

VI JORNADAS INTERDISCIPLINARIAS DE
ANÁLISIS AVANZADO DE
IMÁGENES Y SEÑALES

*VI INTERDISCIPLINARY CONFERENCE
ON ADVANCED IMAGE AND SIGNAL
ANALYSIS*

7 y 8 de noviembre de 2022
Ciudad Autónoma de Buenos Aires, Argentina



Universidad Tecnológica Nacional

VI Jornadas Interdisciplinarias de Análisis Avanzado de Imágenes y Señales ; Coordinación general de Andrea Alejandra Rey. - 1a ed - Ciudad Autónoma de Buenos Aires : Universidad Tecnológica Nacional. Facultad Regional de Buenos Aires, 2023.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-42-0215-8

1. Imágenes. 2. Salud. I. Rey, Andrea Alejandra , coord.

CDD 610



Estos son los Anales de las VI Jornadas Interdisciplinarias de Análisis Avanzado de Imágenes y Señales

Webpage: <https://jiaais.frba.utn.edu.ar/>

ISBN: 978-950-42-0215-8

AJEA (Actas de Jornadas y Eventos Académicos de UTN) no percibe fondos desde los congresos / jornadas, la publicación es gratuita y abierta para eventos realizados u organizados en unidades académicas de la UTN

 Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial 4.0 Internacional.

Índice

Acerca de las VI JIAAIS	1
JIAAIS	1
JIAAIS (in English)	1
Tópicos de Interés	1
Topics of interest	2
Comité Organizador	2
Comité Científico	3
Trabajos	3
Métodos de Aprendizaje Profundo en el Análisis e Interpretación de Imágenes de Microscopía	4
1 Introducción	5
2 Materiales y Métodos	5
2.1 Preprocesamiento	5
2.2 Autoencoder	6
2.3 Distancia de Jensen-Shannon	6
2.4 Transformers	6
3 Resultados y Discusión	7
4 Conclusiones	8
Desarrollo de una Aplicación Web para el Tracking de Filamentos Intracelulares a partir de Imágenes de Microscopía	10
1 Introducción	11
2 Metodología	11
2.1 Backend: Algoritmo de tracking	11
2.2 Frontend: Interfaz web	13
3 Resultados: tracking de filamentos en imágenes de microscopía de fluorescencia	14
4 Conclusiones	14
Procesamiento del habla para la clasificación de personas con enfermedad de Parkinson	16
1 Introducción	16
2 Metodología	17
2.1 Datos	17
2.2 Parámetros acústicos	17
2.3 Métodos estadísticos	17
3 Resultados	18
4 Discusión y perspectivas futuras	19
Detección del Síndrome de Muerte Súbita Empleando Matrices de Transición entre Patrones Ordinales de ECGs	21
1 Introducción	22
2 Materiales y Métodos	23
3 Resultados	23
4 Discusión	23

Extracción de Características de Imagen para Recuperación de formas 3D a partir de Múltiples Vistas		27
1	Introducción	28
2	Metodología	28
3	Resultados	29
4	Conclusiones	30
Detección de Bordos en Imágenes Satelitales Nocturnas usando Grafos y Medidas de Información		32
1	Introducción	33
2	Materiales y Métodos	33
3	Resultados y Discusión	34
4	Conclusiones	35

Acerca de las VI JIAAIS

JIAAIS

Las Jornadas Interdisciplinarias de Análisis Avanzado de Imágenes y Señales (JIAAIS) se han llevado a cabo desde 2016 de manera ininterrumpida, a excepción de los años 2020 y 2021 debido a pandemia de COVID-19. Para su organización han participado instituciones como la Universidad Tecnológica Nacional Facultad Regional Buenos Aires, el Instituto Tecnológico de Buenos Aires, la Universidad Nacional de General Sarmiento y la Universidad Federal de Alagoas (Alagoas, Brasil). El éxito obtenido en el intercambio de conocimiento de temas actuales de investigación en el área, muestran las ventajas de convocar a este tipo de encuentros entre pares, motivando la participación de estudiantes de grado y de posgrado. Las JIAAIS ofrecen una oportunidad de vinculación y comunicación tanto para investigadores como estudiantes, cuyos trabajos están relacionados con el procesamiento y análisis de imágenes, señales y video, computación gráfica, visión computacional, visualización y temas afines, entren en contacto con los últimos desarrollos en el área y exploren posibilidades de colaboración con distintos grupos de Argentina y el exterior. Durante estas jornadas, se invita a especialistas internacionales y nacionales para que dicten charlas sobre temas actuales de investigación en el área.

De esta manera, las JIAAIS permiten la participación de especialistas, investigadores en formación y estudiantes de grado y posgrado, en distintas actividades, posibilitando un efecto sinérgico dentro de la comunidad argentina de investigadores en el campo de Procesamiento de Imágenes, Procesamiento de Señales y Visión por Computadora. El principal objetivo es compartir y discutir nuevas ideas, técnicas y experiencias, así como consolidar vínculos entre diferentes grupos de trabajo.

JIAAIS (in English)

The Interdisciplinary Conference on Advanced Image and Signal Analysis (JIAAIS in Spanish) has been held uninterruptedly since 2016, with the exception of 2020 and 2021 due to the coronavirus pandemic COVID-19. Institutions such as the National Technological University Buenos Aires Regional Faculty, the Technological Institute of Buenos Aires, the National University of General Sarmiento and the Federal University of Alagoas (Alagoas, Brazil). The success obtained in the exchange of knowledge of current research topics in the area shows the advantages of convening this type of meetings between peers, motivating the participation of undergraduate and graduate students. The JIAAIS offer a networking and communication opportunity for both researchers and students, whose work is related to image processing and analysis, signals and video, computer graphics, computer vision, visualization and related topics, enter stay in touch with the latest developments in the area and explore possibilities for collaboration with different groups from Argentina and abroad. During these sessions, international and national specialists are invited to give talks on current research topics in the area.

In this way, the JIAAIS allow the participation of specialists, researchers in training and undergraduate and graduate students, in different activities, enabling a synergistic effect within the Argentine community of researchers in the field of Image Processing, Signal Processing and Computer Vision. The main objective is to share and discuss new ideas, techniques and experiences, as well as consolidate links between different work groups.

Tópicos de interés

Las VI JIAAIS incluyen -pero no se limitan a- los siguientes tópicos relacionados con imágenes, señales y visión.

- Análisis y Procesamiento de Imágenes y Señales
- Análisis Espacio-Temporal
- Análisis Inteligente de Datos
- Medidas Biométricas en Imágenes y Video
- Aprendizaje Profundo en Imágenes y Video
- Compresión, Transmisión, Almacenamiento y Recuperación de la Información
- Estadística y Técnicas de Inteligencia Artificial Aplicadas al Reconocimiento de Patrones
- Métodos de Selección y Extracción de Características, Clasificación, Segmentación y Clustering
- Procesamiento de Color, Textura y Movimiento
- Visión por computadora

Topics of interest

The VI JIAAIS include - but are not limited to - the following topics related to images, signals and vision.

- Images and Signal Analysis and Process
- Spatial-Temporal Analysis
- Intelligent Data Analysis
- Biometric Measures in Images and Video
- Deep Learning in Images and Video
- Compression, Transmission, Storage and Recovery of Information
- Statistics and Artificial Intelligence applied to Pattern Recognition
- Methods for Feature Selection and Extraction, Classification, Segmentation and Clustering
- Colour, Texture and Movement Processes
- Computer vision

Comité Organizador

Dra. Andrea Rey
Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina

Dra. Juliana Gambini
Instituto Tecnológico de Buenos Aires, Argentina
Universidad Nacional Tres de Febrero, Argentina

Comité Científico

Dr. Walter Legnani
Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina

Dr. Alejandro C. Frery
Victoria University of Wellington, Nueva Zelanda

Dr. Claudio Delrieux
Universidad Nacional del Sur, Argentina
CONICET, Argentina

Dra. Magdalena Lucini
Universidad Nacional del Nordeste, Argentina
CONICET, Argentina

Dr. Rodrigo Ramele
Instituto Tecnológico de Buenos Aires, Argentina

Dr. Gustavo Gil
Université de Montpellier, Francia

Trabajos

Los trabajos presentados en estos anales fueron sometidos a un proceso de revisión por simple par ciego. Todas las propuestas fueron evaluadas por dos o tres revisores con amplia experiencia en las áreas de investigación involucradas.

Métodos de Aprendizaje Profundo en el Análisis e Interpretación de Imágenes de Microscopía

Deep Learning Methods in Analysis and Interpretation of Microscopy Images

Eugenia Sol Piñeiro¹ Juliana Gambini²

Resumen


El Retinoblastoma es un tumor ocular caracterizado por células malignas en la retina. Para su tratamiento, los médicos aplican diversos métodos, uno de los cuales es la inyección localizada de un medicamento quimioterapéutico llamado Topotecan. Uno de los intereses de los científicos del área de medicina es estudiar la velocidad con que el Topotecan penetra en los cultivos celulares de Retinoblastoma, para lo cual se tomaron secuencias de imágenes de microscopio [1]. Sin embargo, este proceso es muy complejo debido a que el cultivo celular de Retinoblastoma requiere condiciones muy estrictas [2]. Es por ello que se cuenta un conjunto reducido de imágenes, únicas en el país y los estudios estadísticos son poco confiables. El objetivo de este trabajo es generar imágenes sintéticas que posean las mismas características que las originales, que imiten el proceso de penetración del Topotecan y que nos permitan validar los resultados. Con este propósito, utilizamos modelos de Aprendizaje Profundo como Autoencoders y Transformers. Los resultados son muy alentadores.


Palabras clave: Microscopía, Aprendizaje Profundo, Retinoblastoma, Autoencoders, Machine Learning

Abstract

Retinoblastoma is an ocular tumor characterized by malignant cells in the retina. For its treatment, doctors apply various methods, one of which is the localized injection of a chemotherapy drug called Topotecan. One of the interests of scientists in the area of medicine is to study the speed with which Topotecan penetrates into cell cultures of Retinoblastoma, for which sequences of microscope images were taken. However, this process is very complex because the cell culture of Retinoblastoma requires very strict conditions. That is why it is counted as a reduced set of images, unique in the country and statistical studies are unreliable. This work aims to generate synthetic images with the same characteristics as the original ones, that imitate the penetration process of Topotecan and allow us to validate the results. For this purpose, we use Deep Learning models such as Autoencoders and Transformers. The results are very encouraging.

Keywords: Microscopy, Deep Learning, Retinoblastoma, Autoencoders, Machine Learning

¹Dpto. de Ingeniería Informática, Instituto Tecnológico de Buenos Aires, Argentina. ✉ epineiro@itba.edu.ar  0000-0003-1385-0875

²Dpto. de Ingeniería Informática, Instituto Tecnológico de Buenos Aires; Dpto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero; Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires; Argentina. ✉ mgambini@itba.edu.ar  0000-0002-4534-1402

1 Introducción

El Retinoblastoma es el tumor ocular más frecuente en la infancia, caracterizado por células malignas en la retina. Para evitar la enucleación del órgano ocular y conservar la visión de los niños afectados, los médicos aplican diversos tratamientos, uno de los cuales es la inyección de quimioterapia en forma localizada [1, 3].

Este proyecto surge de la necesidad de evaluar la penetración de la quimioterapia en las células de Retinoblastoma y tiene como objetivo analizar la dinámica de penetración de Topotecan en cultivos de células primarias de Retinoblastoma, mediante el análisis automático de secuencias de imágenes de microscopía confocal. La elección de Topotecan es fundamental, dado que tiene una reacción fluorescente a la excitación por láser y produce imágenes con alto contraste.

El cultivo de estas células requiere condiciones muy estrictas y son muy difíciles de obtener [2]. Las muestras presentadas en ese trabajo eran las únicas en Argentina hasta ese momento. Esto implica que las imágenes tomadas son pocas para realizar un análisis estadístico confiable.

Se cuenta con un conjunto de imágenes de microscopio de cultivos celulares del tumor intraocular, al cual se le inyecta Topotecan, correspondientes a dos pacientes. El primero fue enucleado en la primera línea de tratamiento sin haber recibido previamente quimioterapia. En cambio, el segundo paciente había recibido quimioterapia antes de la enucleación del ojo afectado. En la Figura 1 se puede apreciar una muestra de una célula para cada paciente, en la cual la parte más clara (blanca) representa el Topotecan.

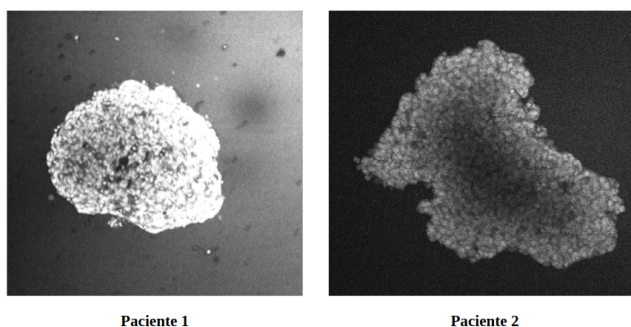


Figura 1: Muestra de una célula de tamaño grande para cada paciente.

Los cultivos celulares, llamados tumoresferas, fueron clasificados por un experto en distintos tamaños: pequeño, mediano y grande. Se tomaron imágenes de las células tras la inyección de Topotecan en intervalos de 30 segundos generando una secuencia de imágenes de microscopía.

El conjunto de datos original es reducido, se trata de 11 secuencias de imágenes de microscopía cada una para cada paciente. A las imágenes dentro de cada secuencia las llamaremos *frames*. El *frame i* representa la penetración de Topotecan en el *Tiempo i*. Siendo tan pocas imágenes se dificulta realizar un análisis predictivo y evaluar los resultados obtenidos para obtener alguna conclusión.

Este trabajo consiste en generar imágenes sintéticas con las mismas características que las originales, que imiten el proceso de penetración del Topotecan y que nos permitan validar los resultados anteriores [4, 5]. Para ello, se utilizarán distintos modelos de Redes Neuronales como Autoencoder Variacional, Transformers y Redes Convolucionales.

2 Materiales y Métodos

2.1 Preprocesamiento

Se analizó la estructura de las imágenes. Se trata de imágenes ruidosas de $1024 \times 1024 \times 3$ en escala de grises con 256 niveles de gris, en formato *.pgm*. Se cuenta con secuencias de imágenes de dos individuos. Cada uno tiene entre 2 y 5 secuencias de imágenes de células grandes, pequeñas y medianas.

En primer lugar, se implementaron técnicas de *Data Augmentation*. Se rotaron las imágenes a 90° , 180° y 270° y se invirtieron respecto del eje vertical y del horizontal.

2.2 Autoencoder

Una vez aplicado el preprocesamiento, se implementaron diversos modelos de Redes Neuronales para solucionar distintos de los problemas encontrados y alcanzar el objetivo. Para ello, se investigaron trabajos relacionados al análisis de imágenes y posibles modelos de Redes Neuronales a utilizar: Autoencoders, Redes Convolucionales y Transformers.

Los Autoencoders se introdujeron por primera vez en el año 1987 [6]. Se trata de una Red Neuronal diseñada para codificar la entrada de forma comprimida, de manera tal que al decodificarla sea lo más similar a la original. El principal objetivo del Autoencoder es aprender de forma no supervisada una representación informativa del conjunto de datos.

Su arquitectura consta de dos partes fundamentales a las que se denominó *Encoder* y *Decoder*. El *Encoder* es un Perceptrón Multicapa y el *Decoder* posee exactamente la misma arquitectura pero de forma espejada; es decir, si el *Encoder* consta de n neuronas en su primera capa, x capas ocultas y m neuronas en su capa de salida, entonces el *Decoder* tendrá m neuronas en su primera capa, x capas ocultas y n neuronas en su última capa.

Hay diversos tipos de Autoencoders dependiendo del problema que se desee resolver, ya sea para eliminación de ruido, identificación de outliers, caracterizaciones o transformaciones de los datos que son útiles para discriminarlos o para generar nuevas muestras de imágenes. Los distintos tipos de usos se comparan en el artículo [7].

En este proyecto se trabaja con el *Autoencoder Variacional* (VAE) [8]. El VAE está basado en un modelo generativo a partir de una distribución de probabilidad. El mismo permite generar imágenes nuevas recorriendo el estado intermedio entre el *Encoder* y el *Decoder*, al cual se lo denomina *Espacio Latente* (Z). La red fue entrenada utilizando el primer *frame* de todas las secuencias de imágenes correspondientes al mismo Individuo y al mismo tipo de célula (grande, pequeña o mediana). Luego, se recorre el *Espacio Latente* tomando como nueva muestra generada el centroide del grupo de puntos pertenecientes al espacio.

De esta forma, se generaron nuevas imágenes, solucionando el problema de que el conjunto de datos original es muy pequeño y se necesita aumentarlo a partir de un modelo generativo válido. Este proceso se repitió para cada uno de los *frames* generando nuevas secuencias de imágenes que simulan la penetración de Topotecan en los cultivos celulares.

2.3 Distancia de Jensen-Shannon

Para evaluar si las muestras generadas poseen las mismas características que las originales, utilizamos distintos métodos. En primer lugar, se utiliza la distancia de *Jensen-Shannon* entre los histogramas de las imágenes originales y las generadas por el VAE. La distancia de *Jensen-Shannon* entre dos funciones de probabilidad p y q , en un espacio de probabilidad \mathcal{X} , se define como:

$$JS(p, q) = \sqrt{\frac{D(p, m) + D(q, m)}{2}}, \quad (1)$$

donde $m = (p + q)/2$ y D es la Divergencia de Kullback-Leibler, dada por:

$$D(p, q) = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}. \quad (2)$$

2.4 Transformers

En el año 2017, se introduce el concepto de *Transformer* [9]. Se trata de una Red Neuronal cuyo objetivo es transformar una secuencia de elementos en otra secuencia, basada en un mecanismo de atención. A diferencia de otros tipos de redes como las Redes Recurrentes (RNN), las Redes Convolucionales o Long Short-Term Memory (LSTM) [10], permite ejecutar tareas de forma paralela reduciendo el tiempo requerido para entrenar. A su vez, puede recordar viejas conexiones en secuencias de gran tamaño evitando el problema del gradiente que va desapareciendo o crece demasiado [11].

La arquitectura de un Transformer consiste en un conjunto de *Encoders* y *Decoders*, que interactúan entre sí para a partir de una secuencia (*input*) y generar otra (*output*). Cabe destacar que se puede contar con tantos *Encoders* y *Decoders* apilados como se desee, considerando que la cantidad debe ser la misma tanto para el *encoding* como para el *decoding*. El *output* de cada *Encoder* se toma como *input* del siguiente, hasta llegar al último *Encoder* cuyo *output* es tomado por todos los *Decoders*, junto con el *output* del *Decoder* anterior. Todos los *Encoders* tienen una estructura idéntica, aunque difieren en los pesos.

La secuencia que describe el artículo original es una secuencia de palabras que forma un texto, permitiendo predecir palabras siguientes en una oración. Pero en el trabajo [12] se muestra que un Transformer es muy eficiente aplicado a la clasificación de imágenes. Incluso se compara con las Redes Neuronales Convolucionales [13], notando que computacionalmente requiere menos recursos y alcanza mejores resultados.

En este trabajo, clasificamos las imágenes utilizando Transformers, siendo el conjunto de entrenamiento las imágenes originales y el de testeo aquellas generadas por el VAE. Cada imagen se etiqueta con su respectivo frame y como métrica de evaluación estándar se utiliza *Accuracy*, la cual se puede definir como:

$$\text{Accuracy} = \frac{\text{Cantidad de predicciones correctas}}{\text{Cantidad total de predicciones}}. \quad (3)$$

3 Resultados y Discusión

Se obtuvieron distintas imágenes recorriendo el espacio latente del VAE. En la Figura 2 se puede ver la distribución de las imágenes en el espacio latente 2D generado en una ejecución del VAE y a su lado una muestra obtenida a partir del centroide del mismo. Una vez obtenidas las muestras generadas de cada uno de los *frames* (*Tiempo i*) se puede armar una secuencia, como se ve en la Figura 3.

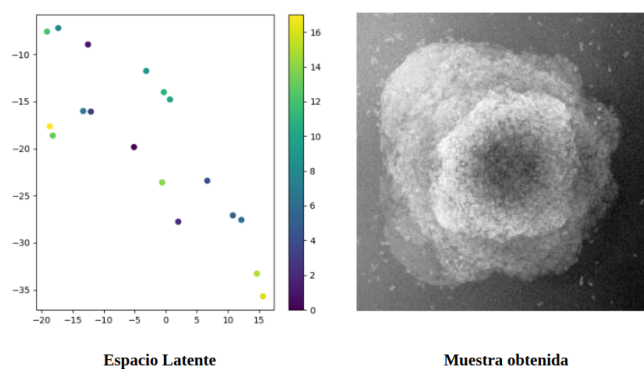


Figura 2: Espacio latente y muestra obtenida a partir del centroide.

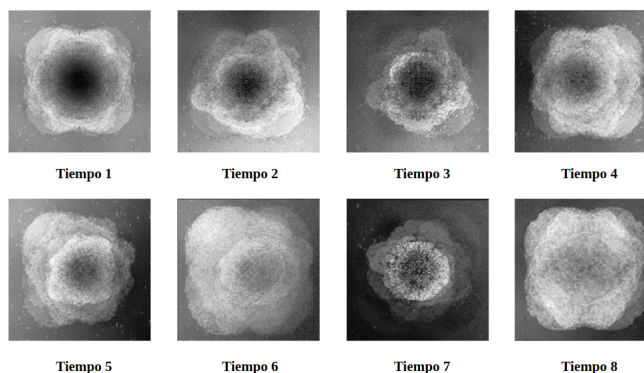


Figura 3: Ejemplo de secuencia generada por el VAE.

Luego, se calcula la distancia de *Jensen-Shannon* usando la ecuación (1) entre las imágenes originales y las obtenidas por el VAE. Para ello se tomaron todas las muestras correspondientes a un *frame* y se calculó la distancia entre ellas. Por ejemplo, para el primer *frame* en promedio se obtuvo una distancia de 0.44. Esto quiere decir que las muestras no son sumamente similares (el resultado es lejano al cero). Aún así, no se desea que lo sean dado que el objetivo es obtener muestras nuevas pero que tengan alguna semejanza con las imágenes originales.

Por otro lado, al clasificar las imágenes con el Transformer se obtuvo un valor de *Accuracy* del 66% usando la ecuación (3). Este resultado se podría mejorar entrenando la red con más épocas y variando parámetros como la tasa de aprendizaje de la misma y la dimensión de las imágenes.

Si bien ya se obtuvieron resultados, aún restan por resolver diversos disparadores, como por ejemplo: ¿Qué arquitectura del VAE es la mejor? ¿Cómo definir el concepto de *mejor*? ¿Qué métrica de evaluación se puede utilizar? ¿Cuántas neuronas y capas de red se necesitan? ¿Cómo se debe recorrer el espacio latente para obtener una muestra: usando el centroide, el punto medio o métodos de Machine Learning como por ejemplo, *K-Means* [14]? ¿Cómo se puede reducir el costo computacional para evitar reducir la dimensión de las imágenes?

4 Conclusiones

En este trabajo, logramos generar nuevas muestras a partir de un Autoencoder Variacional y con ellas generar secuencias que simulen la penetración de Topotecan en grupos celulares de Retinoblastoma.

A su vez, se evalúan las muestras generadas a partir de distintos métodos como la distancia de *Jensen-Shannon* y haciendo uso de Transformers. Los resultados permiten analizar si se pueden utilizar las imágenes generadas como base para predecir cuánto tiempo tardará el Topotecan en penetrar en la célula.

Cabe destacar que aún el proyecto se encuentra en una etapa de evaluación y análisis por lo cual se pueden obtener mejores resultados variando distintos parámetros, ya sea la dimensión de las imágenes o la arquitectura de cada una de las redes neuronales utilizadas.

Referencias

- [1] P. Schaiquevich, A. Ceciliano, N. Millan, P. Taich, F. Villasante, A. C. Fandino, J. Dominguez, and G. L. Chantada, "Intra-arterial chemotherapy is more effective than sequential periocular and intravenous chemotherapy as salvage treatment for relapsed retinoblastoma," *Pediatric Blood & Cancer*, vol. 60, no. 5, pp. 766–770, 2013. DOI:10.1002/pbc.24356.
- [2] U. Winter, R. Aschero, F. Fuentes, F. Buontempo, S. Zugbi, M. Sgroi, C. Sampor, D. Abramson, A. Carcaboso, and P. Schaiquevich, "Tridimensional retinoblastoma cultures as vitreous seeds models for live-cell imaging of chemotherapy penetration," *International Journal of Molecular Sciences*, vol. 20, p. 1077, 2019. DOI:10.3390/ijms20051077.
- [3] P. Chantada, G. Schaiquevich, "Management of retinoblastoma in children: current status," *Pediatric Drugs*, vol. 17, pp. 185–198, 2015. DOI:10.1007/s40272-015-0121-9.
- [4] M. Howlin, D. Chan, R. Ramele, and J. Gambini, "Topotecan penetration assessment in retinoblastoma cells using Shannon entropy and coefficient of variation," *XVIII workshop on Information Processing and Control (RPIC), Bahía Blanca, Argentina*, vol. 1, no. pp. xiii-xvii, 2019. DOI:10.1109/RPIC.2019.8882160.
- [5] D. Chan, U. Winter, P. Schaiquevich, R. Ramele, and J. Gambini, "Time estimation of topotecan penetration in retinoblastoma cells through image sequences analysis," in *VIII Congreso Latinoamericano de Ingeniería Biomédica*, 2019. DOI:10.1007/978-3-030-30648-9_35.
- [6] K. Holyoak, "Parallel distributed processing: explorations in the microstructure of cognition," *Science*, vol. 236, pp. 992–997, 1987.
- [7] N. Mansouri and Z. Lachiri, "Laughter synthesis: A comparison between variational autoencoder and autoencoder," in *5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, 2020. DOI:10.1109/ATSIP49331.2020.9231607.
- [8] D. Kingma and M. Welling, *An Introduction to Variational Autoencoders*. IEEE Now Foundations and Trends, 2019. DOI:10.1561/22000000056.
- [9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 1–15, 2017.

- [10] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, pp. 1735–1780, 1997. DOI:10.1162/neco.1997.9.8.1735.
- [11] K. Afebu, Y. Liu, E. Papatheou, and B. Guo, “LSTM-based approach for predicting periodic motions of an impacting system via transient dynamics,” *Neural Networks*, vol. 140, pp. 49–64, 2021. DOI:10.1016/j.neunet.2021.02.027.
- [12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” in *International Conference on Learning Representations*, 2021. DOI:10.48550/arXiv.2010.11929.
- [13] R. Xin, J. Zhang, and Y. Shao, “Complex network classification with convolutional neural network,” *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 447–457, 2020. DOI:10.26599/TST.2019.9010055.
- [14] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, “An efficient k-means clustering algorithm: Analysis and implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881–892, 2002. DOI:10.1109/TPAMI.2002.1017616.

Desarrollo de una Aplicación Web para el Tracking de Filamentos Intracelulares a partir de Imágenes de Microscopía

Development of a Web Application for Tracking of Intracellular Filaments from Images of Microscopy

Faustino Pannunzio¹ Tobías Brandy² Ignacio Sagüés³ Luciana Bruno⁴


Resumen


Presentamos el desarrollo de una aplicación web que permite realizar el seguimiento de filamentos en imágenes y recuperar las coordenadas de los mismos en los distintos cuadros de la película de manera semiautomática. La aplicación está principalmente pensada para el análisis de imágenes de microscopía de fluorescencia con el fin de estudiar la dinámica temporal de estructuras biológicas tales como filamentos del citoesqueleto, organismos flagelados, filopodios y cilias. Sin embargo, también puede ser utilizada con otro tipo de imágenes y sistemas macroscópicos. El algoritmo de tracking se basa en el análisis de los perfiles locales de intensidad a lo largo del filamento y el ajuste de una función continua lo que permite determinar las coordenadas del centro del filamento con resolución sub-píxel. El algoritmo de tracking tiene asociada una interfaz web que permite subir las imágenes al servidor, modificar los parámetros del algoritmo de ajuste y previsualizar, en tiempo real, el efecto de estas modificaciones sobre los resultados del tracking. Finalmente, el programa permite exportar los datos en diferentes formatos para su posterior análisis y generar una película con los resultados. La aplicación se encuentra desplegada en servidores del ITBA (<https://pf-pipo.it.itba.edu.ar>) y de la FCEN-UBA (<http://fernet.exp.dc.uba.ar>) accesibles al público general.


Palabras clave: Tracking, Filamentos, Fluorescencia, Microscopía Confocal, Microtúbulos


Abstract

In this work, we present the development of a web application that allows tracking filaments in images and recovering the coordinates of the same in the different frames of the film in a semi-automatic way. The application is mainly designed for the analysis of fluorescence microscopy images in order to study the temporal dynamics of structures of Biological organisms such as cytoskeletal filaments, flagellated organisms, filopodia, and cilia. However, it can also be used with other types of images and macroscopic systems. The tracking algorithm is based on the analysis of the profiles' local intensity along the filament and the adjustment of a continuous function that allows the

¹Instituto Tecnológico de Buenos Aires, Argentina. ✉ fpannunzio@itba.edu.ar  0000-0003-1276-6524

²Instituto Tecnológico de Buenos Aires, Argentina. ✉ tbrandy@itba.edu.ar  0000-0003-3508-0306

³Instituto Tecnológico de Buenos Aires, Argentina. ✉ isagues@itba.edu.ar  0000-0003-1960-1247

⁴Instituto de Cálculo, Facultad de Ciencias exactas y Naturales, Universidad de Buenos Aires, Argentina. ✉ lucianabrun@gmail.com  0000-0003-0146-2699

determination of the coordinates from the center of the filament with sub-pixel resolution. The tracking algorithm is associated with a web interface that allows uploading the images to the server, modifying the parameters of the adjustment algorithm, and previewing in real-time the effect of these modifications on the tracking results. Finally, the program allows the export of the data in different formats for its subsequent analysis and generates a movie with the results obtained. The application is deployed on ITBA (Instituto Tecnológico de Buenos Aires) servers (<https://pf-pipo.it.itba.edu.ar>) and the FCEN-UBA (Exact and Natural Science Faculty, Universidad de Buenos Aires, <http://fernet.exp.dc.uba.ar>), which are accessible to the public in general.

Keywords: Tracking, Filaments, Fluorescence, Confocal Microscopy, Microtubules

1 Introducción

Existen numerosos algoritmos para el seguimiento (tracking) con precisión nanométrica de partículas individuales de geometría tipo-esférica. Por otro lado, hacer el seguimiento de filamentos que se desplazan y transforman a lo largo del tiempo es un trabajo más difícil y existen menos estrategias (ver, por ej. [1–4]).

Una alternativa más reciente es el algoritmo AFTER (Automated Filament Tracking and Evaluation Routine) [5], desarrollado en el Grupo de Dinámica Intracelular de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (FCEN-UBA). Esta es una herramienta que permite interactuar con un algoritmo desarrollado en Matlab mediante una interfaz gráfica. El mayor problema encontrado con AFTER es, en primer lugar, el hecho de que Matlab no es un lenguaje de código abierto, por lo tanto se requiere una licencia para utilizarlo, tiene que ser instalado localmente y cuenta con una baja performance a la hora de hacer un seguimiento. Además, se destaca la complejidad que podría representar, debido al lenguaje utilizado, realizar modificaciones sobre la interfaz.

Resolver esta problemática es de particular interés para los investigadores del Grupo de Dinámica Intracelular. Para ello propusimos una rutina de seguimiento implementada en Python, lenguaje de código abierto (Sección 2), a la que se accede vía una aplicación web que minimiza los requisitos necesarios para utilizar la herramienta (Sección 2). Finalmente, mostramos los resultados de la aplicación para el tracking de espermatozoides y microtúbulos en imágenes de microscopía de fluorescencia (Sección 3).

2 Metodología

2.1 Backend: Algoritmo de tracking

El primer paso en la construcción de la aplicación fue desarrollar el algoritmo para realizar el seguimiento del filamento seleccionado por el usuario a través de la secuencia de imágenes recibida. Para la definición de este algoritmo se parte de la hipótesis de que los movimientos de los filamentos a seguir no son bruscos, es decir, que de una imagen a otra su posición no se ve alterada en gran medida, y que los mismos son curvas suaves.

El método propuesto consiste en calcular la posición del filamento a lo largo de una serie de fotogramas. Para esto, usando como valor inicial la última posición conocida del filamento, se realiza un análisis sobre el entorno y se obtiene la nueva posición. En el caso de la primer iteración esta aproximación inicial será provista por el usuario. En los demás casos, se utiliza el valor obtenido en la iteración previa.

En cada iteración del algoritmo, se realiza una interpolación sobre los puntos iniciales, ya sea la selección provista por el usuario en el caso de la primer iteración, o los obtenidos en el paso anterior. Luego, para cada punto analizado, se obtiene una recta en la dirección normal al filamento. El largo de esta podrá ser configurado. Desplazándose sobre dichas rectas, se busca el punto correspondiente con el centro del filamento. Para encontrarlo se traza un perfil de intensidad en esta dirección. Teniendo en cuenta que la difracción de la luz generada en el proceso de captura tiene una distribución gaussiana, se realiza un ajuste a una función de este tipo, buscando la media de la misma. Luego de realizar esto para cada punto, se podrá opcionalmente, aplicar una función de suavizado basada en curvas de Bezier. Se realiza una iteración por cada imagen recibida.

La Figura 1 representa esquemáticamente los principales pasos que realiza el algoritmo para la recuperación de las coordenadas centrales del filamento y que se describen en las secciones siguientes.

Determinación aproximada de la forma del filamento para la primera iteración

El primer paso del algoritmo es realizado a partir del primer fotograma de la secuencia de imágenes y una selección de puntos, provista por el usuario, que representan la aproximación inicial del filamento. Estos puntos provistos

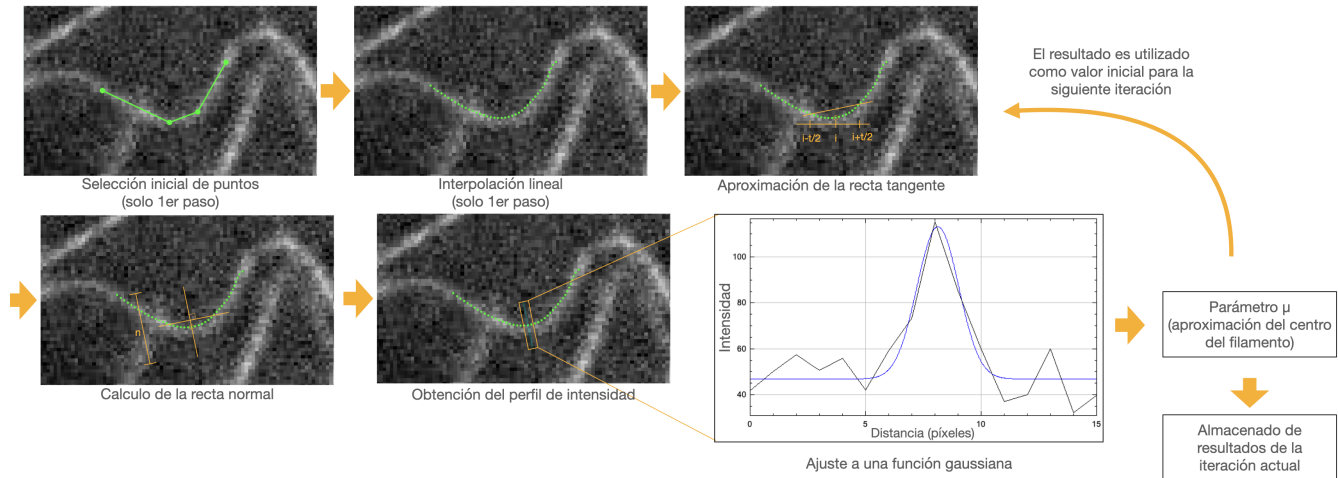


Figura 1: Figura esquemática con los principales pasos del backend.

son los mínimos necesarios para poder detectar la forma del filamento. Con estos puntos se genera la primera aproximación del filamento mediante una interpolación lineal de los mismos. La densidad de puntos permite regular la cantidad de puntos utilizados en el análisis. La misma podrá ser modificada por configuración. De esta manera se podrá ajustar el análisis para balancear entre la precisión y tiempo requerido. Una vez realizada la primer interpolación, comienza el proceso iterativo. Este es repetido por cada fotograma con la aproximación final del fotograma anterior.

Ajuste de los perfiles de intensidad en la dirección normal al filamento

Para el cálculo de la dirección tangente del filamento se hace un análisis del entorno del punto. En particular, se toman los extremos de un entorno de longitud variable, centrado en el punto y se calcula la pendiente entre los mismos. La cantidad de puntos a tener en cuenta se recibe como parámetro del algoritmo. Aproximada la dirección de la tangente, se podrá obtener la dirección de la recta normal. Luego, es necesario definir el segmento de la misma cuyo perfil de intensidad se desea analizar. Estará centrado en la última posición conocida y tendrá una longitud que también será recibida como parámetro del algoritmo. Definido cada segmento normal, será necesario obtener las intensidades de los píxeles asociados a esas posiciones.

Las intensidades obtenidas serán ajustadas a una función gaussiana con 4 parámetros:

$$g(x) = y_0 + b \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \quad (4)$$

donde μ es el parámetro de interés y representa la posición central del filamento, σ será una proporción al ancho del filamento, mientras que y_0 y b corresponden a la intensidad basal debida al ruido y a la amplitud de la señal, respectivamente.

Para realizar el ajuste es necesario tomar buenos valores iniciales. En el caso de estos 4 parámetros utilizamos $\mu = x_{\max}$, $\sigma = 1$, $b = b_{\max}$, $y_0 = b_{\max}/4$. Siendo b_{\max} la intensidad máxima medida en el perfil que se está analizando y x_{\max} cumple que $g(x_{\max}) = b_{\max}$. Es posible que no se puedan ajustar los puntos o que el ajuste logrado no sea bueno. El error del ajuste deberá estar por debajo de un valor máximo que será parámetro del algoritmo. En caso de no encontrarse por debajo del límite, el punto será marcado como inválido e interpolado a partir de los puntos de su entorno. El error del ajuste lo calculamos como la raíz de la variación de la estimación de μ .

Reposición de coordenadas longitudinales por interpolación

Si se descartasen aquellos puntos que no son ajustados correctamente se perdería información lentamente y se generarían discontinuidades en el filamento afectando las aproximaciones siguientes. Para solucionar este problema cada punto o conjunto de puntos no ajustado correctamente es reemplazado por uno nuevo obtenido a partir de

interpolan los puntos vecinos que si fueron ajustados correctamente. Para esto en cada extremo del segmento a interpolar se necesitan una cantidad determinada de puntos correctos la cual también será recibida como parámetro del algoritmo. El conjunto de puntos de ambos extremos será ajustado con una regresión lineal para obtener una aproximación de la recta que los une. Por último, se generan puntos equiespaciados contenidos en la recta calculada.

Suavizado de la forma recuperada

Se incorporó una rutina final (opcional) de suavizado de la forma recuperada de los filamentos basada en curvas de Bezier, tomando las coordenadas recuperadas como los puntos de control. A partir de la curva continua resultante se muestrean las coordenadas finales recuperadas para cada cuadro de la película.

Una consecuencia negativa de este procedimiento es que no permitiría recuperar cambios abruptos en la curvatura del filamento. Para contemplar estos casos, no nos limitamos a tomar una sola curva de Bezier para todo el filamento, sino que partimos el resultado en secciones con un mismo número de puntos, que será recibido por configuración. Luego, ajustamos a una Bezier cada uno, y por último unimos los resultados. Esto es posible gracias a que los puntos extremos de cada intervalo no se ven modificados por el ajuste.

2.2 Frontend: Interfaz web

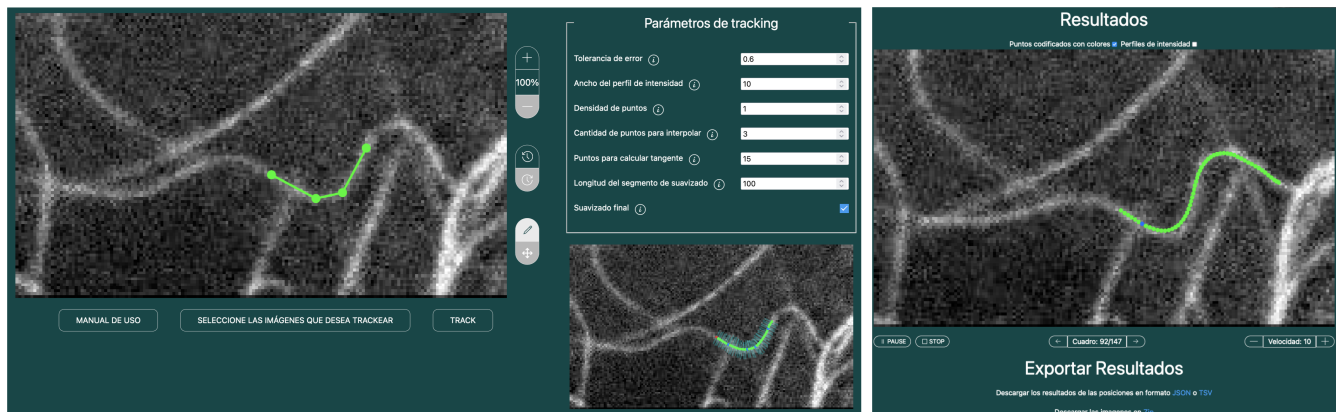


Figura 2: Captura de pantalla de la interfaz. A la izquierda el tablero de parámetros, imagen para la selección y previsualización. A la derecha visualizador de resultados. Se puede acceder a <http://fernet.exp.dc.uba.ar> para visualizar con mayor detalle.

Para poder interactuar con el algoritmo mencionado se diseñó una interfaz web a la que se pueden subir las imágenes y permite al usuario determinar algunos puntos del filamento a trackear en el primer frame de la película. Los formatos soportados son: **tiff**, **jpeg** y **png**. Para facilitar la selección, se traza la recta que los interpola, se brinda la posibilidad de realizar zoom, desplazarse en la imagen, deshacer y rehacer la selección de los puntos.

La interfaz cuenta con una tabla de parámetros definidos previamente (ver detalles en la figura 2). Los valores de los parámetros pueden modificarse para mejorar la performance. Para una rápida inspección de estos cambios sobre el tracking, se ofrece una previsualización. De esta forma puede analizarse si la configuración brinda, en principio, los resultados esperados, evitando analizar la secuencia completa, pudiéndose modificar parámetros y recibiendo un primer resultado en tiempo real.

Una vez obtenida una configuración aceptable se puede ejecutar el algoritmo y se muestra, una vez finalizado el proceso, un visualizador con los resultados. Este cuenta con controles para poder ver toda la secuencia de imágenes junto con los resultados del tracking, pudiendo adelantar y retroceder sobre la misma. Por último, es posible descargar estos resultados en formato **json** y **tsv** o todos los fotogramas de la película en formato **png**, comprimidos en un **zip**.

3 Resultados: tracking de filamentos en imágenes de microscopía de fluorescencia

En la Figura 3A se muestra el resultado de la aplicación al seguimiento del movimiento del flagelo de un espermatozoide. Este escenario cuenta con una imagen de muy buena calidad, donde la figura está claramente delimitada y con un gran contraste respecto al fondo. Por otro lado, este caso presenta mucho movimiento y de forma muy rápida. En este caso el error de ajuste es bajo, pero se esperan dificultades a la hora de seguir los movimientos de mayor amplitud. Esta película de 100 fotogramas fue procesada en 28s.

La Figura 3B muestra una secuencia de imágenes correspondientes a microtúbulos fluorescentes en células vivas. Este caso es una película de mucha menor resolución (166×96 px) y los obstáculos más importantes para la rutina son el ruido y la interferencia con otros filamentos vecinos. Esta película de 330 fotogramas fue procesada en 59s. Cabe destacar que a lo largo de toda la película el filamento fue identificado de forma exitosa. Se puede ver como la rutina es capaz de seguir los grandes cambios de curvatura. Por último, se puede observar en el tercer cuadro una desviación respecto al filamento que se estaba siguiendo debido a la intersección con otros filamentos. Las películas mencionadas, junto con los parámetros utilizados para el tracking, se encuentran alojadas en: https://drive.google.com/drive/folders/1ZHX5K9ye75i7_GHLZXigDM4YoKyvxPno?usp=sharing.

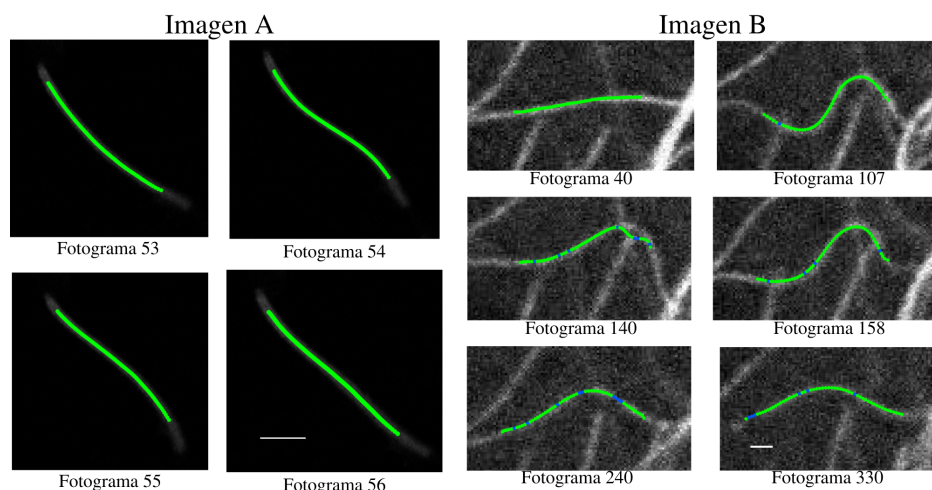


Figura 3: A: Cuatro fotogramas consecutivos en blanco y negro de un segmento del flagelo de espermatozoides de ratón marcados con mitoTracker en solución acuosa tomadas con un microscopio Airyscan. En verde se muestra el segmento recuperado por la aplicación. Los puntos azules indican los que debieron ser interpolados a partir los puntos de su entorno. Escala: 4.26 μ m. Las imágenes son gentileza del BioIng. Mariano González, del laboratorio de Bases moleculares de la fertilización, Instituto de Biología y Medicina Experimental - CONICET. B: Imágenes de microscopía confocal de microtúbulos fluorescentes en células de rana y el tracking de un segmento de microtúbulo. Escala: 820 nm. Imágenes cedidas por el Grupo de Dinámica Intracelular, FCEN-UBA.

4 Conclusiones

Se logró construir un algoritmo que, no sólo está escrito en un lenguaje de código abierto como Python, sino que es configurable y sus tiempos de respuesta se encuentran dentro de los márgenes deseados. El mismo es robusto al ruido, tanto en la totalidad de la imagen como en zonas puntuales durante el seguimiento. Por otro lado, en imágenes de muy alta relación señal/ruido se logra tolerar movimientos bruscos del filamento sin perderlo. Por último, cabe destacar la resistencia del algoritmo ante cruces con otros durante el proceso de seguimiento, fallando bajo condiciones altamente complejas, donde en pocas iteraciones continuará con el seguimiento correctamente.

La interfaz web permite que interactuar con el algoritmo resulte sencillo, permitiendo ajustar los parámetros de la rutina de seguimiento para poder así, obtener rápidamente resultados útiles. Por último, al poder usarse de manera on-line, se reduce la curva de entrada al mismo, ahorrando el trabajo de instalar la aplicación localmente,

con las dificultades que esto puede conllevar.

Referencias

- [1] F. Gittes, B. Mickey, J. Nettleton, and J. Howard, “Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape,” *The Journal of Cell Biology*, vol. 120, no. 4, pp. 923–934, 1993. DOI:10.1083/jcb.120.4.923.
- [2] J. Käs, H. Strey, J. Tang, D. Finger, R. Ezzell, E. Sackmann, and P. Janmey, “F-actin, a model polymer for semiflexible chains in dilute, semidilute, and liquid crystalline solutions,” *Biophysical Journal*, vol. 70, no. 2, pp. 609–625, 1996. DOI:10.1016/S0006-3495(96)79630-3.
- [3] C. P. Brangwynne, G. H. Koenderink, E. Barry, Z. Dogic, F. C. MacKintosh, and D. A. Weitz, “Bending dynamics of fluctuating biopolymers probed by automated high-resolution filament tracking,” *Biophysical Journal*, vol. 93, no. 1, pp. 346–359, 2007. DOI:10.1529/biophysj.106.096966.
- [4] F. Ruhnnow, D. Zwicker, and S. Diez, “Tracking single particles and elongated filaments with nanometer precision,” *Biophysical Journal*, vol. 100, no. 11, pp. 2820–2828, 2011. DOI:10.1016/j.bpj.2011.04.023.
- [5] C. Pallavicini, V. Levi, D. E. Wetzler, J. F. Angiolini, L. Benseñor, M. A. Despósito, and L. Bruno, “Lateral motion and bending of microtubules studied with a new single-filament tracking routine in living cells,” *Biophysical Journal*, vol. 106, no. 12, pp. 2625–2635, 2014. DOI:10.1016/j.bpj.2014.04.046.

Procesamiento del habla para la clasificación de personas con enfermedad de Parkinson

Speech Processing for Classification of People with Parkinson's Disease

Mónica Giuliano¹ Humberto Torres² Verónica Aubin³

Resumen

Aquí se presentan los avances realizados en el procesamiento comparativo del habla de personas con y sin enfermedad de Parkinson del mismo rango etario. Los experimentos se han realizado con muestras de fonación de la vocal /a/ a las cuales se le han extraído 297 parámetros. De este conjunto de variables se seleccionaron 11, utilizando un criterio de selección en etapas, usando sucesivamente las técnicas de análisis de componentes principales, análisis de la variancia y modelos de regresión logística. Se logró un modelo con una precisión del 80% en la clasificación de personas con y sin Parkinson, y se identificaron las 11 variables más relevantes en el proceso de clasificación.

Palabras clave: Habla, Enfermedad de Parkinson, Análisis Acústico, Clasificación automática


Abstract

In this work, advances in comparative speech processing of people with and without Parkinson's disease of the same age range are presented. The experiments have been carried out with phonation samples of the vowel /a/ from which 297 parameters have been extracted. From this set of variables, 11 were selected, using a selection criterion in stages, successively using the techniques of principal component analysis, analysis of variance, and logistic regression models. A model with an accuracy of 80% was achieved in the classification of people with and without Parkinson's disease, identifying the 11 most relevant variables used for the classification.


Keywords: Speech, Parkinson's Disease, Acoustic Analysis, Automatic Classification

1 Introducción

La enfermedad de Parkinson (EP) es un trastorno neurodegenerativo crónico y progresivo del sistema nervioso central, de comienzo gradual, sin sintomatología obvia y progresión lenta, diagnosticándose en edades medias entre 40 y 70 años. Las manifestaciones motoras de la EP están representadas por temblor de reposo, rigidez y bradicinesia [1]. En su evolución natural se agregan otros síntomas como trastornos de la marcha, desequilibrio y alteraciones de la voz, que implican un marcado impacto en la calidad vida [2]. Se estima que entre un 60 – 80% de los pacientes con EP presentan alteraciones de la voz, caracterizados por cambios en frecuencia, duración e intensidad del habla [3].

¹Universidad Nacional del Oeste e Instituto de Inmunología, Genética y Metabolismo, UBA-CONICET, Argentina ✉ mguliano@uno.edu.ar  0000-0001-5883-5535

²Instituto de Inmunología, Genética y Metabolismo, UBA-CONICET, Argentina ✉ hmtorres@conicet.gov.ar  0000-0001-5913-161X

³Universidad Nacional de la Matanza, Argentina ✉ vaubin@unlam.edu.ar  0000-0002-4981-6460

El análisis acústico de la voz permite detectar los cambios de los parámetros vocales, lo cual puede ser utilizado para predecir el avance de la EP y definir una intervención clínica específica. Recientemente, Goyal *et al.* [4] han presentado una revisión comparativa de trabajos realizados con publicaciones entre 2017 y 2019, donde se hace foco en las diferentes técnicas de extracción y clasificación de características.

Aquí se realiza un resumen de algunos análisis de muestras con la fonación de la vocal /a/ en una base de datos propia, utilizando alrededor de 300 parámetros iniciales con algoritmos obtenidos de la bibliografía [5, 6] y con una reducción de variables para la predicción automática de la EP. Se comparan los resultados con los obtenidos por otros autores y se destaca el método de análisis para la reducción y selección de variables.

2 Metodología

2.1 Datos

Se dispone de una base de datos grabada en el año 2019 [7] de voces de personas con y sin EP. Pacientes con EP del Hospital Nacional Profesor Alejandro Posadas de la Provincia de Buenos Aires, tratados en el servicio de Neurología Funcional, se presentaron en el Hospital Bernardino Rivadavia de la Ciudad de Buenos Aires, donde participaron de la construcción de la base de datos de voces de pacientes con EP junto con el apoyo técnico de la Universidad Nacional de La Matanza (UNLaM). Adicionalmente, en la UNLaM se construyó una base de datos con habla de personas sin EP diagnosticado, considerando condiciones demográficas similares. El promedio de edad de los pacientes con EP y de las personas sin EP es similar, de 64 años. Las personas con EP tenían en el momento de la toma de la muestra una antigüedad media de la enfermedad de 6 años con avance leve a moderado, además se encontraban medicados en el momento de toma de la muestra. La base de datos utilizada consta de un total de 108 individuos, 46 con EP y 62 sin EP, 59 mujeres y 49 varones [8].

2.2 Parámetros acústicos

Se utilizaron los parámetros propuestos en [8] para el análisis de la señal de voz, generados con el Voice Analysis Toolbox (VAT) [6, 9], el cual consta de una serie de rutinas en MATLAB optimizadas para la pronunciación de la letra /a/. El VAT entrega 339 medida de disfonía. Se toman los parámetros agrupados según [8] en 4 grupos relacionados con variaciones en la señal acústica y afecciones de la voz del paciente con EP, amplificando el análisis de la voz patológica en cuanto a la variación de amplitud y período de la curva de la frecuencia fundamental (F0), la emisión de ruido por el cierre incompleto de las cuerdas vocales y los problemas de articulación. El grupo 1 (G1) detectan variación en la periodicidad y entrega las medidas típicas de las variaciones de la curva de F0 (Jitter), junto con medidas basadas en análisis no lineal de señales y parámetros derivados de los coeficientes de aproximación y detalles del Análisis Onditas (Wavelets). El grupo 2 (G2) permite análisis en la variación de la amplitud de la curva de F0 (Shimmer). El grupo 3 (G3) involucra problemas relacionados con el cierre incompleto de las cuerdas vocales y medidas de ruido, en métricas tradicionales y no lineales. El grupo 4 (G4) considera problemas en la articulación, reflejados en los coeficientes cepstrales en escala Mel (MFCC, del inglés Mel-Frequency Cepstral Coefficient).

2.3 Métodos estadísticos

En los datos que se analizaron, el número de variables supera ampliamente el número de casos por lo que se hizo necesaria la implementación de modelos estadísticos que eviten el sobreajuste y tomen en cuenta la multicolinealidad. Este análisis se realizó en dos etapas. En la primera, se seleccionaron 35 variables utilizando Análisis de Componentes Principales (ACP) y análisis de la varianza (ANOVA, del inglés ANalysis Of VAriance), y en la segunda, se ajustaron modelos de regresión con distintas opciones de penalización LASSO (del inglés Least Absolute Shrinkage and Selection Operator) y Elastic net.

En primer lugar, se realizó un análisis exploratorio que permitió un filtrado inicial de 297 de las 339 variables, todas presentes en el conjunto de medidas de disfonía que estima el VAT. Luego, se realizaron pruebas de comparación de medias de las 297 variables considerando las categorías de género, condición de EP y la interacción entre ambas. Tomando en cuenta las que mostraban diferencias significativas, fueron seleccionadas 240 variables a las que se aplicó ACP [10] para detectar grupos de alta correlación y se identificaron las 9 primeras componentes. Luego, se consideraron las variables con mayor peso vinculado a cada una de estas componentes, quedando finalmente seleccionadas 29 variables con carga factorial mayor a 0.4, en uno o más componentes, y con comunalidad mayor a

0.7. El procedimiento utilizado coincide con el propuesto anteriormente en [11, 12]. Para realizar una comparación con el trabajo de Tsanas *et. al* [6], se agregaron 6 variables a las 29 anteriormente seleccionadas, quedando 35 variables de los 4 grupos según el tipo de medida de disfonía, descriptos previamente.

En la siguiente etapa se utilizaron modelos de regresión logística, dada la cantidad de variables independientes y la multicolinealidad existente entre ellas, se consideraron métodos de penalización, LASSO y Ridge [5, 13]. La penalización Ridge se aplica generalmente para resolver problemas de multicolinealidad, mientras que LASSO además de esta tarea, introduce una penalización sobre los coeficientes que permite seleccionar las variables de entrada (medida de disfonía) más relevantes para explicar la variable respuesta (condición de EP), logrando modelos interpretables y parsimoniosos. Se estimaron los coeficientes según distintas propuestas de penalización, eligiendo el parámetro λ (tuning parameter o parámetro de sintonización o ajuste) a través de la minimización de devianza en la validación cruzada según las recomendaciones en [13]. Se consideraron dos alternativas óptimas para λ : el mínimo obtenido por validación cruzada (λ_{\min}) y el mínimo obtenido más un desvío estándar (λ_{1se}). Los 4 modelos utilizados fueron: Modelo 1L (penalización LASSO y $\lambda = \lambda_{\min}$); Modelo 2L (penalización LASSO y $\lambda = \lambda_{1se}$); Modelo 1E (Elastic net con 70% de penalización LASSO y 30% Ridge, y $\lambda = \lambda_{1se}$); Modelo 2E (Elastic net con 50% de penalización LASSO y 50% Ridge, y $\lambda = \lambda_{1se}$).

Considerando como objetivo seleccionar las variables más significativas para la clasificación, en cada modelo se procedió iterativamente eligiendo al azar conjuntos de entrenamiento y validación (el 75% y 25% del total, respectivamente), registrando el porcentaje de veces que cada variable queda seleccionada sobre un total de 50 iteraciones. Para medir la capacidad predictiva del modelo se consideró el área bajo la curva ROC (AUC, del inglés Area Under Curve ROC, Receiver Operating Characteristic), resultante en cada iteración del modelo. Esta medida fue utilizada para resumir la precisión diagnóstica del modelo clasificación binaria (con EP o sin EP) bajo estudio [13].

3 Resultados

Se registró la frecuencia con que apareció cada una de las 35 variables consideradas en las 50 iteraciones de cada uno de cuatro modelos ajustados, 2 LASSO (1L y 2L) y 2 Elastic net (1E y 2E). Con las 35 variables y las 50 iteraciones se obtuvo una clasificación con precisiones promedio entre 80% y 84% de AUC. Además, se identificaron las 11 variables que aparecieron con mayor frecuencia, con incidencia promedio mayor al 24.5%, como se observa en la Tabla 1.

Tabla 1: Frecuencia promedio de aparición de las 11 variables más relevantes según los 4 modelos y los 4 grupos. Fuente [8].

		% de veces que queda elegida cada variable en los modelos (de 50 iteraciones)				
Variable		LASSO		elastic net		Promedio
Grupo		Modelo 1L	Modelo 2L	Modelo 1E	Modelo 2E	
G1	V1	26	12	22	40	25.0
	V51	56	18	28	18	30.0
G2	V34	70	40	48	58	54.0
G3	V338	96	92	98	100	96.5
	V59	70	22	26	42	40.0
	V70	36	10	10	42	24.5
G4	V137	96	92	96	100	96.0
	V141	90	62	80	90	80.5
	V152	96	94	100	100	97.5
	V71	66	26	36	54	45.5
	V75	98	92	96	100	96.5

Entre las variables seleccionadas se destacó la presencia de medidas de disfonías en los cuatro grupos teóricos

utilizados como referencia. Las variables más relevantes en la selección fueron las medidas de MFCC (G4) relacionadas con el posicionamiento de los articuladores en el tracto vocal en la pronunciación del fonema /a/. En segundo lugar, se destacó una medida de ruido no lineal que se relaciona con el cierre incompleto de las cuerdas vocales (G3). Esto es consistente con los hallados perceptualmente por profesionales de la salud.

4 Discusión y perspectivas futuras

El criterio de selección en etapas, utilizando ANOVA, ACP y modelos de regresión logística con penalización Ridge y LASSO permitió a partir de 297 métricas acústicas jerarquizar y seleccionar solo 11, logrando una precisión promedio del 80% para la clasificación de personas con y sin Parkinson.

Las técnicas combinadas de ANOVA y ACP se aplicaron con mayor número de variables que individuos, lo que resultó útil para luego aplicar los modelos logísticos con menos variables y menor colinealidad. Esta metodología ya fue utilizada por los autores en trabajos anteriores [11, 12, 14], donde se mostró que la preselección permitió alcanzar posteriormente precisiones promedio de por lo menos el 80% en modelos logísticos aplicados para la clasificación según condición de EP.

Los resultados obtenidos en [8], usando una base de datos de 108 locutores creada para esta tarea, fueron comparados con los resultados obtenidos por Tsanas *et al.* [6] para una muestra de 126 locutores. El número de casos en ambas bases es similar, los resultados son comparables y las variables seleccionadas coinciden parcialmente.

Revisando las variables utilizadas en [8], se observó poco representadas las medidas asociadas al Shimmer, variables del G2 y se profundizó esta línea complementando con medidas propuestas en [14]. Es importante analizar los algoritmos involucrados en los softwares de análisis acústicos, ya que los mismos nombres pueden involucrar diferentes valores para las medidas acústicas, como se observó en [15].

También se ha realizado con la misma base de datos un análisis de las imágenes generadas con los espectrogramas de audios de la base con técnicas de deep learning [16] en busca de diferenciación del habla de personas con y sin EP. En 2022 [17] se comparó las 11 medidas seleccionadas en [8] con metodología de Redes Neuronales Convolucionales (en inglés CNN) a partir de imágenes del espectrograma [17]. Se utilizó la misma muestra de audios y se obtuvo una exactitud de 85% y 88%, respectivamente, que resulta similar con métodos de análisis muy distintos.

Además de la vocal /a/ sería interesante analizar la fonación de las demás vocales y las frases. Por ejemplo, con la misma base de datos se ha analizado medidas vinculadas al triangulo vocal con fonaciones con las vocales /a/, /i/ y /u/ considerando los dos primeros formantes [18]. Con otras variables se volvió a lograr una precisión promedio del 78% en la clasificación, estas variables que podrían compararse con las 11 variables y sumarse a los modelos de clasificación.

Se espera aumentar el número de voces de la base de datos con la utilización de un Sistema de Asistencia para Evaluación del habla y el Diagnóstico Acústico (SEHDA) que está en desarrollo [19] en la Universidad Nacional del Oeste. En general, los softwares proponen algoritmos que comparan su precisión de la fonación comparándola con voces artificiales o voces sin patología, y el objetivo buscado con SADA, en cambio, es tener en cuenta voces con presbifonia y alteraciones que son habituales en las personas con el avance de la edad.

Agradecimientos

Universidad Nacional del Oeste. Proyecto de Investigación. “Estudio del habla de pacientes con Enfermedad de Parkinson para la asistencia al diagnóstico y seguimiento”. Instituto de Inmunología, Genética y Metabolismo UBA-CONICET

Referencias

- [1] J. Jankovic, “Enfermedad de Parkinson, características clínicas y diagnóstico,” *Journal of Neurology, Neurosurgery & Psychiatry*, pp. 368–376, 2008.
- [2] M. Köllensperger, F. Geser, K. Seppi, M. Stampfer-Kountchev, M. Sawires, C. Scherfler, S. Boesch, J. Mueller, V. Koukouni, N. Quinn, *et al.*, “Red flags for multiple system atrophy,” *Movement Disorders: Official Journal of the Movement Disorder Society*, vol. 23, no. 8, pp. 1093–1099, 2008. DOI:10.1002/mds.21992.

- [3] M. De Letter, P. Santens, and J. Van Borsel, "The effects of levodopa on word intelligibility in Parkinson's disease," *Journal of Communication Disorders*, vol. 38, no. 3, pp. 187–196, 2005. DOI:10.1016/j.jcomdis.2004.09.001.
- [4] J. Goyal, P. Khandnor, and T. C. Aseri, "Classification, prediction, and monitoring of Parkinson's disease using computer assisted technologies: a comparative analysis," *Engineering Applications of Artificial Intelligence*, vol. 96, p. 103955, 2020. DOI:10.1016/j.engappai.2020.103955.
- [5] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996. DOI:10.1111/j.2517-6161.1996.tb02080.x.
- [6] A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 181–190, 2013. DOI:10.1109/TNSRE.2013.2293575.
- [7] M. G. Giuliano, D. Adamec, and M. Debas, "Construcción de una base de voz de personas con y sin enfermedad de Parkinson," *Revista Digital del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza (REDDI)*, vol. 6, no. 1, 2021.
- [8] M. Giuliano, L. Fernandez, and S. Pérez, "Selección de medidas de disfonía para la identificación de enfermos de Parkinson," in *2020 IEEE Congreso Bienal de Argentina (ARGENCON)*, pp. 1–8, IEEE, 2020. DOI:10.1109/ARGENCON49523.2020.9505554.
- [9] A. Tsanas, *Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning*. PhD thesis, Oxford University, UK, 2012.
- [10] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.
- [11] M. Giuliano, A. García-López, S. Pérez, F. D. Pérez, O. Sposito, and J. Bossero, "Selection of voice parameters for Parkinson's disease prediction from collected mobile data," in *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pp. 1–3, IEEE, 2019. DOI:10.1109/STSIVA.2019.8730219.
- [12] S. Pérez and M. Giuliano, "Utilización de grabaciones con teléfonos móviles para la predicción de la enfermedad de parkinson," *Actas CONAIISI*, 2019.
- [13] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009. DOI:10.1007/978-0-387-21606-5.
- [14] S. Pérez, L. Fernández, and M. Giuliano, "Aporte de medidas de Shimmer para la detección de enfermedad de parkinson," in *Jornada Avances En Métodos Estadísticos Workshop Advances In Statistical Methods*, 2021.
- [15] N. G. Elisei, "Análisis acústico de la voz normal y patológica utilizando dos sistemas diferentes: ANAGRAF y PRAAT," *Interdisciplinaria*, vol. 29, no. 2, pp. 271–286, 2012.
- [16] R. Guatelli, V. I. Aubin, M. Mora, J. Naranjo-Torres, and A. Sinopoli, "Detección de Parkinson mediante espectrogramas en color y redes neuronales convolucionales," in *II Simposio Argentino de Imágenes y Visión (SAIV 2021)-JAIIO 50 (Modalidad virtual)*, 2021.
- [17] R. Guatelli, M. Giuliano, V. Aubin, L. Fernández, S. Pérez, and L. Pepe, "Análisis comparativo entre CNN y modelos logísticos para detección de la enfermedad de Parkinson utilizando la voz." No publicado, 2022.
- [18] M. Giuliano, L. Fernández, S. N. Pérez, and A. Renato, "Evaluación de áreas del espacio vocal y formantes para caracterizar personas con y sin enfermedad de Parkinson," in *2022 IEEE Biennial Congress of Argentina (ARGENCON)*, pp. 1–8, IEEE, 2022. DOI:10.1109/ARGENCON55245.2022.9939816.
- [19] M. Giuliano, S. N. Pérez, and N. A. Mangiarua, "Estudio del habla de pacientes con enfermedad de parkinson y desarrollo de aplicación web," in *XXIV Workshop de Investigadores en Ciencias de la Computación (WICC 2022, Mendoza)*, 2022.

DetECCIÓN DEL SÍNDROME DE MUERTE SÚBITA EMPLEANDO MATRICES DE TRANSICIÓN ENTRE PATRONES ORDINALES DE ECGs *Sudden Death Syndrome Detection Using Transition Matrices Between Ordinals Patterns of ECG*

Pablo Martínez Coq¹ Andrea Rey² Walter Legnani³

Resumen

En el presente trabajo se propone una novedosa metodología para discriminar señales provenientes de ECGs normales de otras de pacientes que han padecido el síndrome de muerte súbita (SMS). Se han empleado dos bases de datos de PhysioNet, por un lado una con registros de electrocardiogramas (ECGs) normales y otra de sujetos jóvenes, sin medicación ni enfermedades cardiovasculares preexistentes que padecieron el SMS. En base al cálculo de la probabilidad de transición entre patrones ordinales de las señales, se construyen las matrices de probabilidad de transición entre los mismos. Dichas matrices presentan valores medios muy bien diferenciados entre ambas clases, constituyendo de esta manera un poderoso, a la vez que original, método de clasificación computacionalmente eficiente. Dada la importancia de la detección en etapas tempranas del SMS, la aplicación de la presente propuesta tiene un enorme interés tanto en áreas de Ciencias de la Salud, como en el mismo procesamiento de señales.

Palabras clave: Muerte Súbita, Patrones Ordinales, Transición entre Patrones Ordinales, Matrices Estocásticas, Distancia de Manhattan

Abstract

In the present work, a novel methodology is proposed to discriminate signals from normal ECGs of others from patients who have suffered sudden death syndrome (SDS). Two PhysioNet databases have been used, one with normal electrocardiogram (ECG) records and another of young patients, without medication prescription or pre-existing cardiovascular diseases but who suffered from SDS. Based on the computation of the probability of transition between ordinal patterns of the signals, the transition probability matrices between them are constructed. The mean values of these matrices can be very well differentiated between both classes, thus constituting a powerful, as well as original, computationally efficient classification method. Given the importance of detection in the early stages of SDS, the application of this proposal is of great interest both, in the areas of Health Sciences and in signal processing itself.

¹Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉
pmartinez@frba.utn.edu.ar  0000-0002-5677-2021

²Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉
arey@frba.utn.edu.ar  0000-0002-9185-1382

³Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉
walter@frba.utn.edu.ar  0000-0002-6949-0728

Keywords: Sudden Death, Ordinal Patterns, Ordinal Patterns Transitions, Stochastic Matrices, Manhattan's Distance

1 Introducción

El SMS es un problema que está demandando un gran esfuerzo en la comunidad científica internacional. No se dispone hasta el momento de una definición universalmente aceptada de las causas de dicho síndrome, ni de los mecanismos de diagnóstico más apropiados para realizarlo. Jóvenes deportistas con controles médicos absolutamente normales, sin preaviso, sufren de un arresto cardíaco severo y final. En otros casos, adultos mayores con vidas completamente bajo control médico sufren una muerte imprevista. Estas y muchas condiciones sin explicación motivan el desarrollo intensivo de indicadores que permitan, de alguna forma, prever la aparición del SMS. Los métodos de estudio de series temporales basados en patrones ordinales (PO) [1] buscan cuantificar características de una señal midiendo la distribución secuenciada de un subconjunto de valores de misma. En general, se estudia el orden de los patrones a partir de la posición relativa de los valores observados en subsegmentos de datos de longitud fija. De esta manera, se diferencian de los métodos tradicionales de análisis de series temporales no lineales, que en términos generales calculan parámetros en el espacio de fases reconstruido [2], o en aquellos que se obtienen basados en la dinámica simbólica. A partir del trabajo seminal de Bandt & Pompe [3], gran parte de la literatura relativa al análisis ordinal aplicado se ha centrado en la investigación de señales medidas en sistemas biológicos complejos, economía, y en diversidad de aplicaciones (*cf.* [1, 4]).

Las contribuciones más recientes en este campo incluyen el análisis ordinal de ECG e intervalos entre latidos [5–7], y el análisis del electroencefalograma (EEG) para la detección del inicio de convulsiones, la estadificación del sueño y la anestesiología [6, 8]. La amplia diversidad de aplicaciones del método de PO se basa en que no es necesario suponer ninguna característica del sistema que produce la serie temporal (técnicamente se dice que es agnóstica al dominio) [9]. Hay dos aspectos para destacar en la teoría de los PO. Por un lado, bajo ciertas suposiciones existe una relación entre el espacio de fases y la dinámica simbólica de los patrones ordinales [10, 11]. Por otro lado, se ha demostrado con respecto a los mapas discretos, que la distribución de símbolos ordinales enumerados a partir de una serie temporal se rige por las propiedades dinámicas del sistema subyacente [12]. Entre las ventajas de esta metodología se puede mencionar: la simplicidad conceptual y la sencillez de implementación computacional, la velocidad, la robustez y desde un punto de vista más teórico la invariabilidad con respecto a las transformaciones monótonas no lineales de los datos originales [3]. La gran mayoría de la literatura esencialmente se desarrolla en torno a cuantificar algún aspecto de la distribución estacionaria de los PO. Si bien es cierto que cada PO contiene información sobre la estructura temporal de corto alcance de las observaciones contiguas de una serie temporal [1], las cuestiones relativas a la relación entre la secuencia de patrones codificados en la dinámica simbólica ordinal y el comportamiento característico del sistema subyacente siguen siendo en gran medida inexploradas. Sin embargo, recientemente se ha desarrollado una nueva medida denominada la entropía condicional de los PO que fue introducida por Unakafov *et al.* para cuantificar la complejidad transitoria de la dinámica simbólica ordinal mediante el cálculo de la entropía para patrones sucesivos [13]. Recientemente, en forma adicional, se ha propuesto incorporar el concepto de redes ordinales [14] que ha sido complementado con el de redes complejas [15, 16], que a su vez permite cuantificar aspectos de la estructura temporal de la dinámica simbólica ordinal utilizando medidas sencillas del campo de la teoría de las redes. Dado que los PO se calculan sobre segmentos de longitud fija de las señales y éstas se codifican mediante un alfabeto preestablecido, se puede obtener la frecuencia con la que un segmento es seguido por otro de manera tal de recorrer toda la longitud permitida de la serie de datos. A su vez, estas frecuencias relativas se pueden emplear para aproximar la probabilidad de que un segmento dado se comunique con otro presente en la señal. Esto constituye lo que se ha denominado probabilidad de transición entre PO. Con todas las probabilidades de transición estimadas, se puede construir una matriz cuyas columnas contengan la probabilidad de transición de un PO dado hacia todos los otros posibles. La matriz así conformada será una matriz de probabilidad dado que la suma de los elementos de sus columnas dará como resultado 1 (para más detalles *cf.* [9]). Con los ECGs de las bases de datos de registros normales y que han padecido el SMS se construyeron las matrices de transición entre PO (MTPO) para luego calcular una distancia matricial de cada una de ellas a la matriz del valor medio de cada clase de estudio. Los valores de distancia obtenidos han permitido diferenciar claramente ambos grupos de interés.

2 Materiales y Métodos

Las series temporales de ECGs fueron tomadas de datos disponibles en el portal PhysioNet (<https://physionet.org/>), que es de uso muy difundido en la comunidad científica. Se dispuso de 18 ECGs sinusales normales que no presentaban arritmias significativas y 23 registros de ECGs de larga duración pertenecientes a pacientes con diagnóstico de SMS [17].

Para reducir la cantidad de variables que pueden influir en los análisis, en el presente estudio se extrajo de la base de SMS sólo registros correspondientes a personas jóvenes que no tuvieran antecedentes de enfermedades cardiovasculares preexistentes, así como que tampoco estuvieran tomando ningún tipo de medicación. De esta forma, el conjunto de trabajo para los registros diagnosticados con SMS abarcó aquellos identificados, según la base original en PhysioNet, con los números: 30, 33, 42, 43 y 47. Para minimizar el efecto del ruido o la información no relevante, los primeros diez minutos de cada ECG de ambos grupos de pacientes se descartaron. Por la misma razón, el grupo de señales de SMS se filtró previamente con un filtro pasa banda Butterworth de segundo orden, cuya banda de paso se fijó en el rango $[0.5, 4]$ Hz. Para obtener una transformación a PO, se emplea la secuencia que surge naturalmente de la serie temporal, comparando los valores consecutivos en una dada vecindad empleando para ello una ventana deslizante y sustituyéndolos por una secuencia de símbolos, denominado alfabeto. Este proceso requiere de la definición de dos parámetros, la dimensión de incrustación D , que determina el tamaño de la ventana deslizante, y el tiempo de retardo de incrustación τ , que fija el intervalo entre los puntos de serie o señal. En el presente estudio, la dimensión de incrustación se fijó en 4, atendiendo a un compromiso entre el costo computacional y una estimación estadística razonable de los PO a partir de la longitud que poseen las señales de ambos grupos de interés. Esto conduce a que las MTPO resulten de 24 filas y 24 columnas; es decir, de tamaño $D! \times D!$. El tiempo de retardo de incrustación se tomó igual a 1. Se calcularon todas las MTPO correspondientes a los registros de los ECGs de una sola derivación (es decir un solo canal) de ambas bases de datos, junto con el valor medio de las MTPO para cada clase. Finalmente se obtuvo la distancia de Manhattan, definida como el máximo de los valores absolutos de las diferencias entre posiciones, entre las MTPO de cada ECG y las matrices de los valores medios de cada grupo.

3 Resultados

Los grafos que se muestran en la Figura 1 exhiben una clara diferencia entre los valores medios de las MTPO correspondientes a ambas clases. En este sentido se encuentran transiciones con valores más altos en los nodos centrales de grafo de los registros de SMS. Con el fin de cuantificar la diferencia entre las muestras de ambas clases de interés, se propuso el cálculo de la distancia de las MTPO con referencia a los valores medios representados en la Figura 1. Los resultados que se presentan en la Tabla 1 muestran claramente que las señales s30, s33, s42, s43 y s47 de individuos que padecieron el SMS, poseen una distancia de Manhattan a MTPO de valores medios ECGs normales (tercera columna) mucho mayor que a la MTPO de valores medios de la otra clase (segunda columna), lo que induce a incluir dichos ECGs, por proximidad, dentro de la clase de SMS. El caso contrario, procediendo análogamente con los valores de señal del segundo bloque de la tabla, los valores que se aprecian en la tercera columna son extremadamente menores que los correspondientes a la segunda columna, lo que muestra que tales señales se pueden asociar a la clase de ECGs que provienen de sujetos que no han padecido el SMS, o sea con registros normales.

4 Discusión

El presente estudio muestra el avance de resultados que pueden contribuir en la detección temprana del SMS. Por un lado, se emplea una metodología que no necesita de suposición alguna sobre el régimen dinámico del sistema que genera la señal, como los son los patrones ordinales. Sumado a esto último, las redes de transición entre patrones ordinales se encuentran entre los desarrollos más recientes en el análisis de señales (para una lista extensa por ejemplo acudir a [7, 9] y referencias internas). En la selección de los casos empleados con diagnóstico de SMS, el hecho de sólo haber tomado los registros de individuos jóvenes, no medicados y sin diagnóstico de enfermedad cardiovascular alguna, redujo la gama de variación de los factores que pueden influir en el síndrome bajo estudio. La diferencia entre cada MTPO proveniente de un ECG con diagnóstico de SMS resultó ser 75.08% diferente respecto a la distancia de Manhattan de la MTPO del valor medio de todos los registros ECGs con registros normales, mientras

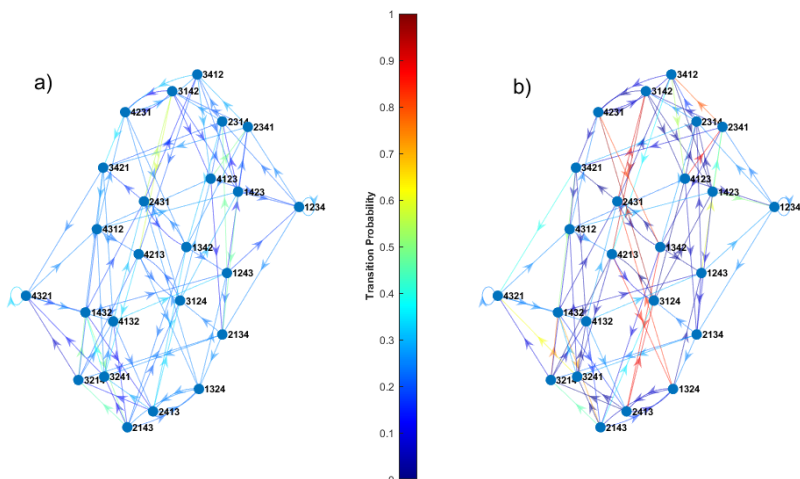


Figura 1: Grafos de valores medios de transición entre PO: a) Normales, b) Afectados por SMS. La intensidad de los colores es proporcional a la frecuencia de las correspondientes transiciones.

que la diferencia en la distancia de Manhattan de las MTPO de registros de ECGs normales resultó ser un 78.39% mayor entre las MTPO de los registros normales a la MTPO del valor medio de los registros con diagnóstico de SMS. Si bien es imprescindible extender la metodología presentada en este trabajo a un número más representativo de casos, el mismo deviene en una prueba de clase piloto que alienta a continuar con esta línea de investigación.

Agradecimientos

Los autores desean expresar su agradecimiento al Dr. Osvaldo Rosso por proponer la metodología de patrones ordinales para realizar los análisis. También destacar el interés y los intercambios llevados a cabo con el desaparecido Prof. Dr. Karsten Keller sobre PO y SMS. Este trabajo se desarrolló en el marco del proyecto PID UTN 8120.

Referencias

- [1] J. M. Amigó, K. Keller, and V. A. Unakafova, "Ordinal symbolic analysis and its application to biomedical recordings," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 373, no. 2034, p. 20140091, 2015. DOI:10.1098/rsta.2014.0091.
- [2] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 9, p. 097610, 2015. DOI:10.1063/1.4917289.
- [3] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002. DOI:10.1103/PhysRevLett.88.174102.
- [4] M. Zanin, L. Zunino, O. A. Rosso, and D. Papo, "Permutation entropy and its main biomedical and econophysics applications: A review," *Entropy*, vol. 14, no. 8, pp. 1553–1577, 2012. DOI:10.3390/e14081553.
- [5] D. Makowiec, A. Kaczkowska, D. Wejer, M. Żarczyńska-Buchowiecka, and Z. R. Struzik, "Entropic measures of complexity of short-term dynamics of nocturnal heartbeats in an aging population," *Entropy*, vol. 17, no. 3, pp. 1253–1272, 2015. DOI:10.3390/e17031253.
- [6] L. Zunino, F. Olivares, and O. A. Rosso, "Permutation min-entropy: An improved quantifier for unveiling subtle temporal correlations," *EPL (Europhysics Letters)*, vol. 109, no. 1, p. 10005, 2015. DOI:10.1209/0295-5075/109/10005.

Tabla 1: Distancias de Manhattan entre la MTPO de cada señal de estudio y las MTPO promedio correspondientes a cada clase.

Señal	Distancia de Manhattan			Porcentaje
	MTPO Media Muerte Súbita	MTPO Media ECG Normal		
s30	3.57	24.92		85.65
s33	13.38	22.44		40.37
s42	4.04	25.06		83.88
s43	4.67	24.50		80.93
s47	3.78	24.60		84.60
s16265	20.87	3.64		82.56
s16272	24.47	3.07		87.45
s16273	21.13	2.84		86.55
s16420	22.05	2.17		90.14
s16483	18.67	4.67		74.98
s16539	22.51	1.89		91.60
s16773	22.06	3.53		83.99
s16786	20.53	3.21		84.37
s16795	22.37	3.18		85.78
s17052	25.19	4.06		83.86
s17453	24.42	3.94		83.84
s18177	20.56	3.94		80.81
s18184	21.07	2.56		87.83
s19088	20.29	3.41		83.18
s19090	20.57	5.22		74.61
s19093	20.22	3.89		80.75
s19140	27.30	9.23		66.18
s19830	21.40	2.90		86.44

- [7] C. W. Kulp, J. M. Chobot, H. R. Freitas, and G. D. Sprechini, “Using ordinal partition transition networks to analyze ECG data,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 7, p. 073114, 2016. DOI:10.1063/1.4959537.
- [8] K. Keller, A. M. Unakafov, and V. A. Unakafova, “Ordinal patterns, entropy, and EEG,” *Entropy*, vol. 16, no. 12, pp. 6212–6239, 2014. DOI:10.3390/e16126212.
- [9] I. Cardoso-Pereira, J. B. Borges, P. H. Barros, A. F. Loureiro, O. A. Rosso, and H. S. Ramos, “Leveraging the self-transition probability of ordinal patterns transition network for transportation mode identification based on GPS data,” *Nonlinear Dynamics*, vol. 107, no. 1, pp. 889–908, 2022. DOI:10.1007/s11071-021-07059-x.
- [10] K. Keller and M. Sinn, “Kolmogorov–Sinai entropy from the ordinal viewpoint,” *Physica D: Nonlinear Phenomena*, vol. 239, no. 12, pp. 997–1000, 2010. DOI:10.1016/j.physd.2010.02.006.
- [11] J. M. Amigó and K. Keller, “Permutation entropy: One concept, two approaches,” *The European Physical Journal Special Topics*, vol. 222, no. 2, pp. 263–273, 2013. DOI:10.1140/epjst/e2013-01840-1.
- [12] J. M. Amigó, S. Zambrano, and M. A. Sanjuán, “True and false forbidden patterns in deterministic and random dynamics,” *EPL (Europhysics Letters)*, vol. 79, no. 5, p. 50001, 2007. DOI:10.1209/0295-5075/79/50001.
- [13] A. M. Unakafov and K. Keller, “Conditional entropy of ordinal patterns,” *Physica D: Nonlinear Phenomena*, vol. 269, pp. 94–102, 2014. DOI:10.1016/j.physd.2013.11.015.
- [14] M. McCullough, M. Small, T. Stemler, and H. H.-C. Iu, “Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 5, p. 053101, 2015. DOI:doi.org/10.1063/1.4919075.

- [15] R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths, “Recurrence-based time series analysis by means of complex network methods,” *International Journal of Bifurcation and Chaos*, vol. 21, no. 04, pp. 1019–1046, 2011. DOI:10.1142/S0218127411029021.
- [16] A. Garcíá Cantú Ros, G. Forti, and G. Nicolis, “Network representation of dynamical systems: Connectivity patterns, information and predictability,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 8, pp. 2018–2034, 2013. DOI:10.1016/j.cnsns.2012.11.007.
- [17] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000. DOI:10.1161/01.CIR.101.23.e215.

Extracción de Características de Imagen para Recuperación de formas 3D a partir de Múltiples Vistas

Image Feature Extraction for Retrieving 3D Shapes from Multiple Views

Marcela Guerrero¹ Juan M. Santos² Juliana Gambini³

Resumen

La reconstrucción 3D a partir de imágenes 2D es un desafío en el campo de imágenes y visión, con múltiples aplicaciones en áreas muy diversas. Algunos métodos se basan en marcadores, los cuales son puntos fluorescentes, estratégicamente ubicados en el objeto de interés. A diferencia de estos métodos, este trabajo se enfoca en encontrar características en imágenes 2D las cuales pueden ser utilizadas como marcadores, permitiendo la reconstrucción automática 3D. Utilizamos el método SIFT (Scale Invariant Feature Transform) para asociar puntos característicos en imágenes de la misma escena provenientes de diferentes puntos de vista. Nos encontramos mejorando el proceso por medio del reconocimiento de *skeletons*, debido a que los puntos relevantes del *skeletons* del cuerpo humano son las articulaciones y las extremidades. El objetivo de este trabajo es que los puntos encontrados en imágenes con diferente punto de vista se utilicen para estimar la coordenada z y luego la estructura 3D. Los resultados obtenidos hasta el momento son alentadores.

Palabras clave: Recuperación 3D, Recuperación de Pose, método SIFT


Abstract

D reconstruction from 2D images is a challenge in the field of images and vision, with multiple applications in very diverse areas. Some methods are based on markers, which are fluorescent dots strategically located on the object of interest. On the other side of these methods, this work focuses on finding features in 2D images which can be used as markers, allowing automatic 3D reconstruction. We use the SIFT method (Scale Invariant Feature Transform) to associate feature points in images of the same scene from different viewpoints. We are improving the process through the recognition of skeletons since the relevant points of the skeletons of the human body are the joints and limbs. The aim of this work is that the points found in images with different viewpoints are used to estimate the z -coordinate and then the 3D structure. The results obtained are so far encouraging.

Keywords: 3D Recuperation, Position Recuperation, SIFT method

¹Depto. de Informática, Instituto Tecnológico de Buenos Aires, Argentina. ✉ mguerrero@itba.edu.ar

²Centro de Investigación y Desarrollo en Informática Aplicada, Universidad Nacional de Hurlingham, Argentina. ✉ juan.santos@unahur.edu.ar

³Dpto. de Ingeniería Informática, Instituto Tecnológico de Buenos Aires; Dpto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero; Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires; Argentina. ✉ mgambini@itba.edu.ar  0000-0002-4534-1402

1 Introducción

La reconstrucción de objetos 3D a partir de imágenes 2D posee múltiples aplicaciones como la recuperación topográfica, la reconstrucción de huesos para la creación de prótesis, el reconocimiento de pose, la industria del cine y de la robótica, entre otras [1]. Este proceso presenta varias dificultades, como la representación de la forma del objeto, las medidas de similitud entre dos formas distintas, la robustez bajo transformaciones afines y ruido o el alto costo computacional [2]. Uno de los enfoques para abordar este problema está basado en múltiples vistas, el cual recupera el objeto 3D utilizando un conjunto de imágenes capturadas a partir de distintos puntos de vista de la misma escena [1, 3]. Éste es el objetivo del presente artículo.

Algunos programas de captura de movimiento recuperan la información 3D de la escena utilizando marcadores, los cuales se ubican en lugares estratégicos del cuerpo, como por ejemplo las articulaciones, la cabeza, los pies, etcétera. Este enfoque se basa en modelos del cuerpo humano previamente definidos y realizan un seguimiento de las marcas, ubicadas en lugares precisos [2, 4]. Los algoritmos más recientes combinan estos modelos con el proceso de reconstrucción 3D, aplicando técnicas de *deep learning* [5, 6].

Otro tipo de métodos son aquellos que estiman la pose del cuerpo reconociendo características de las imágenes sin utilizar información a priori [7]. Inspirados en ese tipo de metodología, asociamos dos imágenes por medio del reconocimiento de correspondencias en esas características, las cuales pueden ser utilizadas como marcadores. Para esto, aplicamos el método SIFT, el cual nos permite detectar características de las imágenes [8]. Las mismas tienen dos atributos: la ubicación centrada en un pixel y un descriptor. El primero es el centro de un parche en la imagen y el segundo es un vector de números reales de 128 elementos que contiene información sobre el histograma de orientaciones dentro del parche. Luego, los descriptores encontrados en cada imagen se asocian por medio de la distancia Euclídea. Dos descriptores cercanos, corresponden a la misma zona de la imagen.

Otro tipo de método que resulta de nuestro interés son aquellos basados en *skeleton* [?], los cuales encuentran una simplificación del cuerpo humano o esqueleto que se utiliza para capturar movimientos, posturas y pose.

Finalmente, queremos reconstruir el objeto en 3D aplicando los métodos mencionados a partir de imágenes capturadas por distintos puntos de vista dentro de un laboratorio. En este artículo presentamos la idea principal del proyecto.

2 Metodología

Las imágenes fueron tomadas en el Laboratorio de Captura de Movimiento, situado en el Instituto Tecnológico de Buenos Aires. Corresponden a ocho cámaras Flex 3 Optitrack, con una resolución de 0.3Mp y con imágenes de (480×640) px. Las cámaras están ubicadas en trípodes alrededor de un círculo y en el centro del mismo está el objeto de interés. La Figura 1 muestra un esquema del sistema. Las ubicaciones y orientaciones de las cámaras se muestran con flechas y el centro es la ubicación del objetivo. De esta forma, es posible tomar imágenes desde ocho puntos de vista diferentes.

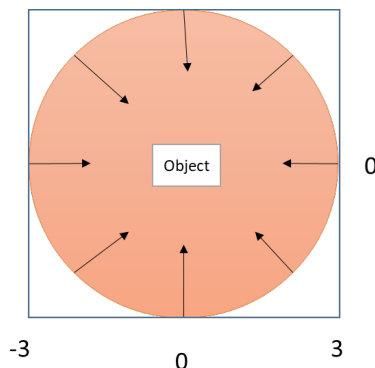


Figura 1: Esquema de posiciones y orientaciones de las cámaras (indicadas con flechas) dentro de la sala del Laboratorio de Captura de Movimiento.

En primer lugar, se calibra el sistema y se calculan los parámetros intrínsecos y extrínsecos de las cámaras. Los

primeros se refieren a las características de la cámara como la distancia focal y la ubicación del punto principal, mientras que los segundos son utilizados para describir la transformación entre la cámara y el sistema de referencia del mundo [9].

El siguiente paso es capturar los movimientos de la persona en la escena, sin utilizar marcadores. La Figura 2 muestra dos imágenes de la misma escena capturadas por dos cámaras consecutivas.



Figura 2: Dos imágenes de la misma escena tomadas por dos cámaras consecutivas.

A partir de esas imágenes, nuestro objetivo es detectar puntos característicos en dos imágenes capturadas con dos cámaras consecutivas y luego las correspondencias entre los mismos. Para esto aplicamos el método SIFT, buscando los puntos cuya distancia euclídea es mínima, como muestra la Figura 3. Debido a que este algoritmo de asociación presenta dificultades para encontrar correspondencias bajo las transformaciones 3D [10], es necesario rotar la segunda imagen para calcular sus puntos característicos y compararlos con la primera. Si bien los resultados hallados son alentadores, observamos que los puntos característicos encontrados no son suficientes para describir la pose o la forma 3D, por ejemplo puede ocurrir que no estén ubicados en todas las extremidades del cuerpo o en las articulaciones, como manos, rodillas, codos, etcétera, los cuales son elementos claves para describir la estructura 3D.

Para resolver este problema, recurrimos al estimador corporal conocido como *skeleton*. La ventaja de utilizar *skeleton* es que puede restringirse la búsqueda de puntos asociados entre aquellos que se encuentren en regiones correspondientes, además de reducir el costo computacional. El método utilizado para hallar el *skeleton*, consiste en encontrar el contorno del cuerpo humano utilizando el método presentado en [11], para luego detectar las extremidades y posteriormente utilizar las métricas corporales para detectar partes del cuerpo, por ejemplo, si h es la altura del sujeto, la *cabeza* mide $\frac{h}{8}$ y el cuello tiene una longitud $0.37 * \text{cabeza}$. De esta forma, los *skeletons* permiten buscar puntos correspondientes entre dos imágenes, utilizando como referencia los puntos principales de ellos en cada una de las imágenes.

El Algoritmo 1 describe el pseudocódigo del método utilizado. Sean I_j , C_j y LI_{C_j} , $j = 1, \dots, 8$ las imágenes capturadas, el contorno del objeto de interés de cada una y la lista de píxeles en el interior del contorno, respectivamente.

3 Resultados

La Figura 3 muestra los resultados que se generan al aplicar el método SIFT a las imágenes de la Figura 2. Puede observarse que los puntos característicos encontrados no están en todas las extremidades o en las articulaciones, como las rodillas, codos o cuello. Esto dificulta la recuperación tanto de la estructura 3D como de la pose o de los movimientos.

La Figura 4 muestra el *skeleton* encontrado en una imagen. Puede observarse que los puntos de unión corresponden a los de flexión del cuerpo humano. Construimos los descriptores con el método SIFT para estos píxeles, solucionando así el problema anterior.

En el paso 8 del Algoritmo 1, se obtiene la profundidad del punto en la escena a partir de los píxeles correspondientes y de los parámetros extrínsecos e intrínsecos del sistema, utilizando triangulación. En el paso 12, el

Algorithm 1 Algoritmo utilizando *skeletons*.

Dado C_1 y las métricas corporales, obtener un *skeleton* S_1 de 14 puntos principales.

- 2: Obtener los *skeletons* S_i , $i = 2, \dots, 8$ de acuerdo a los parámetros extrínsecos calculados en la calibración.
Sobreimprimir los *skeletons* S_i sobre las imágenes I_i , $i = 1, \dots, 8$.
- 4: Inicializar la lista L vacía.
for $i = 1$ **to** 8 **do**
6: **for** $p_i \in LI_{C_i}$ **do**
 Obtener su descriptor d_{p_i} usando SIFT y encontrar la correspondencia p_{i+1} en $LI_{C_{i+1}}$, usando como referencia los puntos principales de los *skeletons* S_i y S_{i+1} .
- 8: Obtener las coordenadas (x, y, z) respecto al centro de referencia del sistema.
 Agregar (x, y, z) a L .
- 10: **end for**
 end for
- 12: **return** L

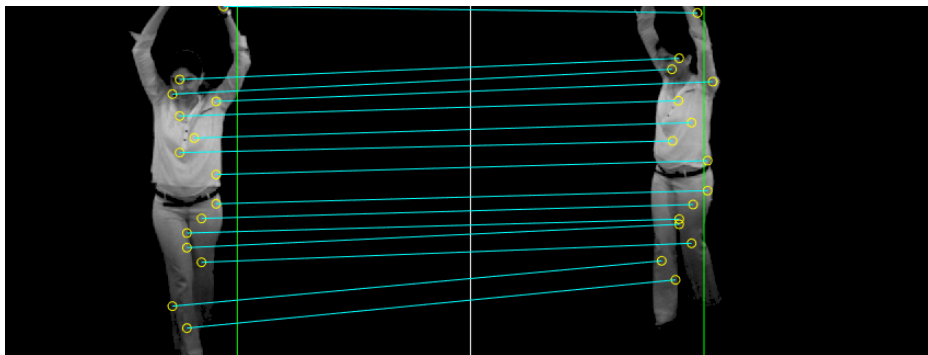


Figura 3: Correspondencias entre dos imágenes provenientes de distintos puntos de vista.

algoritmo devuelve una lista de puntos (x, y, z) con la cual puede calcularse un modelo 3D del sujeto usando, por ejemplo, triángulos o trapecios. Esta operación debe realizarse para cada par de cámaras consecutivas en el sistema, generando una lista de puntos 3D. Luego la triangulación debe ser aplicada para obtener la representación 3D del cuerpo de la persona.

4 Conclusiones

A partir de las pruebas realizadas, es posible encontrar correspondencias entre dos imágenes del cuerpo humano, tomadas desde distintos puntos de vista. Los píxeles que pueden ser asociados, son utilizados como marcadores para la reconstrucción 3D. El problema es que los puntos de asociación deben ubicarse estratégicamente para poder recomponer la figura tridimensional. Con este objetivo, estudiamos el uso de *skeleton*, búsqueda de descriptores sobre los puntos relevantes de *skeleton*, el método para encontrar la coordenada Z utilizando dos vistas del mismo objeto y el algoritmo de triangulación para reconstruir su superficie. Sabemos que tenemos trabajo por delante pero los resultados obtenidos hasta el momento son prometedores. La nueva etapa de trabajo que estamos encarando ahora, incluye agregar cámaras al círculo de manera que el ángulo entre las cámaras se menor.

Referencias

- [1] A. Liu, N. Hu, D. Song, S. Guo, H. Zhou, and T. Hao, “Multi-view hierarchical fusion network for 3D object retrieval and classification,” *IEEE Access*, vol. 7, pp. 153021–153030, 2019. DOI:10.1109/ACCESS.2019.2947245.
- [2] A. Bronstein, M. Bronstein, L. Guibas, and M. Ovsjaniko, “Shape google: Geometric words and ex-

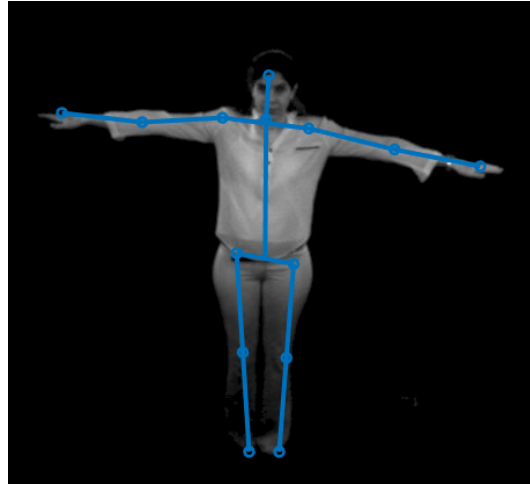


Figura 4: *Skeleton* del cuerpo humano. Pueden identificarse los 14 puntos principales.

pressions for invariant shape retrieval,” *ACM Transactions on Graphics*, vol. 30, no. 1, pp. 11–20, 2011. DOI:10.1145/1899404.1899405.

- [3] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. DOI:10.1023/B:VISI.0000029664.99615.94.
- [4] D. Zhang, Z. Miao, and S. Chen, “Human model adaptation for multiview markerless motion captures,” *Mathematical Problems in Engineering*, vol. 2013, no. 1, pp. 1–7, 2013. DOI:10.1155/2013/564214.
- [5] H. Yasin, U. Iqbal, B. Krugerand, A. Weber, and J. Gall, “A dual-source approach for 3D pose estimation from a single image,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4948–4956, 2016. DOI:10.1109/CVPR.2016.535.
- [6] A. Mathis, S. Schneider, J. Lauer, and M. Mathis, “A primer on motion capture with deep learning: Principles, pitfalls, and perspectives,” *Neuron*, vol. 108, no. 1, pp. 44–65, 2020. DOI:10.1016/j.neuron.2020.09.017.
- [7] J. Park, K. Park, S. Baeg, and M. Baeg, “Reliable feature point detection and object pose estimation using photometric quasi-invariant SIFT,” in *International Conference on Control, Automation and Systems*, pp. 2142–2147, 2008. DOI:.
- [8] M. Guerrero, J. Santos, and J. Gambini, “3D features recovery from images of a multi camera system,” in *International Conference on Image Processing, Computer Vision, & Pattern Recognition (IPC)*, pp. 1–5, 2021.
- [9] Z. Zhang, *Camera Parameters (Intrinsic, Extrinsic)*. Boston, MA: Springer US, 2014. DOI:10.1007/978-0-387-31439-6_152.
- [10] J. M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” *SIAM Journal on Imaging Science*, vol. 2, no. 2, pp. 438–469, 2009. DOI:10.1137/080732730.
- [11] J. Gambini, D. Rozichner, M. E. Buemi, M. Mejail, and J. Berllés, “Occlusion handling for object tracking using a fast level set method,” in *XXI Brazilian Symposium on Computer Graphics and Image Processing*, pp. 61–68, 2008. DOI:10.1109/SIBGRAPI.2008.36.

Detección de Bordos en Imágenes Satelitales Nocturnas usando Grafos y Medidas de Información

Edge Detection in Satellite Images at Night using Graphs and Information Measures

Manuel García Blesa¹ Gisela Caballero² Claudio Delrieux³ Andrea Rey⁴

Resumen

El sensado remoto es una técnica para la adquisición de información espacial, temporal y espectral que no requiere del contacto físico con el objetivo de interés. Dependiendo de la plataforma donde se ubique el sensor, los sistemas de teledetección pueden ser terrestres, aéreos o espaciales. En este último caso es que se encuentran los satélites que orbitan alrededor de la Tierra. Las imágenes satelitales proporcionan una gran fuente de información, cuyo análisis es una herramienta fundamental para el estudio del medio ambiente y la actividad humana. En el caso particular de las imágenes satelitales nocturnas, resulta de interés poder estudiar las zonas urbanas -representadas por valores altos de brillo- con el objeto de analizar áreas de ocupación, crecimiento demográfico o planificación urbana. En el presente trabajo se introduce un algoritmo automático para la detección de bordes de zonas relativas a ciudades aplicado a imágenes satelitales nocturnas. La metodología se basa en calcular una nube de puntos en el espacio causal entropía-complejidad a partir del grafo asociado a una ventana que se desliza a través de la región de interés. Luego, este conjunto de observaciones es particionado mediante la técnica de k -medias lo que origina una imagen binarizada a la cual se le aplica un código sencillo para obtener los puntos en el borde buscado. El algoritmo propuesto se aplica sobre dos imágenes satelitales nocturnas de regiones en Argentina y en Estados Unidos, con resultados que muestran un buen desempeño.


Palabras clave: Imágenes satelitales nocturnas, Detección de bordes, Entropía, Complejidad, Grafos


Abstract

Remote sensing is a technique for the acquisition of spatial, temporal, and spectral information that does not require physical contact with the target of interest. Depending on the platform where the sensor is located, remote sensing systems can be land, air, or space. In this last case, the satellites that orbit around the Earth are found. Satellite images provide a great source of information, whose analysis is a fundamental tool for the study of the environment and human activity. In the particular case of night satellite images, it is of interest to be able to study

¹Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉ hgarciablesa@frba.utn.edu.ar  0000-0003-0491-2018

²Dpto. de Matemática, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉ gcaballero@frba.utn.edu.ar

³Dpto. de Ingeniería Eléctrica y de Computación, Universidad Nacional del Sur y CONICET, Argentina. ✉ cad@uns.edu.ar  0000-0002-2727-8374

⁴Centro de Procesamiento de Señales e Imágenes, Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Argentina. ✉ arey@frba.utn.edu.ar  0000-0002-9185-1382

urban areas -represented by high brightness values- in order to analyze areas of occupation, growing demographic, or urban planning. In the present work, an automatic algorithm applied to night satellite images is introduced for the detection of edges of areas relative to cities. The methodology is based on finding a cloud of points in the entropy-complexity causal plane from the graph associated with a sliding window through the region of interest. Then, this set of observations is partitioned using the k -means technique, which gives rise to a binarized image at which a simple programming code is applied to obtain the points on the sought edge. The proposed algorithm is applied to a pair of night satellite images of regions in Argentina and the United States, and the results show good performance.

Keywords: Nigth Satellite Images, Edge Detection, Entropy, Complexity, Graphs

1 Introducción

Las técnicas de sensado remoto se caracterizan por realizar mediciones de propiedades físicas mediante un dispositivo que no entra en contacto con el objeto, área o fenómeno que está siendo estudiado. La tecnología satelital es un claro ejemplo donde se emplean sensores a bordo de plataformas espaciales, diseñados para estudiar energía reflejada y emitida desde la Tierra [1]. Las imágenes satelitales son el resultado de este proceso de teledetección y aparecen como una de las mayores fuentes de información para el inventario territorial y el monitoreo de recursos naturales.

Una imagen satelital está compuesta por una gran cantidad de puntos llamados píxeles. Cada píxel tiene un color propio, en general dentro de una paleta de tonos de gris, incluyendo los casos extremos representados por los colores blanco y negro. El tono asignado a un píxel representa el nivel de energía -valor de brillo- sensado por el satélite. Típicamente se trabaja con 256 valores de brillo que ayudan al reconocimiento de características de la imagen. La diferencia entre los tonos de gris mide el grado de contraste entre un objeto y su fondo, siendo más fácil la identificación ante la presencia de mayor contraste. Este tipo de imágenes ha sido muy utilizado a lo largo de las últimas décadas con el propósito de obtener información en una gran variedad de campos del conocimiento como aplicaciones militares [2, 3], estudio de áreas de cultivo [4], detección de desastres naturales [5], seguimiento de patrones climáticos globales [6], entre otros. La Teoría de Grafos ha sido ampliamente aplicada a tareas de procesamiento y análisis de imágenes. A modo de ejemplo, se pueden citar los trabajos [7] y [8] que realizan segmentación y clasificación, respectivamente. Un grafo consiste en un conjunto no vacío de vértices y un conjunto de aristas cada una de las cuales unen a dos vértices. Dada una imagen, se puede construir un grafo asociado a la misma de manera tal que por cada píxel en la imagen se define un vértice y cada arista representa la conexión entre píxeles pertenecientes a una vecindad dada. Con el objetivo de describir ciertas características de la imagen, se suele asignar un peso a cada arista.

La entropía es una manera de medir el número de formas distintas de arreglar un conjunto de objetos, considerándose como una medida del grado de equilibrio de un sistema o su tendencia al desorden. Por otro lado, la complejidad se puede definir como la capacidad que tiene un sistema de incorporar información, reflejando la cantidad necesaria de información no aleatoria para describir al sistema. Para más detalles sobre estos conceptos, nos referimos al trabajo [9]. En el presente trabajo se propone aplicar herramientas de la Teoría de la Información al grafo asociado a una imagen satelital nocturna con el fin de detectar el borde de zonas urbanas.

2 Materiales y Métodos

Sea I una imagen de tamaño $M \times N$, se construye el grafo asociado $\mathcal{G}(I)$ de la siguiente manera. Dados P y V los conjuntos de píxeles y de vértices respectivamente, se define la biyección $f : P \rightarrow V$ como $f(x, y) = (x - 1)M + y$ (cf. [10]). Luego, el conjunto de aristas queda definido como $E = \cup_{(x,y) \in P} \{e_{f(x,y)f(x',y')} : (x', y') \in \mathcal{N}_{(x,y)}\}$, donde $\mathcal{N}_{(x,y)}$ es el conjunto formado por los ocho vecinos del píxel (x, y) . Finalmente, se usa un *kernel* radial para asignar a cada arista $e_{ij} \in E$ el peso $w_{ij} = \exp(-|I_{f^{-1}(i)} - I_{f^{-1}(j)}|/0.01)$, donde I_a denota el valor de brillo de la imagen I en el píxel a . Observar que este peso resulta mayor en el caso de valores de brillo muy similares, permitiendo la desconexión de vértices que corresponden a píxeles con diferencias marcadas en su brillo.

Dado un grafo \mathcal{G} con n vértices, se define la matriz de adyacencia $A_{\mathcal{G}} = (w_{ij}) \in \mathbb{R}^{n \times n}$ donde w_{ij} es el peso de la arista que une los vértices i y j . Además, el peso total de \mathcal{G} es $W_{\mathcal{G}} = \sum_{i,j=1}^n w_{ij}$. Considerando la proporción de cada componente de $A_{\mathcal{G}}$ en $W_{\mathcal{G}}$, se construye una función de densidad de probabilidad (fdp) discreta $\mathcal{P}(A_{\mathcal{G}}) = \{p_{ij} = w_{ij}/W_{\mathcal{G}} : i, j = 1, 2, \dots, n\}$.

A partir de la fdp $\mathcal{P}(A_G)$, se calcula la entropía de Shannon definida como:

$$S[\mathcal{P}(A_G)] = - \sum_{i,j}^n p_{ij} \ln(p_{ij}). \quad (5)$$

En este trabajo se considera una versión proporcional a la entropía de Shannon dada por:

$$H^*[\mathcal{P}(A_G)] = S[\mathcal{P}(A_G)]/[n^2 \ln(n-1)]. \quad (6)$$

Para el cálculo de la complejidad, se toma el enfoque propuesto en [11], donde se define:

$$C[\mathcal{P}(A_G)] = \frac{-2}{(1+1/n) \ln(n+1) - 2 \ln(2n) + \ln(n)} \cdot \left[- \sum_{i,j=1}^n \frac{q_{ij}}{2} \ln\left(\frac{q_{ij}}{2}\right) + \frac{1}{2} \sum_{i,j=1}^n p_{ij} \ln(p_{ij}) - \frac{1}{2} \ln(n) \right] \cdot \frac{S[\mathcal{P}(A_G)]}{\ln(n^2)}, \quad (7)$$

donde $q_{ij} = p_{ij} + 1/n$ y el factor del medio corresponde a la divergencia de Jensen entre $\mathcal{P}(A_G)$ y la distribución equiprobable.

La propuesta presentada en este trabajo para la detección automática de bordes en una imagen satelital nocturna, se basa en el siguiente algoritmo:

1. Dada una imagen satelital nocturna, se toma R una región de interés de la misma.
2. Se elige un número natural impar s . En el presente trabajo se toma $s = 5$.
3. Para cada ventana deslizante U de tamaño $s \times s$:
 - (a) Se construye el grafo asociado $\mathcal{G}(U)$.
 - (b) Se calcula la matriz de adyacencia $A_{\mathcal{G}(U)}$ junto con su fdp $\mathcal{P}(A_{\mathcal{G}(U)})$.
 - (c) Se calcula el punto $(H^*[\mathcal{P}(A_G)], C[\mathcal{P}(A_G)])$ en el plano de información causal $H \times C$, donde la entropía proporcional definida por la expresión (6) se representa en el eje horizontal y la complejidad definida por la expresión (7) en el eje vertical.
4. Se aplica el algoritmo de *clusterización k-medias* con $k = 2$, a los puntos obtenidos en el plano $H \times C$. Este método separa las observaciones en k grupos minimizando la varianza dentro de cada grupo.
5. Se construye una imagen binarizada con el resultado obtenido por k -medias.
6. Se aplica un código de programación automático que fue desarrollado en lenguaje R [12], versión 4.1.0, para obtener los puntos en el borde de las zonas urbanas. (En una próxima etapa, el código estará disponible en un repositorio abierto.)

Con el fin de analizar el desempeño de la metodología propuesta, se trabajó con dos imágenes satelitales nocturnas, una correspondiente a la provincia de Buenos Aires en Argentina, y la otra a los Estados Unidos. Los datos corresponden a la misión *Defense Meteorological Satellite Program* de la *National Oceanic and Atmospheric Administration* (NOAA), y fueron descargados en el año 2005 desde el sitio <https://www.ospo.noaa.gov/Operations/DMSP/index.html>. Estas imágenes, junto con las regiones seleccionadas se muestran en la Figura 1. Los tamaños de cada recorte son 201×201 para Buenos Aires y 141×541 para Estados Unidos.

3 Resultados y Discusión

Las Figuras 2(a) y 3(a) presentan las imágenes binarizadas luego de aplicar el algoritmo de k -medias a los puntos obtenidos en el plano $H \times C$ a partir de las ventanas deslizantes. Estas visualizaciones sugieren resultados aceptables pero que deben ser mejorados. A tal fin, se ejecuta el código mencionado en el ítem 6 de la Sección ??, obteniendo los resultados que se muestran en las Figuras 2(b) y 3(b). Como se puede apreciar, la metodología propuesta es capaz de detectar el borde de las zonas urbanas de una manera muy eficiente. Las corridas se realizaron en una computadora con procesador Intel® Core™, i7-6700K CPU 3.40 GHz 3.41 GHz y sistema operativo con 16 GB RAM y System Type 64 bit. En cuanto al costo computacional, el mayor tiempo insumido se registra al calcular los puntos en el plano $H \times C$. Los tiempos transcurridos para cada paso de la metodología propuesta al aplicarse a ambas imágenes, se encuentran en la Tabla 1.

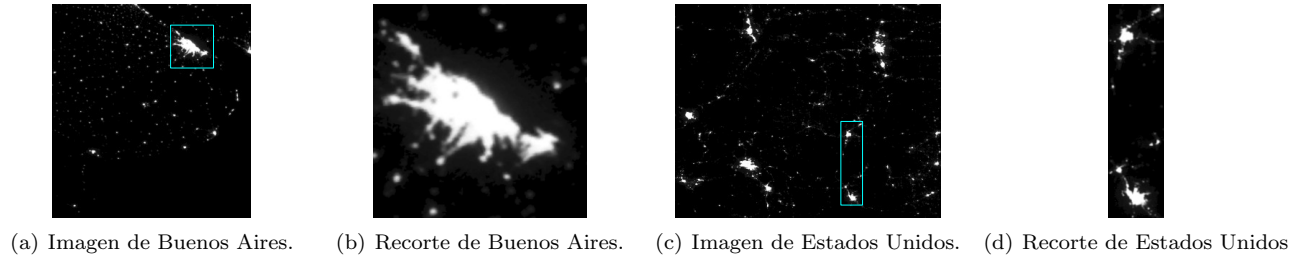


Figura 1: Imágenes satelitales y sus correspondientes recortes que definen las regiones de interés.

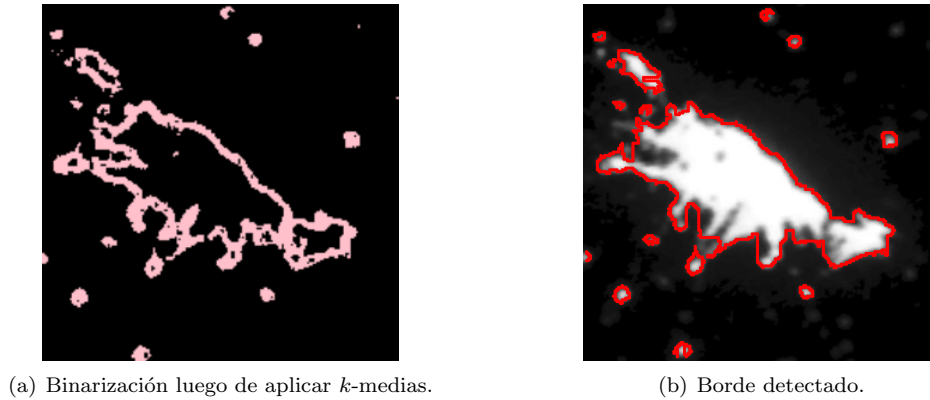


Figura 2: Resultados de la metodología propuesta aplicada al recorte de Buenos Aires.

4 Conclusiones

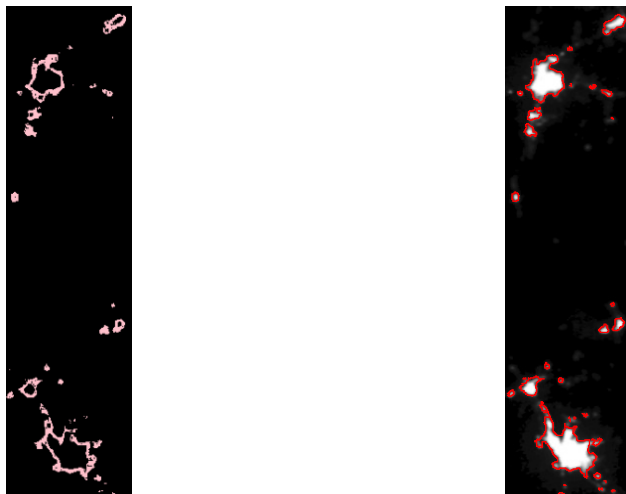
En el presente trabajo se utilizan herramientas de la Teoría de Grafos combinadas con medidas de la información, como lo son una versión proporcional de la entropía de Shannon y la complejidad. La metodología propuesta es aplicable a problemas de detección de bordes en imágenes satelitales nocturnas. La misma ha mostrado resultados interesantes al aproximar contornos de zonas urbanas en dos regiones de interés de distintos puntos geográficos como son los países de Argentina y Estados Unidos. El tiempo de procesamiento no es demasiado alto pero está condicionado por el tamaño de la imagen bajo estudio y por la elección del tamaño de las ventanas deslizantes. La elección de trabajar con ventanas deslizantes de 5×5 proviene del ensayo de otros tamaños que mostraron resultados poco satisfactorios en el caso de 3×3 y de menor calidad que el presentado en el caso de 7×7 , además de un mayor costo computacional para este último. En futuros trabajos, se pretende estudiar el crecimiento demográfico de una región de interés aplicando la presente propuesta a una serie temporal de imágenes satelitales nocturnas. Del mismo modo, se puede aplicar esta propuesta a imágenes aerotransportadas y a imágenes médicas de alto contraste.

Referencias

- [1] E. D. Conway, *An introduction to satellite image interpretation*. JHU Press, 1997.

Tabla 1: Tiempo “real” transcurrido en el sistema por las tareas requeridas en la metodología propuesta.

Recorte	Nro. de píxeles	Tiempo (en segundos) por tarea			
		Puntos en $H \times C$	k -medias	Detección del borde	
Buenos Aires	40401	208.27	0.00	1.75	
Estados Unidos	76281	391.32	0.00	1.60	



(a) Binarización luego de aplicar k -medias.

(b) Borde detectado.

Figura 3: Resultados de la metodología propuesta aplicada al recorte de Estados Unidos.

- [2] H. Bandrupally, H. R. Talusani, and T. Sridevi, "Detection of military targets from satellite images using deep convolutional neural networks," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, pp. 531–535, 2020. DOI:10.1109/ICCCA49541.2020.9250864.
- [3] W. Kartikaningsih, R. H. Triharianto, and D. Gunawan, "Identification of military objects around Indonesia using public domain satellite images," in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 1–7, 2021. DOI:10.1109/ICACSIS53237.2021.9631339.
- [4] S. Skakun, C. O. Justice, N. Kussul, A. Shelestov, and M. Lavreniuk, "Satellite data reveal cropland losses in South-Eastern Ukraine under military conflict," *Frontiers in Earth Science*, p. 305, 2019. DOI:10.3389/feart.2019.00305.
- [5] S. N. K. B. Amit, S. Shiraishi, T. Inoshita, and Y. Aoki, "Analysis of satellite images for disaster detection," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5189–5192, 2016. DOI:10.1109/IGARSS.2016.7730352.
- [6] Y. Zhang, S. Wistar, J. Li, M. A. Steinberg, and J. Z. Wang, "Severe thunderstorm detection by visual learning using satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 1039–1052, 2016.
- [7] S. Zorzi, S. Bazrafkan, S. Habenschuss, and F. Fraundorfer, "Polyworld: Polygonal building extraction with graph neural networks in satellite images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1848–1857, 2022. DOI:10.48550/arXiv.2111.15491.
- [8] J. Kang, R. Fernandez-Beltran, D. Hong, J. Chanussot, and A. Plaza, "Graph relation network: Modeling relations between scenes for multilabel remote-sensing image classification and retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4355–4369, 2020. DOI:10.1109/TGRS.2020.3016020.
- [9] T. Modis, "Links between entropy, complexity, and the technological singularity," *Technological Forecasting and Social Change*, vol. 176, p. 121457, 2022. DOI:10.1016/j.techfore.2021.121457.
- [10] A. Rey, G. Caballero, and A. Arce, "Estudio del crecimiento de una región de interés en una imagen usando medidas de grafos," *Proyecciones*, vol. 19, no. 2, pp. 23–34, 2021.
- [11] M. T. Martin, Á. Plastino, and O. A. Rosso, "Generalized statistical complexity measures: Geometrical and analytical properties," *Physica A: Statistical Mechanics and its Applications*, vol. 369, no. 2, pp. 439–462, 2006. DOI:10.1016/j.physa.2005.11.053.

- [12] R Core Team, “R: A language and environment for statistical computing (R Version 4.1.0, R Foundation for Statistical Computing, Vienna, Austria),” 2022. URL: <https://www.R-project.or>.

C P S I
Centro de Procesamiento
de Señales e Imágenes



UTN.BA
UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES