto, para un grafo de 39 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. El algoritmo propuesto (AP) presenta un comportamiento estable, proponiendo un número de filamentos cercano al número de filamentos correctos, así como individualizando un mayor número de filamentos correctos, en un menor tiempo.

Muestra MT-A3

La muestra MT-A3 se obtiene al superponer las individualizaciones de filamentos realizadas por distintos expertos para la Figura 5.3. Esta fusión permite obtener un grafo de 59 aristas a través del cual se deben identificar los 23 filamentos indicados como *ground truth*. A pesar de encontrar solo 12 filamentos de forma exacta, el algoritmo propuesto encuentra el filamento más extenso, identificado en color rojo en la Figura 6.6, que consiste en 12 aristas. La relevancia de lo anterior radica en la dificultad de encontrar un filamento que contiene múltiples aristas cortas de forma consecutiva.

Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.38	-	0.17	-	0.23	-	9/37	-	9/23	-	4.1
DeFiNe 60°	1	0.25	-	0.19	-	0.22	-	8/24	-	8/23	-	19.6
AP MtP	10	0.32	0.37	0.18	0.25	0.23	0.3	10.9/29.4	12/27	10.9/23	12/23	1.9

Tabla 6.5: Resultados de la individualización de filamentos para la muestra MT-A3 (Figura 6.3c). 23 son los filamentos definidos por la superposición de 2 individualizaciones de filamentos realizadas por distintos expertos, para un grafo de 59 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. El algoritmo propuesto (AP) presenta 29.4 filamentos en promedio, alejándose de los 23 filamentos correctos. Sin embargo, presenta un número mayor de filamentos correctos, en un menor tiempo.

Muestra MT-B

La segunda evaluación de microtúbulos de planta corresponde a la muestra MT-B, que se observa en la Figura 5.4, y cuya individualización de filamentos en el caso del algoritmo propuesto se realiza también mediante la configuración MtP. En base a los resultados del algoritmo propuesto, los que individualizan correctamente 8 filamentos en promedio, se observa la necesidad de establecer rangos de tolerancia para determinar si una solución propuesta corresponde a un filamento. Esto se debe a que el uso de un umbral estricto puede llevar a descartar estas soluciones sin su completa evaluación.

Un aspecto a destacar del algoritmo propuesto radica que 4 de las 5 iteraciones con la configuración MtP encuentran 9 filamentos, proponiendo 15 filamentos en promedio, lo que es menor a ambas evaluaciones realizadas con DeFiNe. A su vez, el algoritmo propuesto muestra un mejor desempeño en su tiempo de ejecución, demostrando poder soportar grafos de mayor número de nodos y aristas.

Muestra MT-C

La individualización de filamentos para la Figura 5.5, que representa la muestra MT-C, entrega los mismos resultados al ejecutar tanto DeFiNe con 60° como el algoritmo propuesto con los parámetros predefinidos para microtúbulos de planta (configuración MtP). Esto puede deberse a que la muestra MT-C es de mayor simplicidad que las muestras anteriores. Por su parte, la ejecución de DeFiNe con 30° individualiza correctamente 3 de 5 filamentos. Los resultados se indican en la Tabla 6.7 y en la Figura 6.8.

El filamento faltante para el algoritmo propuesto corresponde a una sobre-asignación de aristas, similar a lo que ocurre en otras imágenes, llevando a no considerar aquella solución como un filamento válido. La posibilidad de sobre-asignar una arista a un filamento se condice con el comportamiento del algoritmo propuesto al favorecer la búsqueda de caminos/filamentos de mayor longitud que cumplan con el criterio de rectitud. Particularmente en este caso,

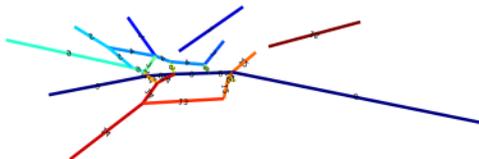
Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.51	-	0.17	-	0.26	-	5/23	-	5/12	-	5.0
DeFiNe 60°	1	0.42	-	0.31	-	0.36	-	5/16	-	5/12	-	16.2
AP MtP	10	0.5	0.52	0.32	0.40	0.39	0.45	8.8/15	9/14	8.8/12	9/12	0.9

Tabla 6.6: Resultados de la individualización de filamentos para la muestra MT-B (Figura 5.4). 12 son los filamentos definidos por un experto, para un grafo de 40 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. El algoritmo propuesto (AP) presenta un comportamiento favorable, presentando un número de filamentos propuestos cercano a la cantidad de filamentos correctos.

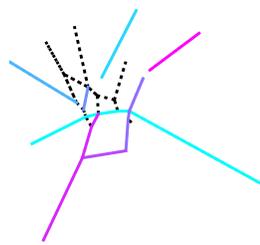
un microtúbulo de planta puede corresponder a la fusión de 2 microtúbulos en uno, denominado **Zippering**, o al nacimiento de un microtúbulo a partir de uno existente, llamado **Nucleación**. Ambas situaciones requieren de información adicional para diferenciarlas, pudiendo usarse características como el grosor de los segmentos de los filamentos involucrados en estas situaciones, así como el ángulo entre los segmentos de filamentos que se separan del segmento común. La variación o continuidad del grosor y que el ángulo mencionado se encuentre en un rango que varía según el tipo de célula son críticos para aclarar cual es el caso observado.

Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.5	-	0.5	-	0.5	-	3/5	-	3/5	-	2.82
DeFiNe 60°	1	0.66	-	0.5	-	0.57	-	4/5	-	4/5	-	2.65
AP MtP	10	0.66	0.66	0.5	0.5	0.57	0.57	4/5	4/5	4/5	4/5	0.29

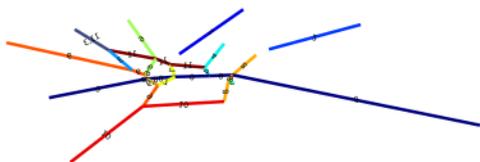
Tabla 6.7: Resultados de individualización de filamentos para Figura 5.5, de la muestra MT-C. Existen 7 aristas en el grafo que representa la red de filamentos, y a su vez el número de filamentos correctos es 5. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. Para esta muestra, todas las iteraciones del algoritmo propuesto (AP) obtienen el mismo resultado, los que también sucede en la ejecución de DeFiNe con 60°, siendo la única diferencia que el algoritmo propuesto realiza la individualización en menor tiempo.



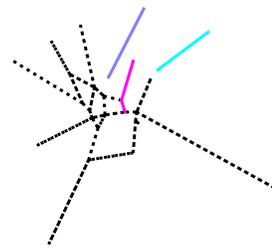
(a) Individualización obtenida mediante DeFiNe con un ángulo de 30° .



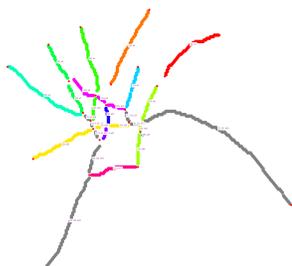
(b) Representación de los filamentos correctamente individualizados en (a), identificados con colores.



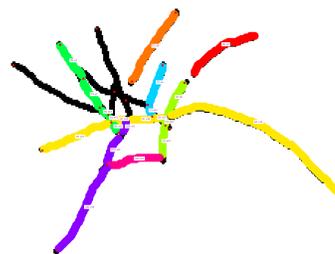
(c) Individualización obtenida mediante DeFiNe con un ángulo de 60° .



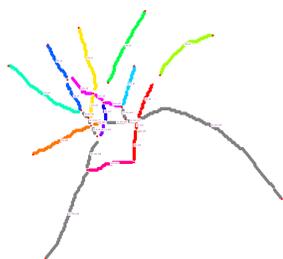
(d) Representación de los filamentos correctamente individualizados en (c), identificados con colores.



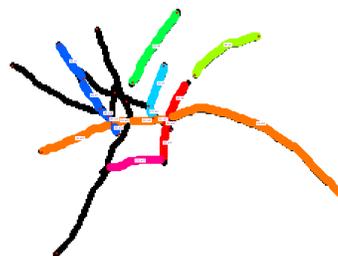
(e) Mejor resultado de la individualización de filamentos usando el algoritmo propuesto con la configuración para microtúbulos de planta.



(f) 8 filamentos correctamente individualizados en base al resultado de (e).

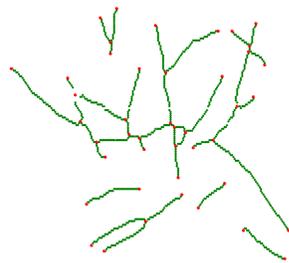


(g) Resultado promedio de la individualización de filamentos usando el algoritmo propuesto con la configuración para microtúbulos de planta.

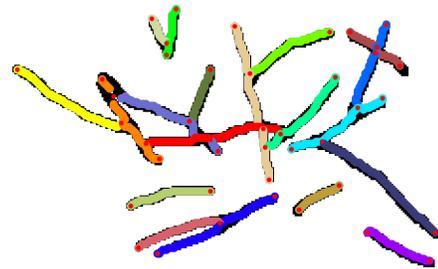


(h) 7 filamentos correctamente individualizados en base al resultado de (g).

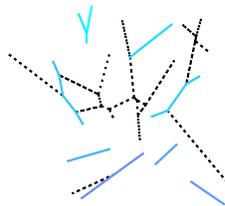
Figura 6.4: Filamentos propuestos y correctamente individualizados a partir la muestra MT-A en la Figura 5.3, obtenidos mediante DeFiNe y el algoritmo propuesto con sus respectivos parámetros. Segmentos marcados en negro representan aristas no asignadas correctamente al filamento correspondiente en el *ground truth*, mientras que los filamentos correctamente identificados se encuentran identificados mediante colores.



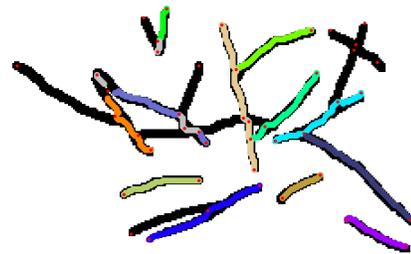
(a) Grafo que representa la red de filamentos de la muestra MT-A2.



(b) Individualización manual por un experto de la muestra MT-A2.

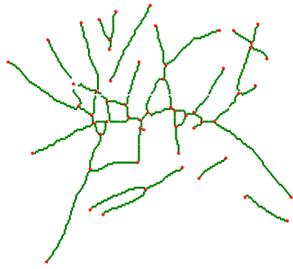


(c) Representación de 9 filamentos correctamente individualizados por el mejor resultado de DeFiNe, correspondiente a la configuración DeFiNe 60°, identificados mediante colores.

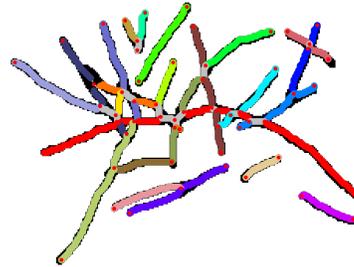


(d) Mejor resultado del algoritmo propuesto, con la configuración MtP, individualizando 12 filamentos correctamente.

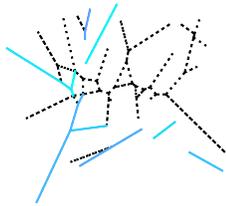
Figura 6.5: Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.3b (Muestra MT-A2). El algoritmo propuesto individualiza correctamente 12 filamentos con la configuración para microtúbulos de planta, en un menor tiempo.



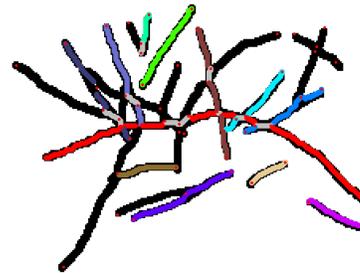
(a) Grafo de 59 aristas que representa la red de filamentos de la muestra MT-A3.



(b) Individualización manual por un experto de la muestra MT-A3.

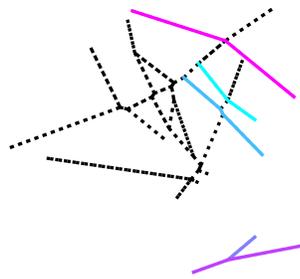


(c) Representación de 9 filamentos correctamente individualizados por el mejor resultado de DeFiNe, correspondiente a la configuración DeFiNe 30°, identificados mediante colores.

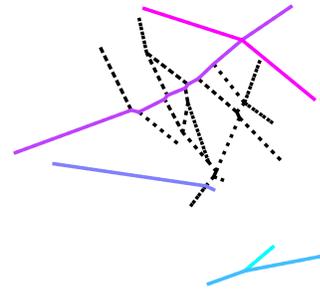


(d) Mejor resultado del algoritmo propuesto, con la configuración MtP, individualizando 12 filamentos correctamente.

Figura 6.6: Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.3c (Muestra MT-A3). El algoritmo propuesto individualiza correctamente 12 de 23 filamentos con la configuración para microtúbulos de planta,



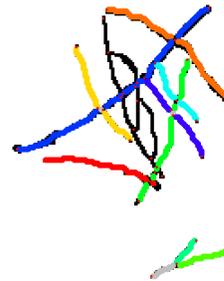
(a) Representación de los 5 filamentos correctamente individualizados por DeFiNe con 30° , identificados con colores



(b) Representación de los 5 filamentos correctamente individualizados por DeFiNe con 60° , identificados con colores

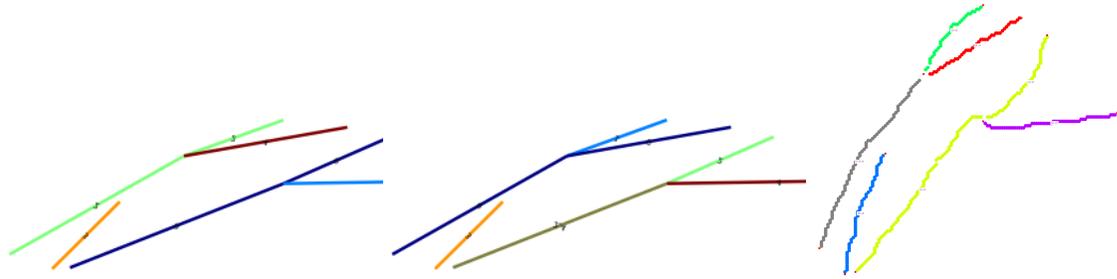


(c) Resultado promedio del algoritmo propuesto utilizando la configuración MtP, con 8 filamentos correctamente individualizados, identificados con colores.

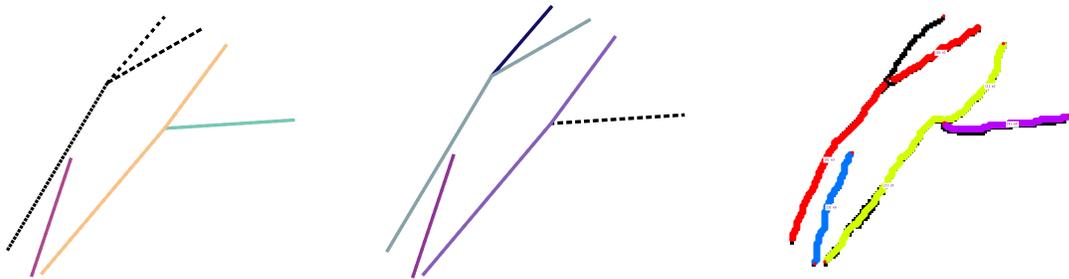


(d) 9 Filamentos correctamente individualizados a partir del mejor resultado del algoritmo propuesto con la configuración MtP, identificados con colores.

Figura 6.7: Filamentos correctamente individualizados en la figura 5.4 correspondiente a la muestra MT-B, mediante el DeFiNe y el algoritmo propuesto. Segmentos marcados en negro representan aristas no asignadas correctamente. El mejor resultado del algoritmo propuesto permite encontrar filamentos no encontrados en a) o b).



(a) Individualización mediante DeFiNe con 30° (b) Individualización mediante DeFiNe con 60° (c) Individualización de filamentos obtenida mediante el algoritmo propuesto, con los parámetros de microtúbulos de planta. El segmento plomo indica una arista seleccionada por múltiples filamentos.



(d) Representación de los filamentos correctamente individualizados en (a), identificados con colores (e) Representación de los filamentos correctamente individualizados en (b), identificados con colores (f) Filamentos correctamente individualizados en (c), identificados con colores

Figura 6.8: Comparación entre filamentos propuestos y correctamente individualizados para la muestra MT-C en la Figura 5.5, con los distintos métodos y parámetros. Segmentos marcados en negro representan aristas no asignadas correctamente al filamento correspondiente en el *ground truth*. A pesar de la igualdad en resultados entre (e) y (f), el algoritmo propuesto lo obtiene en menor tiempo.

Muestra MT-D

La muestra MT-D corresponde a la imagen en la Figura 6.9a, en la que un análisis experto identifica 13 filamentos, y de la que se extrae un grafo de 19 aristas. Al realizar la individualización de filamentos mediante la herramienta DeFiNe con la configuración de 60° , se obtienen 12 filamentos correctamente identificados, siendo el mejor resultado entre las 2 configuraciones de DeFine y las 10 iteraciones del algoritmo propuesto. Por su parte, el algoritmo propuesto obtiene solo 11 de los 13 filamentos indicados por el experto, pudiendo encontrar un filamento que no encuentra DeFiNe con 30° , pero no encontrando 2 filamentos que si encuentra esa herramienta.

Estos 2 filamentos no encontrados corresponden a casos simples, donde la heurística del algoritmo propuesto para construir soluciones prefiere una solución más compleja que se aleja de lo indicado por el experto. Por otra parte, el filamento encontrado solamente por el algoritmo propuesto, indicado en naranja en la Figura 6.9d, es individualizado correctamente a pesar de tener una arista bastante corta, lo cual puede introducir problemas en el calculo de ángulos entre aristas.

Debido a que el grafo utilizado para individualizar filamentos no contiene ciclos así como presenta una baja cantidad de cruces o intersecciones, los tiempos de computo de los métodos evaluados son inferiores al segundo, como se indica en la tabla 6.8.

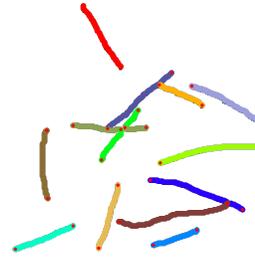
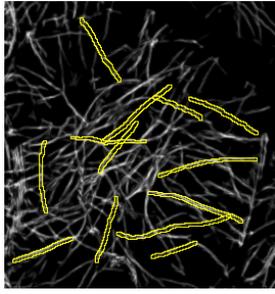
Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.66	-	0.72	-	0.69	-	12/14	-	12/13	-	0.4
DeFiNe 60°	1	0.35	-	0.3	-	0.32	-	9/13	-	9/13	-	0.5
AP MtP	10	0.46	0.46	0.28	0.26	0.35	0.33	10.4/14	11/14	10.4/13	11/13	0.8

Tabla 6.8: Resultados de individualización de filamentos para Figura 6.9a, de la muestra MT-D. Existen 19 aristas en el grafo que representa la red de filamentos, y a su vez el número de filamentos correctos es 13. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. Para esta muestra, el método DeFiNe con la configuración de 30° logra identificar un filamento más que el mejor resultado del algoritmo propuesto.

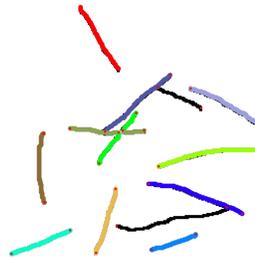
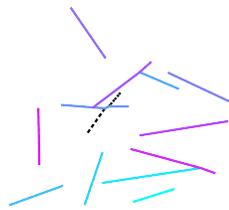
Muestra MT-E

Una segunda muestra adicional a lo indicado en la sección 5.3 corresponde a la imagen en la Figura 6.10a, en la cual un experto individualiza 12 filamentos. El grafo asociado a la red de filamentos de esta muestra esta compuesto por 23 aristas, a partir de las que la herramienta DeFiNe individualiza 8 filamentos correctamente, mientras que el mejor resultado del algoritmo propuesto logra el mejor resultado, con un resultado promedio de las 10 iteraciones de 7.9 filamentos correctamente identificados.

En cuanto al número de filamentos propuestos por cada método, la configuración DeFiNe



(a) Imagen de microtúbulos de la planta *Arabidopsis Marchantia* con 13 filamentos individualizados por un experto, señalado mediante las MT-D. Esta imagen se identifica como la muestra MT-D.



(c) Representación de 12 filamentos correctamente individualizados por el mejor resultado de DeFiNe, correspondiente a la configuración DeFiNe 30°, identificados mediante colores.

(d) Mejor resultado del algoritmo propuesto para la muestra MT-D, con la configuración MtP, individualizando 11 filamentos correctamente.

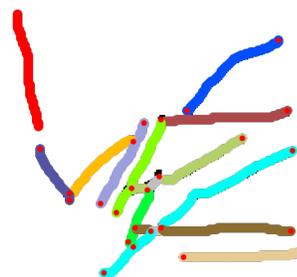
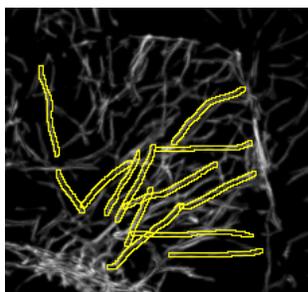
Figura 6.9: Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.9a (Muestra MT-D). El mejor resultado del algoritmo propuesto encuentra un filamento menos que la mejor configuración de DeFiNe. Cada método identifica al menos un filamento que no es encontrado por el otro método, debido a la diferencia en como realizan la individualización.

60° indica que en la red de filamentos existen 13 filamentos (8 de los cuales son correctos), en tanto que el algoritmo propuesto indica en su mejor resultado que son 12 filamentos (8 de los cuales son correctos). Es decir, el comportamiento de ambos métodos es similar, lo que puede ser atribuible a que esta muestra no presenta mayores complejidades en el grafo que representa la red de filamentos.

Una diferencia en la forma en que ambos métodos individualizan filamentos puede observarse a través del filamento identificado por el algoritmo propuesto, de color café oscuro, en la Figura 6.10d. La herramienta DeFiNe, al privilegiar el criterio angular entre aristas del grafo no encuentra aquel filamento. De forma inversa, filamentos simples no son encontrados por el algoritmo propuesto lo cual puede asociarse a una mayor complejidad de las heurísticas y lógicas utilizadas.

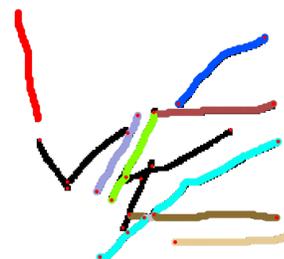
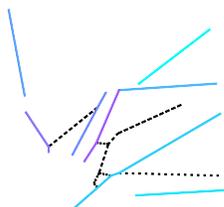
Config	Iters	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	T[s]
DeFiNe 30°	1	0.51	-	0.3	-	0.38	-	8/15	-	8/12	-	0.7
DeFiNe 60°	1	0.48	-	0.37	-	0.42	-	8/13	-	8/12	-	1.0
AP MtP	10	0.44	0.45	0.35	0.48	0.39	0.47	7.9/13.1	8/12	7.9/12	8/12	0.9

Tabla 6.9: Resultados de individualización de filamentos para la muestra MT-E en la Figura 6.10a. El número de filamentos en el *ground truth* es 12, en un grafo de 23 aristas. La columna Iters indica el número de ejecuciones de cada configuración, las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la columna T indica el tiempo de ejecución. (*) Las columnas con asterisco representan el mejor resultado obtenido en la configuración con más de una iteración, mientras que las columnas sin asterisco representan el promedio. Tanto el algoritmo propuesto (AP) como DeFiNe presentan resultados similares, lo que puede atribuirse a que la muestra no presenta una mayor complejidad.



(a) Imagen de microtúbulos de la planta *Ara-* (b) Representación coloreada de la individuali-
bidopsis Marchantia con 12 filamentos indivi- zación manual de un experto para la muestra
dualizados por un experto, señalado mediante MT-E.
las áreas de interes en amarillo. Esta imagen se
identifica como la muestra MT-E.

las áreas de interes en amarillo. Esta imagen se
identifica como la muestra MT-E.



(c) Representación de 8 filamentos correcta-
mente individualizados por ambas configura-
ciones utilizadas con la herramienta DeFiNe, iden-
tificados mediante colores.

(d) Mejor resultado del algoritmo propuesto pa-
ra la muestra MT-E, con la configuración MtP,
individualizando 8 filamentos correctamente.

Figura 6.10: Individualización de filamentos correctos respecto a la evaluación de un experto para la Figura 6.10a (Muestra MT-E). El algoritmo propuesto (AP) y DeFiNe presentan un comportamiento similar, atribuible a la simpleza de la muestra.

6.3.3. Neuronas

En esta sección se muestran los resultados de la evaluación del algoritmo propuesto para las 3 muestras de neurona, obtenidas desde Ampuero et al. [43], y que se denominan N1, N2 y N3. Para estas 3 muestras se utiliza la configuración predeterminada de neurona (AP N), ejecutando el algoritmo propuesto 10 veces para obtener un comportamiento promediado. A diferencia de la sección 6.3.2, en estas evaluaciones las pruebas ejecutadas con DeFiNe presentan cantidades de filamentos propuestos 2 a 6 veces mayores que el número de filamentos individualizados por un experto. Este comportamiento puede atribuirse a la obligación que DeFiNe tiene de asignar todas las aristas al menos a un filamento. En comparación, el algoritmo propuesto al tener mayor flexibilidad, asigna en promedio un 57% de las aristas a filamentos.

Adicionalmente, los tiempos de ejecución de las pruebas con DeFiNe son sustancialmente mayores, encontrándose en el rango de los minutos a las 4 horas, dependiendo de los parámetros utilizados. Lo anterior puede asociarse al número de aristas que las muestras de neurona tienen en su respectivo grafo, las que son 414 para N1 y 161 para N2. En el caso de la muestra N3, representada por un grafo de 67 aristas, esta no pudo ser calculada con DeFiNe ya que el programa arroja un error de falla crítica, impidiendo su ejecución. La información respecto al número de filamentos propuestos como de los tiempos de ejecución se encuentra en la Tabla 6.10.

Muestra	Algoritmo	Fil. Propuestos	% Asignación	Tiempo[s]
N1	DeFiNe 30°	246	100	1514.6
	DeFiNe 60°	192	100	15573.7
	AP N Promedio	59	53.7	32.5
N2	DeFiNe 30°	113	100	82.2
	DeFiNe 60°	85	100	456.4
	AP N Promedio	34.8	59.3	4.9
N3	AP N Promedio	17.4	57.8	4.2

Tabla 6.10: Número de filamentos propuestos y tiempos de ejecución para DeFiNe y el promedio del algoritmo propuesto con configuración de neurona (AP N), para las muestras N1 y N2 en las Figuras 5.6 y 5.7 respectivamente. Estas muestras contienen 24 y 29 filamentos cada una. La muestra N3 en la figura 5.8 es representada por un grafo de 67 aristas y solo considera los resultados promediados del algoritmo propuesto.

Así, los motivos anteriores impiden realizar una evaluación con respecto al criterio experto, por lo que no se consideran los resultados de DeFiNe en esta sección.

En cuanto a los resultados del algoritmo propuesto, indicados en la Tabla 6.11, a pesar de ejecutar todas las evaluaciones en tiempos menores a 40 segundos, las individualizaciones correctas son bajas, mientras que el número de filamentos propuestos se acerca a la cantidad de filamentos individualizados manualmente por un experto, exceptuando el caso de la Muestra N1. Es posible asociar un comportamiento razonable del número de filamentos propuestos con la restricción especial para neuronas considerada dentro de las penalización de anti-feromonas, lo que permite mantener acotado el número de filamentos propuestos mediante el descarte de soluciones infactibles. Sin embargo, el bajo número de aciertos entre los

filamentos propuestos refleja que la heurística que influye en la probabilidad de elección de una arista durante la construcción de caminos, indicada en la ecuación 4.2, no considera todas las características necesarias para explorar exitosamente el grafo, en el caso de neuronas.

Como se menciona en las evaluaciones previas, una de las mediciones principales es encontrar un calce exacto entre los filamentos propuestos y los que identifica el experto. Es por esto, que los filamentos que se sobre o sub asignan en 1 arista o más no son considerados. Sin embargo, esta información puede ser útil para entender el motivo por el cual un filamento propuesto queda tan cerca de ser un calce exacto con respecto a lo indicado por un experto. Esta información adicional se encuentra en la columna F.S de la Tabla 6.11, indicando que aún cuando los calces exactos son bajos, existen varios filamentos propuestos que se encuentran en un rango de 1 a 3 aristas faltantes o sobrantes, con respecto a un filamento correcto. Una comparación visual entre filamentos propuestos y correctos para las pruebas de esta sección se encuentra en la Figura 6.11.

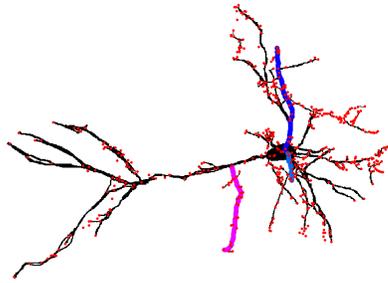
En las neuronas es observable la existencia de una dendrita de mayor longitud que el resto, denominada axón, a partir de la que nacen nuevas dendritas. A su vez, otras dendritas que nacen del centro de la neurona o soma, y se extienden de forma significativa también sirven como lugar de nacimiento para nuevas dendritas. Este comportamiento implica la necesidad de considerar la selección de aristas que no solo respeten el criterio de rectitud, sino que a la vez aporten con la mayor longitud al camino en construcción.

Muestra	P	P*	R	R*	F1	F1*	C/P	C/P*	C/GT	C/GT*	F.S.	F.S.*
N1	0.62	0.6	0.15	0.13	0.24	0.22	1.6/59	3/66	1.6/24	3/24	1.4	1
N2	0.2	0.21	0.09	0.1	0.13	0.14	4.6/34.8	6/34	4.6/29	6/29	8	7
N3	0.5	0.47	0.32	0.29	0.39	0.36	2/17.4	2/17	2/14	2/14	7.8	8

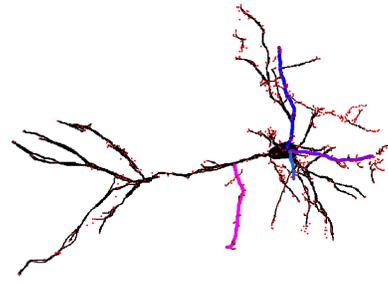
Tabla 6.11: Resultados de la individualización de filamentos mediante el algoritmo propuesto con los parámetros para neuronas, para las muestra N1, N2 y N3, en las Figuras 5.6, 5.7 y 5.8 respectivamente. Las columnas P y R representan *Precision* y *Recall* respectivamente, la columna C/P refleja el número de filamentos correctos con respecto a los propuestos por cada método, mientras que la columna C/GT indica la relación entre los filamentos correctamente individualizados por el método y el criterio del experto. Finalmente la La columna F.S. indica el número de filamentos cuya diferencia con respecto a ser un calce exacto es a lo más de 3 aristas. (*) Las columnas con asterisco representan el máximo obtenido tras las 10 ejecuciones del algoritmo propuesto, mientras que las sin asterisco representan el promedio.

6.3.4. Métrica VI e índices Rand y Jaccard

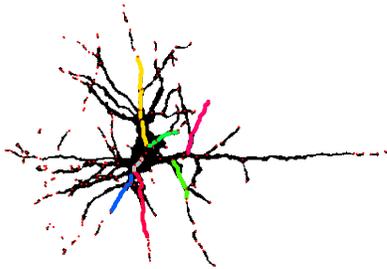
Un análisis particular recae en las métricas y medidas utilizadas para comparar los resultados obtenidos, en imágenes sintéticas y reales, dado que en el caso de la métrica VI y los índices Rand y Jaccard, se pueden obtener resultados que no permiten concluir cual es la calidad del resultado obtenido. Lo anterior se puede observar en la Tabla 6.12, en la que los resultados del índice Rand aparecen como de buena calidad, mientras que lo contrario sucede con lo indicado por el índice Jaccard. Paralelamente, la métrica VI no refleja correctamente los casos donde hay superposición de filamentos. En base a lo anterior, las otras medidas utilizadas como *Precision*, *Recall*, y la medida F1 que se deriva de las 2 anteriores, son las que pueden entregar mayor claridad del comportamiento de los métodos probados, ya que



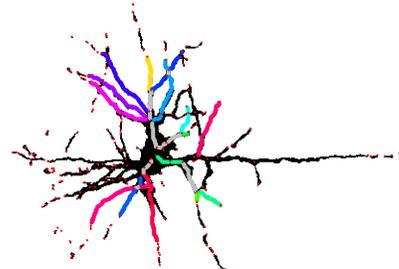
(a) Filamentos correctamente individualizados por el algoritmo propuesto, para la muestra N1 en la Figura 5.6, identificados por colores.



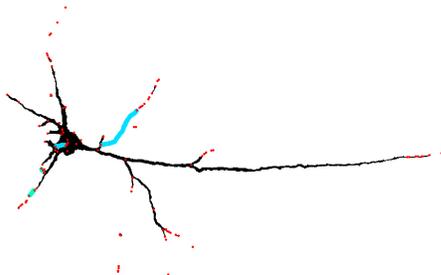
(b) Filamentos correctamente individualizados en (a) más un filamentos sobre o sub asignados por 1 arista, a partir del resultado del algoritmo propuesto.



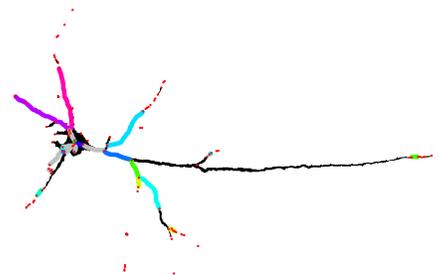
(c) Filamentos correctamente individualizados por el algoritmo propuesto, para la muestra N2 en la Figura 5.7, identificados por colores.



(d) Filamentos correctamente individualizados en (c) más los filamentos sobre o sub asignados entre 1 a 3 aristas, a partir del resultado del algoritmo propuesto.



(e) Filamentos correctamente individualizados por el algoritmo propuesto, para la muestra N3 en la Figura 5.8, identificados por colores.



(f) Filamentos correctamente individualizados en (e) más los filamentos sobre o sub asignados entre 1 a 3 aristas, a partir del resultado del algoritmo propuesto.

Figura 6.11: Filamentos propuestos e individualizados correctamente para las muestras N1, N2 y N3, por el algoritmo propuesto. Se puede observar en los filamentos propuestos que existe superposición en las zonas grises, lo que puede identificarse como dendritas que nacen a partir de otras de mayor longitud. Al incluir filamentos sobre o sub asignados en hasta 3 aristas, se puede observar que para las muestras N2 y N3 existen varios filamentos propuestos cerca de ser calces exactos con los definidos por un experto.

permiten observar la individualización de filamentos como un problema de clasificación de aristas en uno o más categorías, siendo cada categoría un filamento individualizado por un

experto.

Figura	Config. de Parámetros	VI Max	VI	Rand	Jaccard
5.1	MtP	2.3978	1.6360	0.7646	0.2485
5.1	S1	2.3978	0.7091	0.8637	0.4750
5.2	MtP	3.0445	2.2296	0.7276	0.1806
5.2	S2	3.0445	2.5878	0.7276	0.1673
5.3	MtP	3.4965	2.1656	0.8658	0.2407
5.4	MtP	3.7612	2.6285	0.8683	0.2488
5.5	MtP	1.9459	0.4286	0.8929	0.40
5.6	N	6.0258	1.7950	0.8864	0.1389
5.7	N	5.0814	3.7256	0.8775	0.0703
5.8	N	4.2046	1.0060	0.8794	0.2157
6.3b	MtP	3.6635	1.4749	0.9305	0.2781
6.3c	MtP	4.0775	3.5850	0.8994	0.1331
6.9a	MtP	2.9444	1.1781	0.8739	0.2144
6.10a	MtP	3.1354	1.4533	0.2446	0.4477

Tabla 6.12: Resultados de la métrica VI y los índices Rand y Jaccard para las evaluaciones del algoritmo propuesto en imágenes sintéticas y reales, para diferentes configuraciones de parámetros. Los valores de los índices Rand y Jaccard se encuentran en el rango $[0, 1]$, siendo 1 el valor ideal. Se observa que los resultados del índice Rand se contradicen con los del índice Jaccard. Por su parte, la métrica VI penaliza los casos en que una aristas pertenece a más de un filamento propuesto, lo que dificulta la evaluación mediante esta métrica para los casos de superposición de filamentos.

6.4. Resultados Generales

Con la finalidad de comprender el comportamiento general del algoritmo propuesto, se realiza una evaluación estadística de los resultados para determinar la similitud entre el número de filamentos propuestos con respecto a la cantidad de filamentos indicada por un experto. Para realizar la medición, se consideran las 10 iteraciones de cada figura mediante el algoritmo propuesto. Si estos resultados se comportan de acuerdo a una distribución normal, se evalúa su similitud con respecto a lo indicado por un experto a través de la prueba t de Student, de lo contrario se utiliza una prueba de rangos con signo de Wilcoxon. La hipótesis nula (H_0) consiste en que no existe diferencia estadísticamente significativa entre el número de filamentos individualizados por el algoritmo propuesto y lo definido por el experto que realizó la individualización manual respectiva. En el caso de que el valor p o p -value sea menor al 5%, se rechaza H_0 .

En la Tabla 6.13 se muestran los resultados del análisis estadístico, obteniendo un p -value mayor que el 5% en 3 figuras, para las que se puede mantener la hipótesis nula. En los casos en que existe una diferencia estadísticamente significativa, es posible observar que los resultados se encuentran en su mayoría cercanos a la cantidad esperada.

En base a los resultados se observa que el comportamiento del algoritmo propuesto es estable, tendiendo a individualizar los mismos filamentos en cada iteración, y como se indica

Figura	Configuración	Filamentos Correctos	Promedio Filamentos Propuestos 10 Iters	Distribución Normal	p-value
5.1	S1	6	5.9	No	1
5.2	S2	5	8.5	No	0.003
5.3	MtP	12	12.4	No	0.125
5.4	MtP	12	15	Si	0
5.5	MtP	5	5	No	1
5.6	N	24	60.2	Si	0
5.7	N	29	34.3	Si	0
5.8	N	14	17.6	No	0.003
6.3b	MtP	19	21	No	0.001
6.3c	MtP	23	29.4	Si	0
6.9a	MtP	13	14	No	0.001
6.10a	MtP	12	13.1	Si	0

Tabla 6.13: Análisis estadístico del número de filamentos propuestos en cada figura, al realizar 10 iteraciones con su respectiva configuración de parámetros, del algoritmo propuesto. Se obtiene en una figura sintética y en dos figuras de microtúbulos de planta que no existe diferencia significativa entre el número de filamentos propuestos con respecto a lo identificado por un experto ($p\text{-value} > 0.05$).

en las Tablas 6.2 y 6.6, encontrando filamentos que no aparecen en DeFiNe. Se debe tener en consideración que el último aspecto puede relacionarse con el uso de una ponderación distinta a la ideal presentada por los autores de DeFiNe, que puede ocasionar la merma en los resultados en aquel método. La estabilidad del algoritmo propuesto, en conjunto con los parámetros predefinidos apunta a favorecer la experiencia del usuario, ya que evita la sintonización de parámetros, simplificando el trabajo del experto. Además, la utilización de múltiples características en el algoritmo propuesto permite al usuario ajustar parámetros específicos, así como integrar nuevos parámetros para manejar información adicional que pueda ser incorporada. Un aspecto adicional del algoritmo propuesto es el uso de la curvatura de las aristas, en comparación a representarlas mediante una línea recta entre 2 nodos. Esto no solo aporta más información, sino que además mejora la visualización de los resultados.

En resumen, los resultados obtenidos al individualizar filamentos a partir de imágenes sintéticas y reales indican que en casos simples existe un comportamiento similar entre el algoritmo propuesto y DeFiNe, marcando una diferencia a favor del algoritmo propuesto a medida que aumenta la complejidad, tanto en tiempo de resolución como en el menor número de filamentos propuestos. La diferencia entre los tiempos de calculo y la cantidad de filamentos propuestos de cada método puede ser atribuida a la forma en que se explora el espacio de búsqueda, así como a la suavización de la restricción de asignación de cada arista a un filamento propuesto, lo que permite al algoritmo propuesto encontrar filamentos que en algunos casos no son obtenidos por el otro método. Otro aspecto relacionado a la flexibilidad del algoritmo propuesto se refleja la posibilidad de incorporar información adicional, al basarse en la metaheurística ACO.

Capítulo 7

Conclusiones

La individualización automática de filamentos es un problema que se encuentra limitado por la extracción de información que se pueda realizar desde una imagen. Entre las limitantes de este paso previo se puede considerar: (1) la resolución y el ruido de la imagen, (2) la capacidad de la herramienta que genera un grafo a partir de la misma imagen, de conservar la estructura topológica con la menor cantidad de discontinuidades.

La obtención de múltiples características de los filamentos es crítica para descartar rápidamente soluciones de baja calidad, algo especialmente necesario en un problema en el que la combinación de aristas que conforman un camino que representa un filamento puede crecer exponencialmente. Estas características pueden provenir no sólo de aspectos geométricos de los filamentos, sino que también de información topológica y/o espacial de los mismos. Aún con aquella información que permite explorar de mejor forma las combinaciones de aristas contiguas o caminos que representan un filamento en un grafo, la información de la célula observada disponible *a priori*, como el grado máximo de los nodos o la presencia de ciclos en los filamentos, facilita el establecimiento de umbrales y/o cotas que describen parcialmente aspectos del comportamiento de los filamentos.

El análisis derivado de las múltiples características se enfoca en obtener información cuantitativa respecto al comportamiento de los filamentos, marcando una diferencia con otras investigaciones con análisis de tipo cualitativo. De forma similar, el algoritmo propuesto cuenta con una estructura flexible que permite incorporar información y/o lógicas adicionales, las que dependiendo de la etapa en la metaheurística ACO en que se encuentren, pueden variar la forma en que se explora el espacio de búsqueda o en que se aplican restricciones a los resultados parciales obtenidos. Otro aspecto relativo a la flexibilidad del algoritmo propuesto radica en contar con parámetros predefinidos dependiendo del tipo de célula, permitiendo la personalización de estos parámetros por parte del experto, lo que busca facilitar el análisis y a su vez permite automatizar la extracción de información para los mismos tipos de células.

Los resultados obtenidos en esta investigación incorporan los elementos mencionados previamente, permitiendo individualizar filamentos de forma similar o mejor con respecto a métodos en el estado del arte. A su vez, los resultados permiten validar la hipótesis de identificar cada filamento resolviendo un modelo de optimización. A pesar de no lograr una

individualización correcta de la totalidad de los filamentos, se obtiene un comportamiento estable en las distintas iteraciones de cada evaluación, así como se logra un número de filamentos propuestos que en la mayoría de las ocasiones se acerca al número de filamentos indicado por el experto. Cabe volver a destacar lo mencionado en el capítulo 3.3 en relación a que es posible que 2 expertos indiquen diferentes filamentos a partir de la misma imagen, por lo que es necesario incorporar un mayor número de pruebas a este trabajo para descartar la existencia de un sobre-ajuste o sesgo del algoritmo propuesto a la información provista por los expertos que realizaron las individualizaciones manuales de las imágenes utilizadas.

De forma similar, mientras que la flexibilidad del algoritmo propuesto y el uso de 2 tipos de células permite disminuir algún posible sesgo con respecto a los datos utilizados para desarrollar o evaluar el algoritmo propuesto, es necesario realizar evaluaciones adicionales que permitan generalizar el comportamiento del algoritmo propuesto ante otros tipos de células que presentan filamentos.

Con respecto al uso de la metaheurística ACO como base del algoritmo propuesto, esta propone una diferencia con respecto del planteamiento de PCP o FCP, el que obliga a cada arista a pertenecer al menos a un camino. En cambio, el algoritmo propuesto relaja aquella restricción dado que pueden existir aristas aisladas producto una extracción de grafo con pérdida de información, ocasionando discontinuidades. La captación de aquella información es relevante dado que existen métodos en el estado del arte que permiten realizar la unión de una arista aislada y el filamento al que potencialmente puede pertenecer, pudiendo incorporarse como etapas adicionales al modelo.

En general, el algoritmo propuesto permite resolver un modelo de optimización para individualizar filamentos a partir de un grafo con pesos que representa una red de filamentos, cumpliendo con lo indicado como primer objetivo específico. En cuanto al cumplimiento de los demás objetivos específicos, se tiene:

- **Implementar un algoritmo que resuelva el modelo de optimización, entregando como salida la identificación de filamentos, considerando casos de solapamiento y/o cruce:** el algoritmo propuesto cumple con este objetivo, de acuerdo a lo presentado en los capítulos 4 y 6.
- **Identificar la ponderación de propiedades que entregue mejores resultados para grafos que representen una neurona, una bacteria y una célula eucariota de planta:** se realiza una ponderación de las propiedades utilizadas en el algoritmo propuesto en la sección 4.7, la que varía dependiendo del tipo de célula y se encuentra asociado a la distribución de las características en las distintas etapas de la metaheurística ACO.
- **Evaluar técnicas en el estado del arte que realicen individualización de filamentos:** entre Las técnicas descritas en el capítulo 3, solo el método desarrollado por (Breuer and Nikoloski [2]) realiza individualización de filamentos, siendo el método evaluado, en conjunto con el algoritmo propuesto, en el capítulo 6.

Como trabajo futuro de esta investigación se encuentra la opción de intensificar la extracción de información topológica y espacial que permita realizar la asociación de secciones importantes de los filamentos o elementos alrededor de estos en la célula observada, con el

fin de seguir añadiendo información al modelo. También es posible vislumbrar una extensión de este trabajo en la identificación de las interacciones entre los filamentos individualizados.

Bibliografía

- [1] J. Qiu and F. LI, “Quantitative morphological analysis of curvilinear network for microscopic image based on individual fibre segmentation (ifs),” *Journal of microscopy*, vol. 256, no. 3, pp. 153–165, 2014.
- [2] D. Breuer and Z. Nikoloski, “Define: an optimisation-based method for robust disentangling of filamentous networks,” *Scientific reports*, vol. 5, no. 1, pp. 1–14, 2015.
- [3] P. Asgharzadeh, B. Özdemir, R. Reski, O. Röhrle, and A. I. Birkhold, “Computational 3d imaging to quantify structural components and assembly of protein networks,” *Acta biomaterialia*, vol. 69, pp. 206–217, 2018.
- [4] Z. Zhang, Y. Nishimura, and P. Kanchanawong, “Extracting microtubule networks from superresolution single-molecule localization microscopy data,” *Molecular biology of the cell*, vol. 28, no. 2, pp. 333–345, 2017.
- [5] A. Boudaoud, A. Burian, D. Borowska-Wykręt, M. Uyttewaal, R. Wrzalik, D. Kwiatkowska, and O. Hamant, “Fibriltool, an imagej plug-in to quantify fibrillar structures in raw microscopy images,” *Nature protocols*, vol. 9, no. 2, pp. 457–463, 2014.
- [6] M. Louveaux and A. Boudaoud, “FibrilTool Batch: an automated version of the ImageJ/Fiji plugin FibrilTool,” Dec. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.2528872>
- [7] M. Alioscha-Perez, C. Benadiba, K. Goossens, S. Kasas, G. Dietler, R. Willaert, and H. Sahli, “A robust actin filaments image analysis framework,” *PLoS computational biology*, vol. 12, no. 8, p. e1005063, 2016.
- [8] I. Usov and R. Mezzenga, “Fiberapp: An open-source software for tracking and analyzing polymers, filaments, biomacromolecules, and fibrous objects,” *Macromolecules*, vol. 48, no. 5, pp. 1269–1280, 2015. [Online]. Available: <https://doi.org/10.1021/ma502264c>
- [9] N. Lichtenstein, B. Geiger, and Z. Kam, “Quantitative analysis of cytoskeletal organization by digital fluorescent microscopy,” *Cytometry Part A*, vol. 54, no. 1, pp. 8–18, 2003.
- [10] M. Cerda, J. Toledo, and S. Hartel, “Geometrical constraints to improve elongated structures segmentation applied to actin filaments,” *IET Conference Proceedings*, pp. 4–8, 2014. [Online]. Available: <https://digital-library.theiet.org/content/conferences/10>.

- [11] T. Xu, D. Vavylonis, F.-C. Tsai, G. H. Koenderink, W. Nie, E. Yusuf, I.-J. Lee, J.-Q. Wu, and X. Huang, “Soax: a software for quantification of 3d biopolymer networks,” *Scientific reports*, vol. 5, p. 9081, 2015.
- [12] P. Llanos, “Analysis and modeling of microtubule self-organization,” Ph.D. dissertation, School of Medicine, Univ. of Chile, Independencia, Chile, 2020.
- [13] T. Xu, C. Langouras, M. A. Koudehi, B. E. Vos, N. Wang, G. H. Koenderink, X. Huang, and D. Vavylonis, “Automated tracking of biopolymer growth and network deformation with tsoax,” *bioRxiv*, p. 316489, 2018.
- [14] L. Pizarro, “Algoritmo para individualizar filamentos,” <https://gitlab.com/LeoXDXp/graph-crawler/>, 2021, [Online; accessed 29-May-2021].
- [15] A. Caprara, P. Toth, and M. Fischetti, “Algorithms for the set covering problem,” *Annals of Operations Research*, vol. 98, no. 1-4, pp. 353–371, 2000.
- [16] S. C. Ntafos and S. L. Hakimi, “On path cover problems in digraphs and applications to program testing,” *IEEE Transactions on Software Engineering*, no. 5, pp. 520–529, 1979.
- [17] G. Lin, Z. Cai, and D. Lin, “Vertex covering by paths on trees with its applications in machine translation,” *Information processing letters*, vol. 97, no. 2, pp. 73–81, 2006.
- [18] A. Lavado, “Comparación de algoritmos de cálculo del skeleton y su aplicación en biología,” M.S. thesis, Dept. Comp. Sci., Univ. of Chile, Santiago, Chile, 2018.
- [19] P. K. Saha, G. Borgefors, and G. S. di Baja, “Skeletonization and its applications—a review,” in *Skeletonization*. Elsevier, 2017, pp. 3–42.
- [20] J. Nunez-Iglesias, A. J. Blanch, O. Looker, M. W. Dixon, and L. Tilley, “A new python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton,” *PeerJ*, vol. 6, p. e4312, 2018.
- [21] A. Wang, X. Yan, and Z. Wei, “Imagepy: an open-source, python-based and platform-independent software package for bioimage analysis,” *Bioinformatics*, vol. 34, no. 18, pp. 3238–3240, 2018.
- [22] A. Hagberg, P. Swart, and D. S Chult, “Exploring network structure, dynamics, and function using networkx,” Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.
- [23] J. Flusser, B. Zitova, and T. Suk, *Moments and moment invariants in pattern recognition*. John Wiley & Sons, 2009.
- [24] F. Chaumette, “Image moments: a general and useful set of features for visual servoing,” *IEEE Transactions on Robotics*, vol. 20, no. 4, pp. 713–723, 2004.

- [25] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [26] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *Medical Image Computing and Computer-Assisted Intervention — MICCAI’98*. Springer, 1998, pp. 130–137.
- [27] W. Fu, K. Breininger, R. Schaffert, N. Ravikumar, T. Würfl, J. Fujimoto, E. Moulton, and A. Maier, “Frangi-net,” in *Bildverarbeitung für die Medizin 2018*. Springer, 2018, pp. 341–346.
- [28] R. J. Wilson, *Introduction to graph theory*. Pearson Education India, 1979.
- [29] Y. Liu, B. Cao, and H. Li, “Improving ant colony optimization algorithm with epsilon greedy and levy flight,” *JSP*, vol. 24, no. 25, p. 54, 2020.
- [30] K. Socha and M. Dorigo, “Ant colony optimization for continuous domains,” *European journal of operational research*, vol. 185, no. 3, pp. 1155–1173, 2008.
- [31] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, “Hybrid metaheuristics in combinatorial optimization: A survey,” *Applied Soft Computing*, vol. 11, no. 6, pp. 4135–4151, 2011.
- [32] K. Buchin, C. Knauer, K. Kriegel, A. Schulz, and R. Seidel, “On the number of cycles in planar graphs,” in *Computing and Combinatorics*. Springer, 2007, pp. 97–107.
- [33] S. Biswas, S. Durocher, D. Mondal, and R. I. Nishat, “Hamiltonian paths and cycles in planar graphs,” in *Combinatorial Optimization and Applications*. Springer, 2012, pp. 83–94.
- [34] J. Montgomery and M. Randall, “Anti-pheromone as a tool for better exploration of search space,” in *International Workshop on Ant Algorithms*. Springer, 2002, pp. 100–110.
- [35] S. Stam, S. L. Freedman, S. Banerjee, K. L. Weirich, A. R. Dinner, and M. L. Gardel, “Filament rigidity and connectivity tune the deformation modes of active biopolymer networks,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 47, pp. E10 037–E10 045, 2017.
- [36] Ö. Civalek and Ç. Demir, “Bending analysis of microtubules using nonlocal euler–bernoulli beam theory,” *Applied Mathematical Modelling*, vol. 35, no. 5, pp. 2053–2067, 2011.
- [37] M. Li, “Efficiency improvement of ant colony optimization in solving the moderate lisp,” *Journal of Systems Engineering and Electronics*, vol. 26, no. 6, pp. 1300–1308, 2015.
- [38] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge University Press, 2008. [Online]. Available: <https://cds.cern.ch/record/2135372>

- [39] M. Meilă, “Comparing clusterings—an information based distance,” *Journal of Multivariate Analysis*, vol. 98, no. 5, pp. 873–895, 2007.
- [40] A. Ben-Hur, A. Elisseeff, and I. Guyon, “A stability based method for discovering structure in clustered data,” in *Biocomputing 2002*. World Scientific, 2001, pp. 6–17.
- [41] S. Wagner and D. Wagner, *Comparing clusterings: an overview*. Universität Karlsruhe, Fakultät für Informatik Karlsruhe, 2007.
- [42] N. X. Vinh, J. Epps, and J. Bailey, “Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance,” *The Journal of Machine Learning Research*, vol. 11, pp. 2837–2854, 2010.
- [43] E. M. Ampuero, M. Cerda, S. Hartel, F. J. Rubio, S. Massa, P. Cubillos, L. Abarzua, R. Sandoval, and U. Wyneken, “Chronic fluoxetine treatment induces maturation-compatible changes in the dendritic arbor and in synaptic responses in the auditory cortex.” *Frontiers in pharmacology*, vol. 10, p. 804, 2019.
- [44] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [45] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid *et al.*, “Fiji: an open-source platform for biological-image analysis,” *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.

Apéndice A. Material Suplementario

A.1. Recuperación de Información mediante Superpixels

La captura o recuperación de la información del vecindario de los píxeles del esqueleto se realiza utilizando la misma imagen a partir de la cual se obtiene el esqueleto, en conjunto con el grafo derivado del esqueleto. El procedimiento consta de los siguientes pasos:

1. Se generan clusters, también llamados *super pixels*, mediante una agrupación en base a un kernel de tamaño 3x3. El kernel busca separar los pixels que no aportan información, como los que se encuentran en segundo plano o *background*, con respecto a los que son parte de la célula observada y se encuentran en el primer plano o (*foreground*). Los píxeles en el primer plano son los que concentran el interés para el análisis. La creación de clusters permite establecer la primera información de vecindarios.
2. Cada cluster es representado por un nodo, ubicado en el centro de masa del *super pixel*. En conjunto con el centro de masa es posible obtener información geométrica mediante los *raw image moments* [24] del cluster. En este punto también se procede a realizar fusiones de clusters en base a 2 criterios: aquellos clusters con un número de píxeles inferior a un umbral definido por el parámetro *Max_Thickness*, y los clusters que sólo tengan 2 vecinos y se encuentren bajo un umbral de cercanía definido por el parámetro *Connectivity_Threshold*. El parámetro *Max_Thickness* hace referencia al ancho promedio de un filamento en la imagen y es definido por el usuario, mientras que *Connectivity_Threshold* es el resultado del múltiplo entre *Max_Thickness* y un factor que depende de la célula observada, variando entre 0.4 y 0.5. Una de las ventajas de fusionar clusters bajo los criterios mencionados radica en identificar y eliminar ciclos triangulares entre vecinos.
3. Para asociar los super píxeles con los nodos del grafo se realiza un procedimiento similar al del paso 2, donde los nodos que representan clusters son fusionados con los nodo del grafo. Se utiliza la posición del nodo en el grafo como centro de una matrix de tamaño 3x3, la que es recorrida para obtener los clusters a los que pertenece cada píxel en la matriz. Luego, los clusters encontrados son absorbidos por el nodo del grafo, el que representa un nuevo *super pixel* que reemplaza a los que absorbe, heredando los vecinos que estos tenían y manteniendo la posición del nodo del grafo.

A.2. Individualización manual de filamentos

La individualización manual de filamentos realizada por un experto se puede realizar mediante el programa *Fiji* (Schindelin et al. [45]). El procedimiento implica construir una o más regiones de interés sobre una imagen con uno o más filamentos, mediante la herramienta de selección a mano alzada. El único requisito de cada región de interés, también llamada *ROI* por su sigla en inglés, es conformar un área cerrada. Una vez se obtiene una ROI cerrada, se puede almacenar su ubicación en el *ROI Manager*, disponible en el menú *Analyze*, submenú *Tools*, en *Fiji*. Un ejemplo de lo anterior se puede observar en la Figura 1.

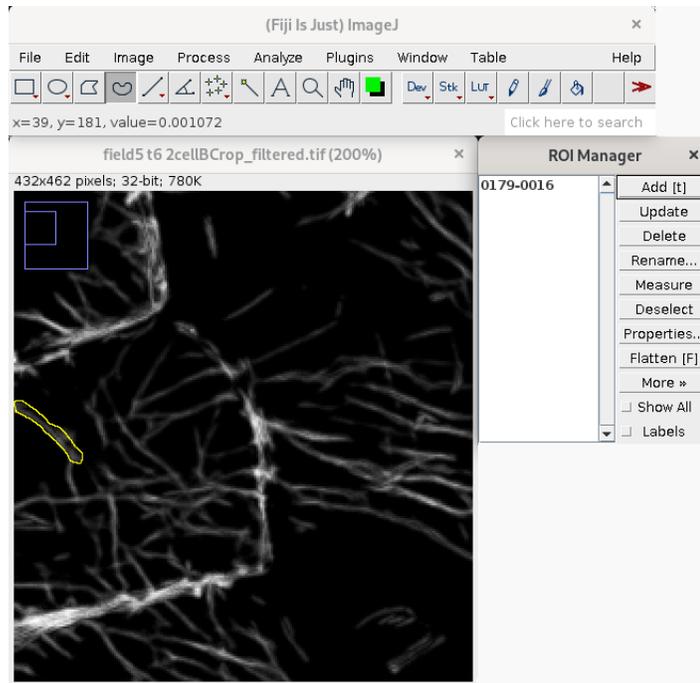


Figura 1: Ejemplo del procedimiento para individualizar manualmente un filamento realizado por un experto en Fiji. Fuente: Elaboración Propia.

Una vez generadas todas las ROIs que indican los filamentos individualizados por un experto, se genera una segmentación binaria de la imagen de filamentos. Así, los píxeles fuera de las ROIs se clasifican como parte del *background* o de segundo plano, con lo que no son consideradas en análisis futuros. Esta imagen resultante es la que puede ser utilizada para extraer un grafo mediante el procedimiento descrito en la sección 3.3.1.

A.3. Resultados algoritmo propuesto en Imágenes Sintéticas

La Tabla 1 indica el detalle de las 5 iteraciones utilizadas para promediar el resultado del algoritmo propuesto para las Figura 5.1 con la configuración Modo 2. Por su parte, la Tabla 2 muestra los resultados de la individualización de filamentos del algoritmo propuesto para la Figura 5.2, con la configuración definida como Modo 3. Información detallada de otros resultados puede encontrarse en el repositorio git del algoritmo propuesto.

Semilla	VI	TP	FP	TN	FN	R	J	P	R	F1	C/P	C/GT
0	0.81	9	5	67	10	0.83	0.37	0.64	0.47	0.54	4/6	4/6
10	0.81	9	5	67	10	0.83	0.37	0.64	0.47	0.54	4/6	4/6
1271	0.27	7	1	37	0	0.97	0.87	0.87	1	0.93	4/5	4/6
3389	0.81	9	5	67	10	0.83	0.37	0.64	0.47	0.54	4/6	4/6
6821	0.81	9	5	67	10	0.83	0.37	0.64	0.47	0.54	4/6	4/6

Tabla 1: Resultados de 5 iteraciones de algoritmo propuesto para individualizar filamentos en la figura 5.1, con la configuración Modo 2. El valor máximo de VI en este caso es de 2.3978, ya que el tamaño del *data set* es de 11 aristas. El número de filamentos en el *ground truth* es 6. Las columnas TP, FN, TN y FN corresponden a *True Positive*, *False Positive*, *True Negative* y *False Negative* respectivamente. Por su parte, R y J son los índices Rand y Jaccard, P y R son *Precision* y *Recall*, C/P indica la cantidad de filamentos correctos respecto a los filamentos propuestos, mientras que C/GT son los filamentos correctos en relación a los filamentos individualizados por un experto.

Semilla	VI	TP	FP	TN	FN	R	J	P	R	F1	C/P	C/GT
0	2.14	28	53	263	62	0.71	0.19	0.34	0.31	0.32	3/9	3/6
10	2.09	20	37	261	60	0.74	0.17	0.35	0.25	0.29	3/9	3/6
1271	2.84	29	63	373	96	0.71	0.15	0.31	0.232	0.26	3/10	3/6
3389	2.14	28	53	263	62	0.71	0.19	0.34	0.31	0.32	3/9	3/6
6821	1.91	19	35	223	48	0.74	0.18	0.35	0.28	0.31	3/9	3/6

Tabla 2: Resultados de 5 iteraciones de algoritmo propuesto para individualizar filamentos en la figura 5.2, con la configuración Modo 3. El valor máximo de VI en este caso es de 2.8332, ya que el tamaño del *data set* es de 17 aristas. El número de filamentos en el *ground truth* es 5. Las columnas TP, FN, TN y FN corresponden a *True Positive*, *False Positive*, *True Negative* y *False Negative* respectivamente. Por su parte, R y J son los índices Rand y Jaccard, P y R son *Precision* y *Recall*, C/P indica la cantidad de filamentos correctos respecto a los filamentos propuestos, mientras que C/GT son los filamentos correctos en relación a los filamentos individualizados por un experto.