

Simulación de movimientos vermiformes mediante caminatas aleatorias entre n-símplex vecinos

Simulation of vermiform movements through
random walks between n-simplex neighbors

Jorge Arroyo-Hernández¹, José Pablo Alvarado-Moya²

Arroyo-Hernández, J; Alvarado-Moya, J. Simulación de movimientos vermiformes mediante caminatas aleatorias entre n-símplex vecinos. *Tecnología en Marcha*. Vol. 33-2. Abril-Junio 2020. Pág 54-66.

 <https://doi.org/10.18845/tm.v33i2.4329>

Fechas de recepción: 1 de mayo de 2019
Fecha de aprobación: 25 de agosto de 2019

1 Escuela de Matemática, Universidad Nacional. Costa Rica. Correo electrónico: jarroy@una.ac.cr.

 <http://orcid.org/0000-0002-3890-5011>

2 Escuela de Ingeniería Electrónica, Instituto Tecnológico de Costa Rica. Costa Rica. Correo electrónico: palvarado@tec.ac.cr.



Palabras clave

Movimiento vermiforme; n-símplex; caminatas aleatorias.

Resumen

La aplicación de técnicas de visión computacional en contextos biológicos requiere generalmente la anotación manual de datos del contexto específico, lo que hace que estas etapas sean costosas en inversión de tiempo y recurso humano especializado. En este artículo, se presenta un método novedoso de generación de datos aleatorios dentro de un subdominio restringido a variedades de formas permitidas. Se particulariza el proceso a modelos de forma explícitos basados en hitos de frontera para estructuras vermiformes. Mediante este proceso, se logran simular deformaciones secuenciales de movimientos de gusanos por medio de caminatas aleatorias entre los subdominios.

Keywords

Vermiform movement; n-simplex; random walks.

Abstract

Application of computational vision techniques in biological contexts usually requires manual annotation of data from the specific context, which makes these stages costly due to investment in specialized human resource time. In this article we present a novel method of generating random data within a subdomain restricted to a variety of allowed shapes. The process is particularized to explicit models based on border landmarks for vermiform structures. Through this process, sequenced deformations of worm movements are simulated by random walks between the subdomains.

Introducción

La economía en Centroamérica depende de la agricultura. En países como Costa Rica, en donde indicadores macroeconómicos muestran un incremento en la diversificación de bienes y servicios, un 10,3% de las actividades económicas se atribuyen al sector agropecuario [1]. La supervivencia de este sector depende de la mejora continua que permita alcanzar, mantener e innovar los procesos, superando los estándares de calidad y cantidad demandados por los mercados locales e internacionales [2].

En la búsqueda de biocontroles para enfermedades en plantas se ha perseguido automatizar parcial o totalmente el análisis de muestras con modernos métodos de campo de visión por computador y aprendizaje de máquina, para así potenciar el impacto del tiempo invertido por expertos en el análisis de dichas muestras [3], [4].

Métodos que utilizan explícitamente información de la forma de organismos para determinar indicadores de interés requieren suficiente material de entrenamiento, que es en general costoso de preparar manualmente. Este artículo propone una técnica novedosa de emulación de la variación de siluetas, que se produce con un bajo costo computacional. El método propuesto parte de un conjunto de datos inicial que define el dominio válido de formas, y a partir de allí sigue un recorrido aleatorio por dicho dominio. El proceso genera una sucesión de formas sintéticas nuevas que complementan el conjunto de datos original, con el objetivo de poder utilizarlo en procesos de entrenamiento y validación de sistemas de aprendizaje automático.

Si bien es cierto, el concepto presentado en este artículo se desarrolla en el contexto de análisis de imágenes de nematodos como caso particular, el concepto es aplicable a cualquier otra área en que se utilicen modelos explícitos de forma, como el análisis facial [5] o el análisis de imágenes médicas [6], entre otros.

El artículo se estructura de la siguiente manera: una revisión de trabajos relacionados se presenta en la siguiente sección.; el fundamento teórico y algorítmico utilizado en la generación de las nuevas formas se describe en la tercera sección; en la cuarta, se analizan los resultados del método propuesto. Conclusiones y trabajo futuro presentados en la última sección cierran el artículo.

Trabajos relacionados y modelo propuesto

Varias estrategias se han propuesto para clasificar, detectar, simular y extraer estructuras en imágenes de objetos con formas de gusano. En “Active shape models...” [7] se sientan las bases de los métodos capaces de expresar explícitamente formas válidas por medio de modelos estadísticos de distribución de puntos, con base en el análisis de componentes principales, para la reducción de la dimensión de vectores representativos de la formas. Sin embargo, en ese mismo trabajo se demuestra que una distribución normal de puntos es incapaz de restringir las formas representadas a formas vermes válidas y deja por fuera siluetas anatómicamente plausibles en nematodos y otros organismos vermiformes.

Un método de descripción de las contorsiones de la cabeza del *C. elegans* se expone en “Automated detection...” [8]. Dicho método propone cómo medir los ángulos de contorsión en la cabeza, en la descripción del *forrageo* (movimientos exploratorios). La información espaciotemporal se aprovecha en “Tracking the swimming motions...” [9] para predecir la posición y ajustar la forma del organismo en procesos de segmentación y rastreo, utilizando para ello la evolución de los ángulos y vértices del esqueleto morfológico. Estos conceptos se toman como base en este trabajo para ser utilizados en la dirección contraria, y sintetizar con ello nuevas formas.

Otros métodos para representar el espacio de formas válidas de nematodos basados en técnicas de aprendizaje de diccionarios y de aprendizaje de variedades se exploran en “Active dictionary models...” [10] y “Segnema: nematode segmentation...” [11]. Un concepto basado en un modelo generativo que aprende la geometría de la variedad a partir de los datos se propone en “Geometry based data generation” [12], en donde se procura ignorar la densidad de estos, utilizando mapas de difusión para establecer la estructura geométrica de la variedad. Dicha estrategia no es apta para el tipo de datos presentes en este artículo, donde la dimensión del espacio es elevada con respecto al número de datos disponible, pues la difusión no cuenta con suficiente soporte para ser estimada de forma robusta.

Como alternativa para dar más peso a la geometría de la variedad que a la densidad de puntos disponibles de entrenamiento, se utiliza en la presente propuesta una técnica similar a los k vecinos más cercanos móviles, expuesta en “Analysis and evaluation of V*-kNN...” [13] para un contexto de consultas en bases de datos espaciales. El método es similar al presentado aquí en el hecho de que usa tanto la información de la posición actual en la caminata como del vecindario para la decisión del siguiente paso; pero en el presente caso la estrategia elegida para saltar a la siguiente posición está orientada a producir formas disímiles al punto de partida, manteniendo su validez, en vez de procurar la descripción geométrica de la variedad.

En otra dirección, la simulación completa del organismo de un nematodo de la especie *C. elegans* se persigue en el proyecto OpenWorm [14]. En ese proyecto ya se ha generado el movimiento locomotivo de un espécimen, por medio de la simulación de 302 neuronas

y 95 células musculares. Los datos producidos en dicha simulación se pueden acoplar a renderizadores que generen las siluetas o las imágenes, según se requiera. Si bien es cierto, esta simulación de abajo hacia arriba (*bottom-up*) permite alcanzar altos niveles de realismo anatómico en los detalles, el consumo de recursos computacionales no la hace viable en contextos donde la generación de formas es parte de otros procesos, como la segmentación de nematodos o su rastreo en imágenes.

En este trabajo se propone una estrategia para la generación de una sucesión de formas a partir de una caminata aleatoria estimada sobre un conjunto de datos de entrenamiento que se asume pertenecen a una variedad topológica embebida en un espacio de alto número de dimensiones, en donde el reto es generar formas suficientemente disímiles de las existentes en el cuerpo de entrenamiento, pero que sigan siendo válidas en el dominio de aplicación, todo ello con un bajo consumo de recursos computacionales, que permita insertar el proceso generativo en otras tareas, como la segmentación o rastreo de nematodos en imágenes.

Simulación de movimientos vermiformes

La generación de nuevas formas aleatorias con el método propuesto persigue restringir los resultados a estructuras vermiformes válidas. Se parte de un modelo donde el dominio completo de D_f formas válidas se particiona en subdominios S_{D_f} , para los que se define una relación de vecindad. Cada paso realizado en una caminata aleatoria discreta produce saltos a subdominios vecinos. Esto asegura que cada forma generada en cada paso de la caminata es única en su estructura, diferenciándose de los pasos previos, pero además que es una forma válida. En las siguientes secciones se detalla el proceso de generación de formas y la definición del camino aleatorio.

Conjunto de datos para la simulación

Los modelos de forma son prototipos paramétricos representados a través de un conjunto D_f de formas cuyo i -ésimo elemento se representa vectorialmente por $\hat{f}_i = [x_{i0} \ y_{i0} \ x_{i1} \ y_{i1} \ \dots \ x_{im-1} \ y_{im-1}]^T$ donde cada par (x_{ij}, y_{ij}) representa las coordenadas cartesianas de un hito P_{ij} en la imagen. Los hitos deben estar ubicados secuencialmente a lo largo del contorno del objeto que representan (ver figura 1).

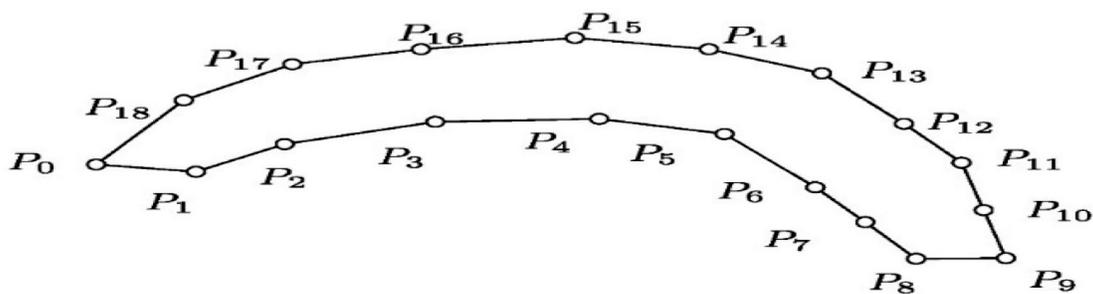


Figura 1. Secuenciación de hitos de una forma verme f_i . Cada forma en el conjunto de datos se construye a partir de una forma *cruda* representada con un número arbitrario de hitos colocados manualmente por un anotador humano sobre el borde de cada organismo en imágenes digitales. Las imágenes provienen de muestras reales de suelo, capturadas con un microscopio óptico en trece secuencias con características visuales variadas.

Las formas crudas son normalizadas en varios pasos. Primero se asegura que los hitos anotados sigan una secuencia en sentido antihorario [15], [16], iniciando siempre en la cabeza o la cola. Un número constante m de hitos equidistantes para cada forma del conjunto D_f se genera por medio de interpolación con trazadores segmentarios cúbicos [17], [18] a partir de los hitos anotados. Además, el conjunto de datos se extiende (o aumenta) con formas generadas por reflexiones de las formas interpoladas, e intercambiando las posiciones de cabeza y cola, con

$$\tilde{f}_i = \left[\begin{array}{cccccccccccccccc} x_{\frac{i}{2}^m} & y_{\frac{i}{2}^m} & x_{\frac{i}{2}^m+1} & y_{\frac{i}{2}^m+1} & \text{L} & x_{i_{m-1}} & y_{i_{m-1}} & x_{i_0} & y_{i_0} & x_{i_1} & y_{i_1} & \text{L} & x_{\frac{i}{2}^m-1} & y_{\frac{i}{2}^m-1} \end{array} \right]^T$$

Finalmente, todo el conjunto de formas se alinea y re-escala utilizando el análisis de Procrustes descrito en “Active shape models...” [7]. En adelante, la i -ésima forma normalizada del conjunto total de datos se denota con \hat{f}_i .

Dominio y subdominios de formas vermes

El dominio D_f es un conjunto de cardinalidad infinita en el que cada elemento $\hat{f}_i \in D_f$ es una forma verme válida. Asimismo, se define un subdominio S_{D_f} como un subconjunto de individuos \hat{f}_i , que por medio de algún estadístico de proximidad o distancia, agrupa formas similares. De este modo el dominio se particiona en un conjunto de subdominios disjuntos. En el presente trabajo, cada subdominio se representa por medio de un n -simplex [19], es decir, una envoltura convexa de n dimensiones engendrada por $n + 1$ puntos en \mathbb{R}^{2m} , dada por $\{v_0, v_1, \text{L}, v_n\}$ con vectores $\{v_1 - v_0, v_2 - v_0, \text{L}, v_n - v_0\}$ linealmente independientes, de modo que $\{f_0, f_1, \text{L}, f_n\}$ sean formas vermes válidas similares. Se parte de la premisa de que formas válidas \hat{f}_i dentro del subdominio (n -simplex) se engendran como combinaciones baricéntricas:

$$\hat{f}_i \approx \alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2 + \text{L} \alpha_n f_n$$

Con $0 \leq \alpha_i \leq 1$ y $\sum_{j=0}^n \alpha_j = 1$, y que dicho subdominio embebido en el espacio \mathbb{R}^{2m} es (aproximadamente) tangente a una variedad topológica n -dimensional correspondiente al dominio D_f que contiene todas las formas válidas. Dos subdominios que poseen al menos un vértice del n -simplex en común se consideran vecinos.

Por medio del algoritmo de k - medias [20] se encuentran k conglomerados \mathbb{R}^{2m} , que agrupan todas las formas disponibles de acuerdo a su similitud. Los k centroides asociados a los conglomerados conforman el conjunto C de posibles vértices de los n -simplex (ver figura 2).

Generación de movimientos aleatorios vermes

Los movimientos vermes se simulan con una caminata aleatoria [21], definida en este caso como una sucesión de saltos aleatorios entre subdominios $S_{D_f}^{(i)}$ vecinos (como se muestra en la figura 3), que deforman gradualmente la silueta original. Este recorrido entre subdominios vecinos permite que cada elemento \hat{f}_i conserve información de la forma de al menos uno de los vértices que los subdominios tienen en común, obligando así en cada cambio de subdominio a una deformación paulatina, sin cambios abruptos.

Se parte entonces de una forma verme inicial \hat{f}_0 . A partir de ella se calcula el n -simplex $S_{D_f}^{(0)}$ engendrado por el conjunto V_0 de los n centroides en C más cercanos a \hat{f}_0 . Esta tarea utiliza

el método aproximado de búsqueda de vecinos más cercanos expuesto en “Fast approximate nearest neighbors...” [22].

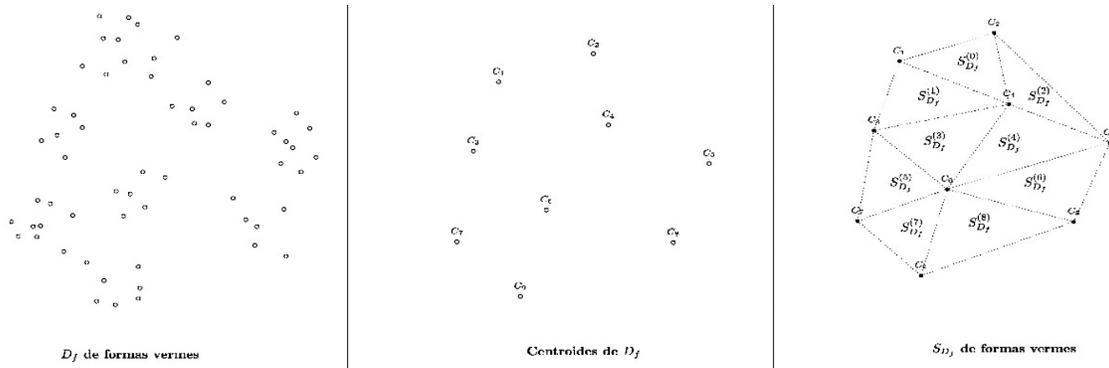


Figura 2. Obtención de S_{D_f} a través de k -medias. El subconjunto de n vértices se elige de modo que el n -simplex engendrado no contenga ningún otro vértice. La elección del número $k > n$ de conglomerados define entonces indirectamente el volumen de los simplex, y qué tan disímiles serán las formas en subdominios vecinos.

Ahora, se selecciona el siguiente subdominio $S_{D_f}^{(1)}$ determinado por el conjunto V_1 de los n centroides más próximos a uno de los vértices (seleccionado aleatoriamente) de V_0 . Este nuevo subdominio de formas permitidas corresponde a un n -simplex vecino del anterior. Dentro de este subdominio $S_{D_f}^{(1)}$ se toma aleatoriamente la siguiente forma $\hat{f}^{(1)}$ de la sucesión, utilizando coeficientes baricéntricos generados aleatoriamente.

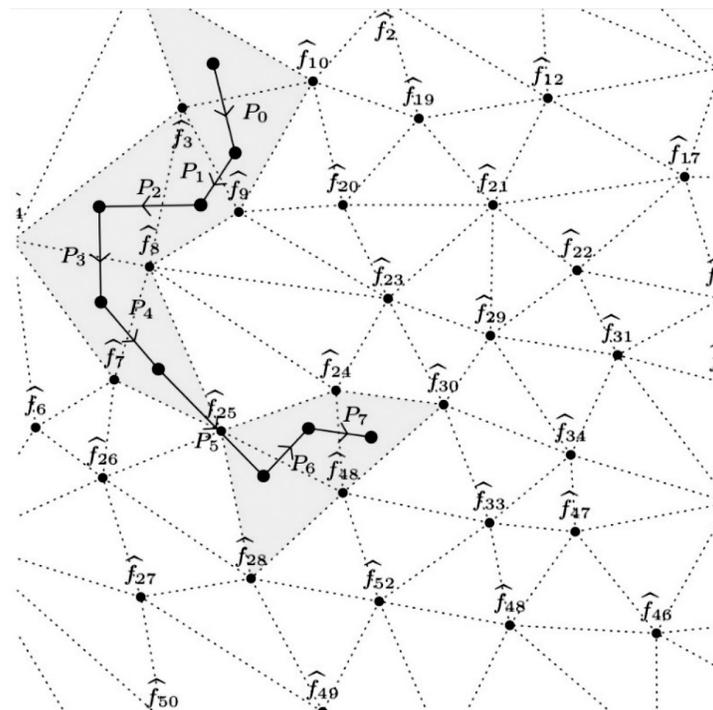


Figura 3. Caminata aleatoria de \hat{f}_i entre S_{D_f} vecinos. El proceso se repite iterativamente hasta satisfacer un criterio de parada, como por ejemplo, un número predeterminado de pasos, o haber superado alguna medida de distancia con el punto de partida. El algoritmo completo se resume en la figura 4.

```

Input :  $\hat{f}_0 \leftarrow$  Forma inicial
           $D_f \leftarrow$  Dominio de formas permitidas
1  $C \leftarrow kmeans(D_f)$  /* centroides de  $D_f$  calculados por  $k$ -means */
2  $M_0 \leftarrow \hat{f}_0$  /*  $M_0$  vértice inicial */
3  $i \leftarrow 1$ 
4 mientras  $i$  menor a algún  $p$  hacer
5    $V_i \leftarrow knn(C, M_{i-1})$  /*  $V_i$  vecinos más cercanos de  $M_{i-1}$  por  $k$ -nn */
6    $S_{D_f}^{(i)} \leftarrow ksíplex(V_i)$  /*  $i$ -ésimo subdominio de formas permitidas */
7    $\hat{f}_i \leftarrow random(S_{D_f}^{(i)})$  /* Cálculo de la  $i$ -ésima deformación  $\hat{f}_i \in S_{D_f}^{(i)}$  */
8    $M_i \leftarrow random(V_i)$  /*  $M_i$  es un vértice al azar de  $S_{D_f}^{(i)}$  */
9    $i \leftarrow i + 1$ 
10 fin

```

Figura 4. Algoritmo de caminata aleatoria de formas válidas. El algoritmo propuesto permite simular las deformaciones a partir de una instancia inicial y que continúen de manera natural conservando información de forma entre las iteraciones, y dando a la secuencia el carácter de movimiento aleatorio de un organismo vermiforme. En la figura 5, se aprecia un ejemplo del movimiento generado por el recorrido aleatorio basado en el salto entre subdominios vecinos. Se parte de una forma inicial arbitraria y se muestran para nueve iteraciones las deformaciones generadas, que el método asegura sean formas vermes válidas.

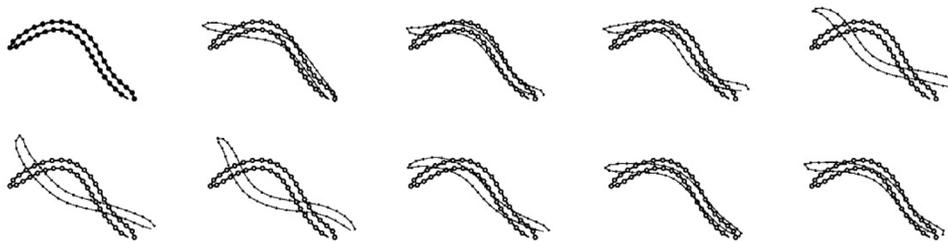


Figura 5. Deformaciones de movimientos continuos mediante el traslado aleatorio entre S_{D_f}

Transformación de formas y proyección a la variedad

Para poder evaluar el alcance de las formas generadas en la caminata aleatoria, es necesario resolver dos problemas:

- ¿Cómo generar eficientemente formas arbitrarias que se puedan considerar válidas?
- ¿Cómo proyectar una forma verme cualquiera al dominio de formas válidas?

El primer problema se ataca para el caso de los nematodos, con la simulación de deformación de la anatomía. El segundo problema se resuelve con el algoritmo de la estimación de las coordenadas baricéntricas correspondientes a una forma dada dentro del n -símplex más cercano. Los detalles se desglosan en las siguientes dos secciones.

Método de deformación de nematodos

Para deformar un nematodo \hat{f}_i , se utiliza el esqueleto E_i (ver figura 6) definido como un vector compuesto por las posiciones de las vértebras, cada una dada por el promedio de los dos hitos

correspondientes opuestos P_{ij} y P_{im-j} . Se toman los hitos sobre la cabeza y cola como la primera vértebra $V_{i0} = (E_{i0_x}, E_{i0_y}) = (x_{i0}, y_{i0})$ y última vértebra $V_{iu} = (E_{iu_x}, E_{iu_y}) = (x_{iu}, y_{iu})$, respectivamente, donde $u = \frac{m}{2}$ y m es el número de hitos de \hat{f}_i . El conjunto de todos los esqueletos E_i de $\hat{f}_i \in D_f$ se denota \bar{E} , y cada uno se expresa vectorialmente por

$$E_i = \left[E_{i0_x} \ E_{i0_y} \ E_{i1_x} \ E_{i1_y} \ L \ E_{iu_x} \ E_{iu_y} \right]^T$$

La posición de cada vértebra intermedia está dada por

$$E_{ij_x} = \frac{x_{ij} + x_{i(m-j)}}{2} \quad E_{ij_y} = \frac{y_{ij} + y_{i(m-j)}}{2} \quad \text{para } 0 < j < u$$

La modificación de forma se concentra en dos fragmentos del esqueleto, definidos por las u' vértebras $V_{ij} = (E_{ij_x}, E_{ij_y})$ más cercanas a la cabeza y a la cola de \hat{f}_i . Para las vértebras V_j del fragmento de la cabeza se cumple $1 \leq j < u'$, y para el fragmento de cola $u - u' \leq j < u$.

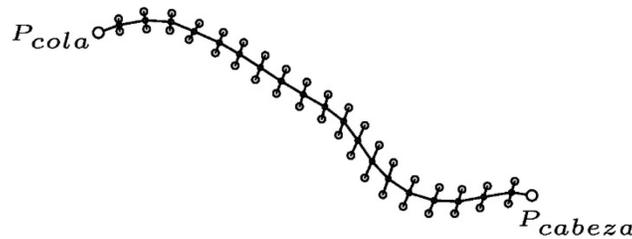


Figura 6. Esqueleto de una forma verme \hat{f}_i

Para lograr la deformación deseada (e inspirada en el *forrageo* [8]) se toman los fragmentos u' con vértebras de cola y cabeza, y se manipulan (junto a los hitos correspondientes a cada vértebra) como una cadena cinemática directa. En dicha cadena, cada vértebra cuenta con un grado de libertad rotacional. Los ángulos de rotación θ están restringidos a ser iguales en todas las vértebras.

Dado el número u' de vértebras que modificar, el algoritmo toma inicialmente la vértebra del fragmento de cabeza $V_{iu'} = (t_x, t_y)$ como fija y aplica la transformación al resto de vértebras e hitos en la cadena cinemática.

$$M_v \begin{bmatrix} x_{ij} \\ y_{ij} \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & \text{sen} \theta & -(t_x \cos \theta + t_y \text{sen} \theta) + t_x \\ -\text{sen} \theta & \cos \theta & -(t_y \cos \theta - t_x \text{sen} \theta) + t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{ij} \\ y_{ij} \\ 1 \end{bmatrix}$$

Luego, dada la nueva posición de la vértebra $V_{iu'-1}$ se le aplica M_v al resto de elementos de la subcadena cinemática hacia la cabeza. Posteriormente, se repite el proceso a partir de la vértebra $V_{iu'-2}$ y así sucesivamente hasta llegar al hito de la cabeza. De forma análoga, se sigue el procedimiento de la vértebra colocada en la posición $u - u' + 1$ hacia la cola. Las contorsiones vermes de la cola resultante se ilustran en la figura 7.

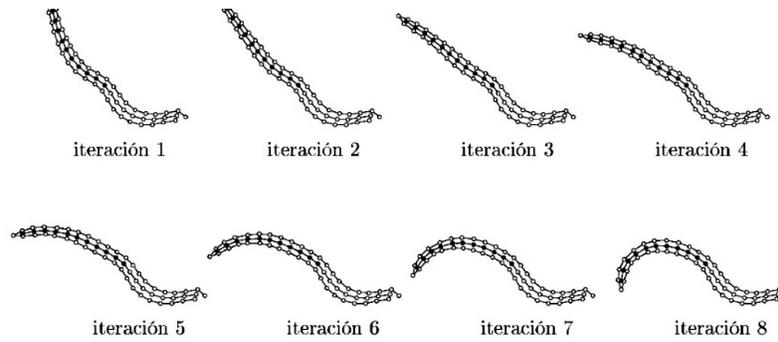


Figura 7. Iteraciones del método de serpiente para la cola de \hat{f}_i para $u' = 11$, $\theta = 0,2$ rad

Proyección al dominio de formas válidas

La proyección de una forma \hat{f}_α al dominio de formas válidas se divide en dos etapas. La primera es determinar el subdominio al que pertenece \hat{f}_α , y la segunda es calcular los pesos baricéntricos que mejor aproximan la forma en dicho subdominio.

Para la primera etapa se calcula el conjunto de los $n+1$ vecinos más cercanos $\{\hat{f}_0, \hat{f}_1, \dots, \hat{f}_n\}$ a \hat{f}_α ordenados según su lejanía. Estos vecinos son elementos del conjunto de posibles vértices obtenido con el algoritmo de k -medias (sección 3.2). Luego, se aplica la variación del método *matching pursuit* [25] propuesta en “Active dictionary models...” [10] para estimar los pesos baricéntricos $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_n]$ que mejor aproximan la forma \hat{f}_α dada por $\hat{f}_\alpha \approx \alpha_0 f_0 + \alpha_1 f_1 + \alpha_2 f_2 + \dots + \alpha_n f_n = \hat{f}_c$

Así \hat{f}_c es la aproximación más cercana a \hat{f}_α que se encuentra dentro del n -símplex con vértices f_i .

Resultados y discusión

En esta sección, se muestran resultados sobre la versatilidad del método propuesto para la generación de sucesiones aleatorias de formas vermes que simulen el movimiento de serpiente. Primero, se analiza la similitud entre formas de la secuencia en función del número de centroides utilizados. Luego se revisa el método utilizado para estimar los pesos baricéntricos.

Para la experimentación se utilizó una base de datos de entrenamiento con 10 932 formas vermes distintas, generadas a partir de un conjunto inicial de formas crudas, normalizadas y aumentadas según lo descrito en la sección 3.1, con $m=40$ hitos por contorno.

Los algoritmos se implementaron en el lenguaje C++, utilizando la biblioteca FLANN [23] para el cálculo eficiente de los centroides y de búsqueda de vecinos cercanos, así como la biblioteca LTI-Lib2 [24] como base para el procesamiento de imágenes.

Distancia entre formas en función de la dimensión del subdominio

En el primer experimento se evaluó la distancia entre las formas producidas con los saltos entre subdominios, en función de la dimensión de dichos subdominios. Dicha distancia entre formas se cuantificó con el promedio de la distancia euclídea entre hitos correspondientes de dos

formas sucesivas \hat{f}_i y \hat{f}_{i+1} , dada por $d(\hat{f}_{i+1}, \hat{f}_i) = \frac{1}{m} \sum_{j=0}^{m-1} \sqrt{(x_{i+1j} - x_{ij})^2 + (y_{i+1j} - y_{ij})^2}$

El cuadro 1 resume los resultados obtenidos usando 1000 iteraciones del algoritmo de generación de formas (figura 4) con subdominios n -dimensionales (n -simplex), con $n = 5, 10, 20, \dots, 100$ vértices tomados de k centroides, definidos sobre el conjunto de datos con el método de k medias (sección 3.2) para $k > 200$ en múltiplos de 100.

El resultado confirma la correlación existente entre la distancia entre las formas sucesivas y la dimensión n de los subdominios. Entre mayor la cantidad de vértices utilizada, mayor la distancia entre los hitos correspondientes a dos deformaciones sucesivas.

Esto puede explicarse así: manteniendo el número k de puntos disponibles como vértices constante, entonces mientras mayor sea el número n , mayor será la distancia entre dichos vértices, de modo que también aumente el tamaño del paso realizado de una forma a otra.

Cuadro 1. Promedio de distancias entre hitos correspondientes a \hat{f}_i y \hat{f}_{i+1} en una sucesión de 1000 formas

$C_j \setminus$ Nodos	5	10	20	30	40	50	60	70	80	90	100
200	9.234	10.912	12.290	14.024	13.299	14.149	15.750	14.748	16.111	17.904	16.833
300	9.772	11.263	12.270	12.142	13.477	15.999	16.252	16.175	17.001	17.576	18.475
400	8.734	10.261	11.417	12.571	13.813	13.415	15.108	15.911	16.614	17.124	17.395
500	8.870	10.999	11.530	11.995	12.223	13.122	14.145	14.925	16.685	18.693	18.583
600	7.647	9.9830	11.314	12.416	13.591	14.164	14.560	15.969	16.823	19.612	20.405
700	7.683	9.5160	11.193	12.497	12.639	13.226	12.569	14.097	15.279	16.482	19.400
800	6.232	9.0780	10.791	12.318	12.559	13.558	13.182	13.921	14.517	14.568	17.801
900	6.283	9.3140	11.217	12.206	13.095	14.642	14.992	15.347	15.347	16.359	19.029
1000	6.483	9.2060	10.794	12.207	13.324	14.218	15.074	15.902	16.522	16.791	18.385

Nota: Se tomaron los k centroides correspondientes a los conglomerados encontrados por el algoritmo de k -medias, y de ellos se seleccionaron subdominios de n dimensiones (n -simplex) embebidos en un espacio de \mathbb{R}^{80} (40 hitos por forma).

Si el número de vértices del subdominio es 5, el desplazamiento de los hitos será menor a 8 píxeles en promedio, lo que exhibe baja variabilidad promedio de cambios independientemente del número de centroides.

Si se elige un número de vértices de los subdominios mayor a 40, las deformaciones sucesivas con distancias superiores a 13 píxeles por hito se perciben abruptas y se pierde la noción de continuidad de la simulación, independientemente del número k de centroides utilizado.

Simulaciones aceptables con distancias promedio entre hitos de alrededor de 12 píxeles se pueden generar con n -simplex con dimensión n entre 10 y 30, independientemente del número de centroides.

Validez y alcance del dominio de formas

El segundo experimento persigue cuantificar de qué modo cualquier forma verme \hat{f}_α puede ser aproximada en el dominio de formas válidas D_f .

Se utilizó una muestra de 1000 formas vermes \hat{f}_α generadas aleatoriamente usando el método de deformación de nematodos ya descrito, normalizadas según el método indicado para la normalización de las formas crudas, utilizando como punto de partida formas no presentes en la base de datos de entrenamiento.

Se utilizó el método para la proyección al dominio de formas válidas, para calcular la aproximación \hat{f}_c de cada forma \hat{f}_α . Se evaluó entonces la distancia promedio entre hitos (dada en la sección 4.1) en función de la dimensión n de los n -simplex, para varios valores de k centroides de donde se tomaron los $n + 1$ vértices de los n -simplex. En los resultados que se muestran en la figura 8 hay tendencia a una disminución de la distancia entre la aproximación \hat{f}_c de \hat{f}_α cuando aumenta el número k de centroides que determinan los subdominios. Desde una partición con $k = 200$ conglomerados del D_f para formar los subdominios, se muestra que aproximadamente la distancia por hito entre las siluetas originales y las calculadas es menor a 1 pixel promedio, por lo que se concluye que son muy cercanas en forma. Por otro lado, al aumentar la dimensión de los n -simplex, la distancia de desvío entre hitos tiende a aumentar también para un mismo número de centroides k , puesto que si se mantiene k , entonces la distancia entre los vértices crece con el aumento de la dimensión n de la variedad. Obsérvese que este efecto se reduce conforme aumenta el número de centroides.

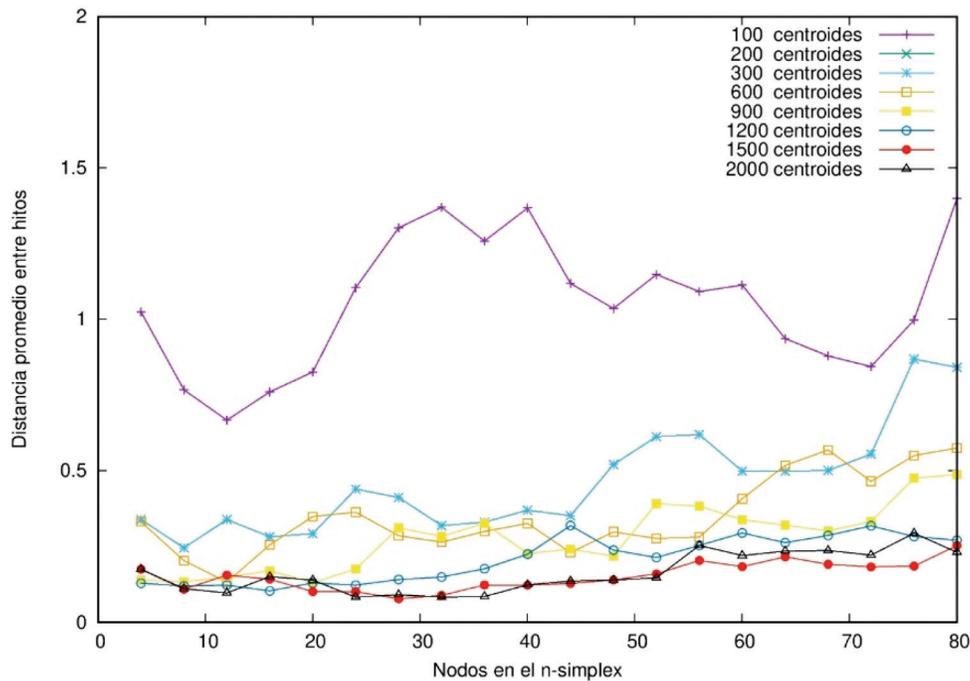


Figura 8. Distancia promedio por hito entre \hat{f}_c y \hat{f}_α en un S_{D_f} de k vértices

Conclusiones y trabajo futuro

Los experimentos realizados conducen a dos resultados: primero, el modelo basado en n -simplex para aproximar la variedad topológica que cubre el dominio D_f de formas válidas permite alcanzar cualquier forma con un desvío promedio de menos de un pixel por hito entre las formas válidas y sus proyecciones a dicha variedad. Segundo, la estrategia de caminata aleatoria propuesta, que utiliza saltos entre subdominios vecinos, permite generar secuencias de movimientos de serpiente realistas y (como consecuencia directa del primer punto) capaces de alcanzar en el proceso cualquier forma válida.

Las características de las formas generadas en la caminata aleatoria dependerán tanto de la dimensión n de la variedad, que determina el número de vértices de cada símplex, como del número k de centroides utilizados como candidatos a vértices de los símplex.

Estas características le permiten al método propuesto generar formas aleatorias que pueden ser utilizadas en procesos de segmentación y clasificación de siluetas vermes. De igual manera, el método de deformación propuesto es capaz de generar formas vermes distintas a las existentes en el cuerpo de entrenamiento.

La distancia entre formas consecutivas en la secuencia de pasos es ajustable con el método propuesto por medio de los parámetros n (dimensión de la variedad topológica) y k (total de conglomerados cuyos centroides constituyen los posibles vértices de los n -símplex). Dichos parámetros controlan además la complejidad total del proceso, para integrar el método en otras estrategias de segmentación y rastreo.

Referencias

- [1] L. F. Arauz Cavallini, A. I. Gómez De Miguel, M. Valverde Díaz, G. Rodríguez Muñoz y E. Mora Monge, "Informe de gestión del sector agropecuario y rural (mayo 2014 – abril 2018)", SEPSA, Costa Rica, inf. téc., abr. de 2018.
- [2] L. Pratt y L. Rivera, "Tendencias en el desarrollo de la agricultura en Centroamérica: Nuevos retos para el Sector Privado y el diseño de políticas públicas", Centro Latinoamericano de Competitividad y Desarrollo Sostenible, CLACDS – INCAE, inf. téc., sep. de 2003.
- [3] A. Suprem, N. Mahalik, and K. Kim, "A review on application of technology systems, standards and interfaces for agriculture and food sector," *Computer Standards & Interfaces*, vol. 35, no. 4, pp. 355-364, 2013.
- [4] A. Tatem, H. Lewis, P. Atkinson, and M. Nixon, "Increasing the spatial resolution of agricultural land cover maps using a hopfield neural network," *International Journal of Geographical Information Science*, vol. 17, no. 7, p. 647, 2003, issn: 13658816.
- [5] X. Cao, Y. Wei, F. Wen, J. Sun, "Face alignment by explicit shape regression," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 177-190, April 2014, issn: 0920-5691. doi: 10.1007/s11263-013-0667-3. Dirección: <http://link.springer.com/10.1007/s11263-013-0667-3>.
- [6] R. Sparks and A. Madabhushi, "Explicit shape descriptors: Novel morphologic features for histopathology classification," *Medical Image Analysis*, vol. 17, no. 8, pp. 997-1009, Dec. 2013, issn: 1361-8415. doi: 10.1016/J.MEDIA.2013.06.002. Dirección: <https://www.sciencedirect.com/science/article/abs/pii/S136184151300087X>.
- [7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models, their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.
- [8] K.-M. Huang, P. Cosman, and W. R. Schafer, "Automated detection and analysis of foraging behavior in *C. elegans*," in 2008 IEEE Southwest Symposium on Image Analysis and Interpretation, IEEE, 2008, pp. 29-32.
- [9] C. Restif and D. Metaxas, "Tracking the swimming motions of *C. elegans* worms with applications in aging studies," in International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2008, pp. 35-42.
- [10] M. Grüner Monzón, "Active dictionary models: a framework for non-linear shape modeling," master thesis, Electronics Engineering Department, Tecnológico de Costa Rica, Cartago, Costa Rica, May 2015.
- [11] J. Jiménez Chavarría, "Segnema: nematode segmentation strategy in digital microscopy images using deep learning and shape models," master thesis, Escuela de Ingeniería en Computación, Tecnológico de Costa Rica, Cartago, Costa Rica, Feb. 2019.
- [12] O. Lindenbaum, J. Stanley, G. Wolf, and S. Krishnaswamy, "Geometry based data generation," in *Advances in Neural Information Processing Systems*, 2018, págs. 1407-1418.
- [13] S. Nutanong, R. Zhang, E. Tanin, and L. Kulik, "Analysis and evaluation of V*-kNN: An efficient algorithm for moving kNN queries," *The VLDB Journal*, vol. 19, no. 3, pp. 307-332, June 2010.
- [14] B. Szigeti, P. Gleeson, M. Vella, S. Khayrulin, A. Palyanov, J. Hokanson, M. Currie, M. Cantarelli, G. Idili, and S. Larson, "OpenWorm: an open-science approach to modeling *Caenorhabditis elegans*," *Frontiers in Computational Neuroscience*, vol. 8, p. 137, Nov. 2014, issn: 1662-5188. doi: 10.3389/fncom.2014.00137. Address: <http://journal.frontiersin.org/article/10.3389/fncom.2014.00137/abstract>.
- [15] R. Sedgewick and K. Wayne, *Algorithms*. Addison-Wesley Professional, 2011.

- [16] C. Garza-Hume, M. C. Jorge, and A. Olvera, "Areas and shapes of planar irregular polygons," in *Forum Geometricorum*, vol. 18, 2018, pp. 17-36.
- [17] K. E. Atkinson, *An Introduction to Numerical Analysis*. John Wiley & Sons, 2008.
- [18] T. A. Foley and G. M. Nielson, "Knot selection for parametric spline interpolation," in *Mathematical Methods in Computer Aided Geometric Design*. Elsevier, 1989, 261-CP4.
- [19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [20] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl and colls., "Constrained k-means clustering with background knowledge," in *ICML*, vol. 1, 2001, pp. 577-584.
- [21] W. Woess, "Random walks on infinite graphs and groups, ép.," *Cambridge Tracts in Mathematics*, vol. 138. Cambridge University Press, 2000.[22] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP'09*, INSTICC Press, 2009, pp. 331-340.
- [23] M. Muja and D. G. Lowe, "Fast matching of binary features," in *2012 Ninth Conference on Computer and Robot Vision*, IEEE, 2012, pp. 404-410.
- [24] P. Dörfler and P. Alvarado-Moya, "Lti-lib – a c++ open source computer vision library," in *Advanced Man-Machine Interaction: Fundamentals and Implementation*, ép. *Signals and Communication Technology*, K. F. Kraiss, Ed., Berlin-Heidelberg: Springer Verlag, 2006, pp. 399-421.
- [25] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094-1121, 2012.