

Abordando los Desafíos de la Detección de Ciclos en Entornos Agrícolas

Nicolas Soncini

Laboratorio de Robótica
CIFASIS (UNR-CONICET)
Rosario, Santa Fe, Argentina
soncini@cifasis-conicet.gov.ar

Javier Civera

Robotics, Perception and Real-Time Group
Universidad de Zaragoza
Zaragoza, España
jcivera@unizar.es

Taihú Pire

Laboratorio de Robótica
CIFASIS (UNR-CONICET)
Rosario, Santa Fe, Argentina
pire@cifasis-conicet.gov.ar

Resumen—La detección de ciclos es un componente fundamental de los sistemas de localización de los robots autónomos, especialmente en operaciones a largo plazo. Sin embargo escasa literatura puede encontrarse sobre el tema de detección de ciclos en campos agrícolas. En este trabajo abordamos la detección de ciclos en campos agrícolas, donde se espera que el uso de robots autónomos para diversas tareas agrícolas tenga un impacto fundamental en los próximos años, basada únicamente en información visual. En particular mostramos resultados de evaluar métodos de detección de ciclos en entornos agrícolas y proponemos mejoras basadas en información por cámaras estéreo para robustecer los mismos.

I. INTRODUCCIÓN

En robótica móvil, el problema de cierre de ciclos (*Loop Closure*) hace referencia al problema de detectar cuando un robot móvil se encuentra transitando por un área previamente visitada, y utilizar dicha información para ajustar la trayectoria y el mapa estimado. La tarea de cierre de ciclos es un paso fundamental para resolver el problema de SLAM (*Simultaneous Localization and Mapping*), en el cual se aborda la tarea de estimación de pose del robot (localización) y construcción de mapa (mapeo) de manera simultánea [1], [2], [3], y donde el proceso de cierre ciclos permite reducir el error acumulado (*drift*) tanto en la localización del sistema como en la reconstrucción del mapa del entorno.

Es común encontrar asociadas las tareas de cierre de ciclos y reconocimiento visual de lugares (VPR) en la literatura. Los métodos tradicionales de reconocimiento de lugares basados en información visual consisten en extraer características visuales presentes en las imágenes (llamados *features*) y buscar correspondencias de características entre imágenes de diferente momento, soliendo utilizarse técnicas derivadas de *Bag of Words* (BoW) [4], [5], [6], [7], o utilizando redes neuronales con aprendizaje profundo (*Deep Learning*) para encontrar ciclos ya sea mediante características visuales profundas o descriptores globales de cada imagen [8], [9], [10]. Sin embargo no podemos utilizar dichos métodos directamente para la detección de ciclos, ya que a la hora de realizar el cierre de ciclos cualquier detección espúrea puede ser irremediable en la tarea de SLAM, volviendo la localización del sistema, como así también al mapa, inutilizable. Para esto necesitamos de sistemas de detección de ciclos que sean precisos a la hora de encontrar y verificar los ciclos, y cuya estimación sea robusta a

los diversos problemas que pueden encontrarse en el entorno, como son: la repetitividad visual, la falta de características visuales del entorno, los cambios de iluminación, y los cambios temporales.

Por otro lado, podemos encontrar una vasta literatura abocada a resolver el problema de localización visual y cierre de ciclos en situaciones de interiores, urbanas o semi-urbanas. Entre ellas se encuentran City-Scale [11] que muestra resultados positivos a nivel de una ciudad, FAB-MAP [12] que se focaliza en conjuntos de datos urbanos, Cadena *et al.* [13] que muestra precisión en entornos de interiores y exteriores urbanos, NetVLAD [14] que se prueba en conjuntos de imágenes de ciudades tomadas desde la calle, SeqNetVLAD [15] que funciona muy bien en secuencias vehiculares a nivel calle y Garg *et al.* [16] que muestra resultados en entornos interiores y urbanos. Sin embargo es sumamente escasa la literatura respecto a entornos agrícolas o de campo abierto [17], [18], [19], donde existe una alta repetitividad visual y una falta de características visuales discriminativas (ver Fig 1).

En el presente trabajo presentamos avances en la tarea de detección visual de ciclos en entornos agrícolas mediante la modificación de un sistema de detección monocular para hacer uso de información visual estéreo que nos permita aumentar la confianza en sus resultados. En la sección II describimos el dataset agrícola público y abiertamente disponible que utilizaremos para evaluar los métodos de detección de ciclos. Luego en la sección III describimos un método de detección de ciclos que analizaremos y, luego de plantear las problemáticas propias del uso de tal sistema en los entornos agrícolas o de campo abierto, desarrollamos una mejora para el mismo a partir del uso de información estéreo. En la sección IV mostramos y analizamos los resultados al correr dichos métodos, y finalmente en la sección V damos una breve conclusión sobre los resultados obtenidos y planteamos posibles trabajos futuros que se desprenden de los mismos.

II. DATASET

La escasez de datasets de entornos de agricultura y de campo abierto nos lleva a tener que utilizar al máximo los datos que logramos colectar. Hacemos uso del dataset Field-SAFE [20], que hasta donde entendemos es el único dataset

abierto en entornos agrícolas que posee ciclos e información de ground-truth de precisión.

FieldSAFE es un dataset compuesto por un conjunto de sensores montados en un tractor que recorre un campo agrícola en el país de Dinamarca transcurriendo el año 2016. Se concibe con la idea de capturar datos relevantes a la detección de personas y objetos en entornos de agricultura para mejorar la seguridad de los vehículos autónomos en dichos entornos. Nosotros aprovechamos el hecho de que contiene todos los sensores necesarios para testear un sistema de detección de ciclos, dado que el sistema montado en el tractor posee los siguientes sensores:

- GNSS: utilizan un *Trimble BD982 GNSS*¹ que nos provee con una posición altamente precisa del vehículo en el campo, el cual utilizamos para derivar el ground-truth de ciclos.
- 3D Stereo Camera: monta también una cámara estéreo *Multisense S21 CMV2000*² que nos provee con un par de imágenes estéreo sincronizadas y de shutter global.
- Web Camera: posee una cámara *Logitech HD Pro C920*³ que podemos utilizar como alternativa a la cámara estéreo, de bajo precio y más fácil obtención.

El dataset FieldSAFE comprende 5 sesiones de grabación en las cuales el tractor recorre el entorno agrícola. Hacemos uso de 3 de las 5 sesiones en las cuales el tractor hace recorridos más largos y que poseen mayor número de superposiciones en las trayectorias (lo cual buscamos para detectar ciclos en las mismas): “Dynamic obstacle session #1”, “Dynamic obstacle session #2” y “Static obstacle session #2”.

La posición ground-truth se obtiene a partir de los datos de latitud y longitud provistos por el GPS RTK. Evitamos utilizar la orientación que provee este mismo sensor ya que al revisar los datos podemos ver que en los momentos de giro del robot los valores de la orientación tienen cierto retraso respecto a los que se pueden apreciar en las imágenes, y les toma un tiempo apreciable hasta converger a la orientación correcta, probablemente causado por el uso de un filtro sobre los datos crudos del sensor IMU contenido en el módulo GPS.

III. MÉTODO PROPUESTO

Analizamos el funcionamiento del sistema de detección de ciclos monocular presentado por Gálvez-López et.al. [5] en entornos agrícolas haciendo uso de su implementación abierta⁴. El mismo consiste en una detección y descripción de puntos característicos en la imagen, seguido de una búsqueda de imágenes similares mediante un método de “Bag of Words” jerárquico [21], [22]. En caso de encontrar imágenes similares se realiza un chequeo con imágenes temporalmente cercanas a la encontrada y en caso de estar todas de acuerdo se selecciona la que más certeza provee y se realiza un chequeo geométrico que consiste en calcular una matriz fundamental entre los

puntos en común de ambas imágenes, y en el caso positivo se acepta dicha imagen como ciclo.

El problema con este método de estilo clásico reside en que al utilizarlo en entornos de campo abierto pueden detectarse ciclos para poses del robot muy distantes, lo que no necesariamente representa un mal funcionamiento del sistema. Las métricas que se utilizan al definir ciclos en la literatura usualmente conciernen a entornos urbanos o estructurados, donde la detección de ciclos asume que el robot, o la cámara, transitan por caminos predefinidos, usualmente angostos, donde los elementos estructurales se encuentran en cercanía al mismo. Esto ha permitido asumir que la información visual es rápidamente cambiante respecto a la distancia recorrida, por ejemplo al moverse pocos metros en la ciudad podemos notar cómo los edificios cercanos desaparecen y dan lugar a nuevos edificios cercanos al horizonte. Sin embargo al tratar con entornos de terreno abierto o extensiones agrícolas de cultivo son escasos los caminos demarcados o las estructuras cercanas, con lo cual el reconocimiento visual de ciclos suele encontrarse con situaciones visualmente similares pertenecientes a posiciones y orientaciones completamente distintas, problema que conocemos como “perceptual aliasing”. Mostramos un ejemplo en la Fig. 1 donde se toman dos fotogramas de un conjunto de datos urbanos y dos fotogramas de un conjunto de datos agrícolas, a la misma distancia posicional, en los cuales se aprecia que el entorno urbano cambia drásticamente su apariencia visual, mientras que el entorno agrícola presenta únicamente cambios de escala de las estructuras lejanas.

Es claro entonces que no podemos evaluar usando métricas arbitrarias sobre distancia y posición, sino que debemos pasar a evaluar nuestros sistemas de detección en el espacio métrico en el que se encuentran, y que nos provean información de que tan cercana o que tan lejana se encuentra nuestra detección respecto a la correcta. Para esto es necesario contar con una estimación de la pose por parte del sistema de detección de ciclos, con lo cual entre los pasos a seguir deberemos adicionar al sistema VPR una estimación de pose y evaluar a ambos como un solo elemento, además de necesitar de nuevas métricas que nos agreguen y nos permitan analizar el error en estimación de pose que resulta de cada una de las detecciones. Creemos que esto además enrobustecerá al sistema, ya que al realizar un subsecuente paso de estimación de pose podemos eliminar aquellos candidatos a ciclos espurios que resultan de correr un sistema VPR sin estimación de pose final.

Para mitigar este problema realizamos cambios en el sistema de detección de ciclos que mencionamos anteriormente para acomodar la utilización de imágenes estéreo, con la intención de mejorar la calidad de las detecciones y potencialmente estimar la pose relativa entre las cámaras estéreo. Mostramos un diagrama de flujo simplificado en la Fig. 2 donde marcamos con texto en negrita los procesos modificados o re-implementados para permitir el trabajo con imágenes estéreo.

El proceso de emparejamiento estéreo consiste en la búsqueda de características visuales comunes a ambas imágenes según la distancia entre descriptores de los mismos, en este caso dado que las imágenes se encuentran rectificadas pode-

¹<https://oemgnss.trimble.com/product/trimble-bd982/>

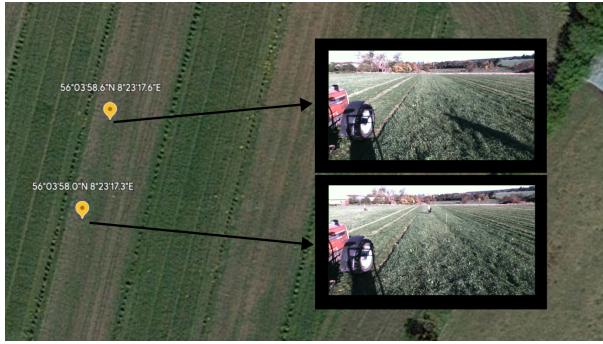
²<https://docs.carnegierobotics.com/S21/index.html>

³<https://www.logitech.com/en-eu/products/webcams/c920-pro-hd-webcam-960-001055.html>

⁴<https://github.com/dorian3d/DLoopDetector>



(a) Imagen satelital de una ciudad urbana. Karlsruhe, Alemania.



(b) Imagen satelital de una serie de campos agrícolas. Zavalla, Santa Fe, Argentina.

Figura 1. Se muestra que la diferencia visual entre imágenes tomadas a la misma distancia (~ 17 m) en entornos urbanos es alta, como muestran las imágenes en 1a tomadas del dataset KITTI, mientras que en entornos agrícolas presentan gran similitud, como muestran las imágenes en 1b tomadas del dataset FieldSAFE [20].

mos restringir la búsqueda a aquellas que se encuentren en la misma fila de píxeles o cercanas a la misma (parámetro configurable del sistema).

Por su parte el proceso de chequeo de consistencia geométrica realiza un filtrado de las características visuales que se calcularon en el emparejamiento estéreo tanto del par estéreo de entrada como del par estéreo que resulta de la búsqueda de similitud. Dicho filtrado consiste en encontrar emparejamientos entre las imágenes izquierda y derecha de ambos pares estéreo y restringirlos a aquellos que también son parte de los emparejamientos estéreo de cada par. De esta forma conseguimos restringir las características visuales a aquellas visibles en las cuatro imágenes (ambos pares estéreo), y utilizarlas para estimar la pose relativa entre ellas mediante la triangulación de los puntos estéreo del primer par y la proyección de dichos puntos 2D en el segundo par de imágenes, proceso conocido como “Perspective-n-Point”.

Habiendo logrado la estimación de pose relativa ya podemos decir que verificamos que ambos pares estéreo están observando el mismo escenario, sin embargo es a partir de la estimación de la pose relativa entre los mismos que podemos terminar nuestra detección de ciclo en un entorno de campo abierto. Sin embargo, como veremos en los resultados, esta estimación de la pose relativa es poco exacta en entornos agrícolas, por lo que todavía resta trabajo en lo que concierne a la misma.

IV. RESULTADOS

Realizamos experimentos de detecciones de ciclo tanto con el sistema monocular como con el sistema estéreo (el cual es nuestro aporte) en las distintas sesiones elegidas del dataset FieldSAFE. Se mantuvieron los mismos parámetros comunes a ambos sistemas salvo la limitación al detector de características visuales, donde al sistema estéreo se le permitieron detectar el doble de características por imagen⁵ ya que al realizar el primer refinamiento estéreo (características en común entre la imagen derecha e izquierda) se solían reducir a la mitad.

Como mencionamos en la sección anterior, no fue posible una estimación de la pose relativa a la detección de ciclo en el dataset actual, ya que en las imágenes, las características visuales de mejor calidad se encuentran en la zona del horizonte, donde se encuentran los elementos distintivos de las imágenes. Dichas características del horizonte, al tener descriptores más robustos suelen predominar luego de filtrarlas mediante emparejamiento estéreo, sin embargo encontrarse a gran distancia de la cámara (en relación al “baseline” estéreo) constituyen un caso degenerado para el proceso de triangulación y por lo tanto la estimación de la pose relativa de las cámaras resulta en valores poco fidedignos. Por lo tanto realizamos un análisis cuantitativo de los resultados de las detecciones de ciclo de nuestro sistema sin tener en cuenta dicha estima, proponiendo como trabajo el mejorarlo e incluirlo en el método en un futuro.

En la Fig 3 se pueden apreciar los resultados agregados en forma de diagrama de caja y bigote (boxplot), que comparan las distancias de posición y orientación para las detecciones de ciclos resultantes de correr los métodos mencionados. Se puede observar que la adición de información estéreo permite al detector mejorar notablemente la calidad de las detecciones obtenidas, reduciendo la distancia media y máxima en todos los casos. En particular, para las distancias posicionales, se redujeron las medianas en torno al 24 % y las distancias máximas pasaron de 58 m a 30 m, mientras que para las distancias angulares no hubo cambios significativos. Sin embargo no podemos dejar de notar que nuestro sistema no es perfecto, ya que todavía existen detecciones de ciclo que superarían hasta los límites más grandes propuestos en la literatura (25 m y 20°). No será hasta tener una estima de la pose relativa a la detección, problema que mencionamos como trabajo futuro, que podamos realmente filtrar las detecciones por distancia.

Notamos además que existen ciertas detecciones con una distancia angular muy alta, cercana a los 180°, tanto para el sistema monocular como para el sistema estéreo, que supondrían ser ciclos detectados en los cuales la cámara se encuentra en direcciones diametralmente opuestas, los cuales suponemos que pueden surgir de algún problema en la orientación dada por el GPS que usamos de ground-truth, pero no hemos investigado aún su causa.

A modo de ilustración mostramos las trayectorias junto a las detecciones del sistema estéreo en la Fig 4.

⁵Este es 1000 características al monocular y 2000 al estéreo.

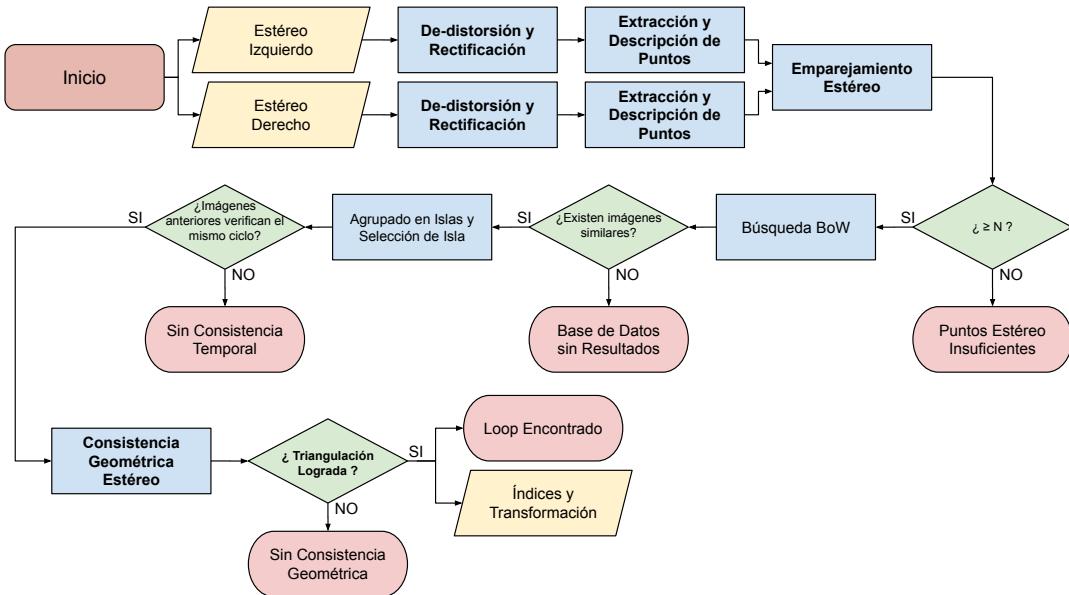


Figura 2. Gráfico de flujo del método de detección de ciclos, se re-implementaron los procesos demarcados en negrita para adaptarlos al procesamiento de imágenes estéreo.

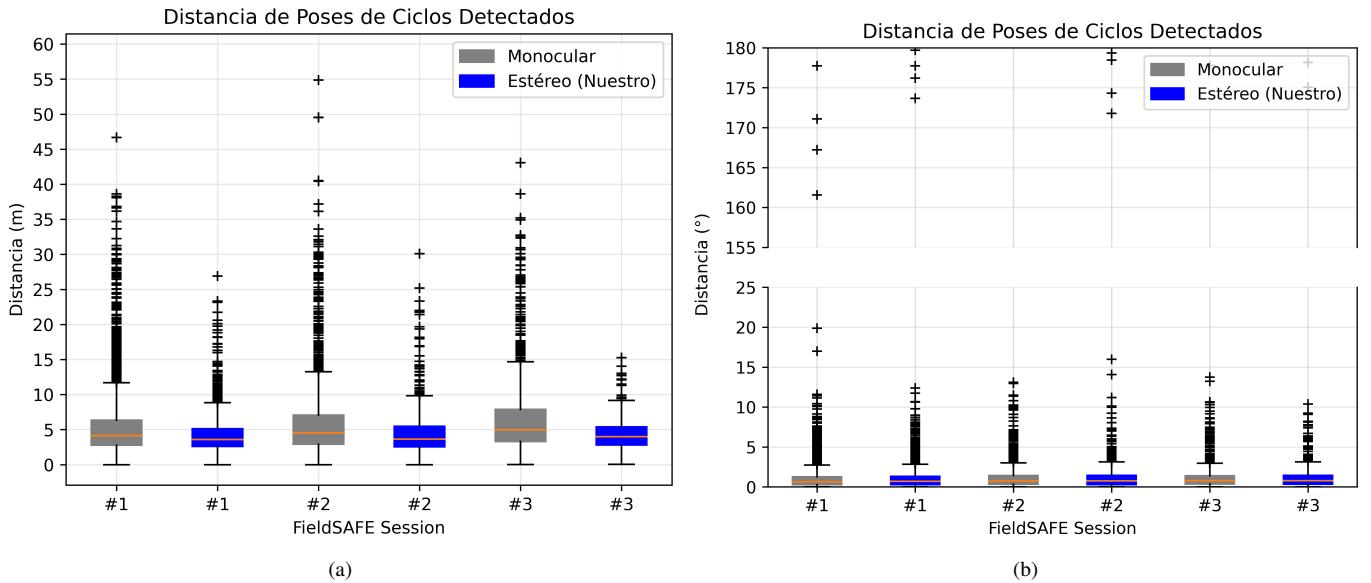


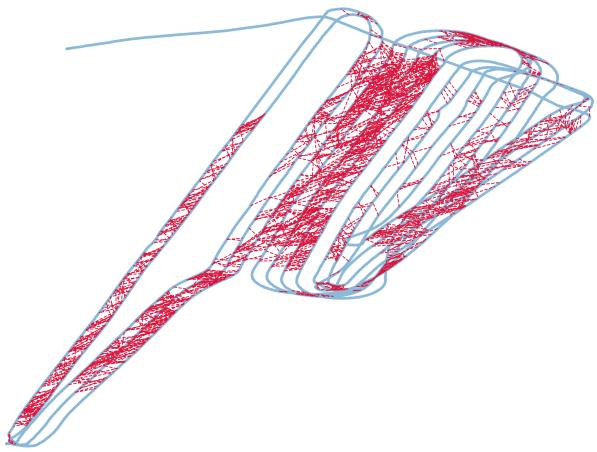
Figura 3. Comparativa de distancia posicional y angular agregadas, respectivamente, para los resultados obtenidos con los detectores monoculares y estéreo (nuestro aporte). En el gráfico de distancia angular se rompe el eje vertical para mejor visualización, sin pérdida de información.

V. CONCLUSIONES Y TRABAJO FUTURO

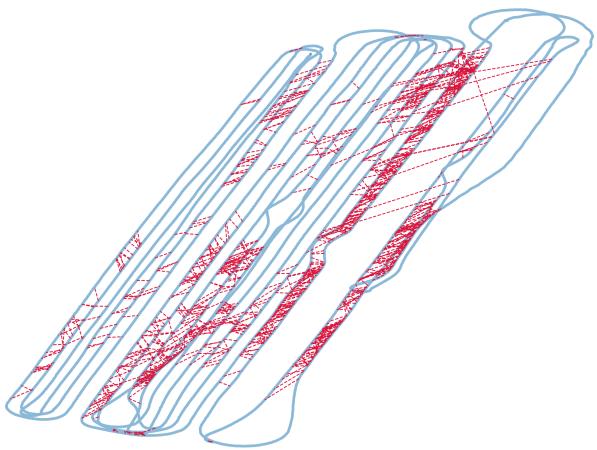
Mostramos un nuevo enfoque en la forma de trabajar para lograr una detección de ciclos en entornos agrícolas o de campo abierto para mitigar los problemas que no suelen darse en los entornos usuales de la literatura. Planteamos una modificación en un sistema clásico de detección de ciclos para incorporar información estéreo y mostramos cómo dicha información mejora la calidad de las detecciones acercándonos cada vez más al objetivo.

Sin embargo, como planteamos en el desarrollo, no po-

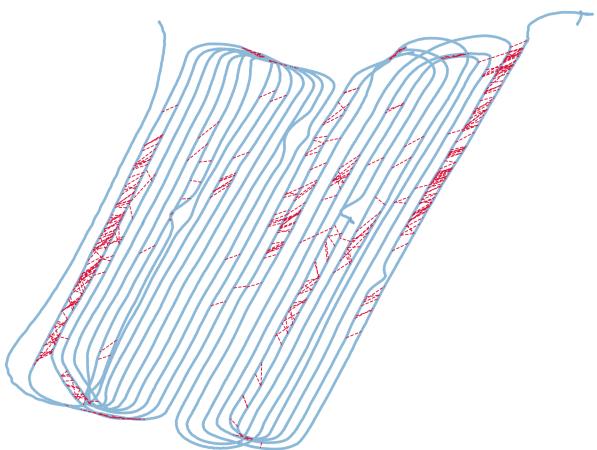
demos dejar de tener en cuenta que la detección de ciclos visual en entornos abiertos puede darnos positivos aún cuando existe una gran distancia entre las poses, por lo tanto debemos continuar nuestro trabajo para poder estimar con precisión la transformación de la pose respecto a la detección y así mejorar el filtrado de detecciones lejanas. Para tal fin estamos trabajando en hacer uso de los elementos presentes en el horizonte del entorno abierto y su semántica para poder triangular mejor la posición, ya sea por diferencia en tamaño o paralelo de dichos elementos, y si no es posible triangular la



(a) Dynamic obstacle session #1



(b) Dynamic obstacle session #2



(c) Static obstacle session #2

Figura 4. Resultados para el método propuesto en las sesiones de FieldSAFE. Se muestra la trayectoria del sistema en azul y las detecciones de ciclo en línea punteada roja conectando una pose su detección de ciclo, cuando ésta existe.

posición con robustez al menos dar una cota en la misma que sea útil para un sistema de cierre de ciclos, que es el objetivo final mayor.

AGRADECIMIENTOS

Este trabajo fue apoyado por CONICET (PIBAA N° 0042), AGENCIA I+D+i (PICT 2021-570) y Universidad Nacional de Rosario (PCCT-UNR 80020220600072UR).

REFERENCIAS

- [1] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, 06 2006.
- [2] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): part II," *IEEE Robotics and Automation Magazine*, vol. 13, no. 3, pp. 108–117, 09 2006.
- [3] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, Present, and Future of Simultaneous Localization and Mapping: Towards the Robust-Perception Age," *IEEE Trans. Robotics*, vol. 32, no. 6, pp. 1309—1332, 2016.
- [4] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 2161–2168.
- [5] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Trans. Robotics*, vol. 28, no. 5, pp. 1188–1197, October 2012. [Online]. Available: <https://ieeexplore.ieee.org/document/6202705>
- [6] T. Pire, T. Fischer, G. Castro, P. De Cristóforis, J. Civera, and J. Jacobo Berlles, "S-PTAM: Stereo Parallel Tracking and Mapping," *Journal of Robotics and Autonomous Systems*, vol. 93, pp. 27–42, 2017.
- [7] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM," *IEEE Trans. Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.
- [8] X. Gao and T. Zhang, "Unsupervised learning to detect loops using deep neural networks for visual SLAM system," vol. 41, no. 1, pp. 1–18, January 2017. [Online]. Available: <https://doi.org/10.1007/s10514-015-9516-2>
- [9] N. Merrill and G. Huang, "Lightweight Unsupervised Deep Loop Closure," in *Robotics: Science and Systems (RSS)*, Pittsburgh, Pennsylvania, June 2018.
- [10] A. R. Memon, H. Wang, and A. Hussain, "Loop closure detection using supervised and unsupervised deep neural networks for monocular SLAM systems," *Journal of Robotics and Autonomous Systems*, vol. 126, p. 103470, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889019308425>
- [11] R. Szeliski, G. Schindler, and M. Brown, "City-Scale Location Recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, june 2007, pp. 1–7. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/CVPR.2007.383150>
- [12] M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," *Intl. J. of Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008. [Online]. Available: <https://doi.org/10.1177/0278364908090961>
- [13] C. Cadena, D. Galvez-Lopez, J. Tardos, and J. Neira, "Robust place recognition with stereo sequences," *IEEE Trans. Robotics*, vol. 28, pp. 871–885, 08 2012.
- [14] R. Arandjelovic, P. Gronát, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," *Computing Research Repository (CoRR)*, 2015. [Online]. Available: <http://arxiv.org/abs/1511.07247>
- [15] S. Garg and M. Milford, "Seqnetvlad vs pointnetvlad: Image sequence vs 3d point clouds for day-night place recognition," CVPR 2021 Workshop on 3D Vision and Robotics (3DVR), Jun 2021.
- [16] S. Garg, N. Suenderhauf, and M. Milford, "Semantic–geometric visual place recognition: a new perspective for reconciling opposing views," *Intl. J. of Robotics Research*, vol. 41, no. 6, pp. 573–598, 2022. [Online]. Available: <https://doi.org/10.1177/0278364919839761>
- [17] R. Comelli, T. Pire, and E. Kofman, "Evaluation of visual slam algorithms on agricultural dataset," in *Workshop on Information Processing and Control (RPIC)*, 09 2019.

- [18] F. Shu, P. Lesur, Y. Xie, A. Pagani, and D. Stricker, "Slam in the field: An evaluation of monocular mapping and localization on challenging dynamic agricultural environment," in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2021, pp. 1760–1770.
- [19] H. Ding, B. Zhang, J. Zhou, Y. Yan, G. Tian, and B. Gu, "Recent developments and applications of simultaneous localization and mapping in agriculture," *Journal of Field Robotics*, vol. 39, no. 6, pp. 956–983, 2022. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22077>
- [20] M. F. Kragh, P. Christiansen, M. S. Laursen, M. Larsen, K. A. Steen, O. Green, H. Karstoft, and R. N. Jørgensen, "FieldSAFE: Dataset for Obstacle Detection in Agriculture," *Sensors*, vol. 17, no. 11, 2017. [Online]. Available: <https://www.mdpi.com/1424-8220/17/11/2579>
- [21] Sivic and Zisserman, "Video google: a text retrieval approach to object matching in videos," in *Proceedings Ninth IEEE International Conference on Computer Vision*, 2003, pp. 1470–1477 vol.2.
- [22] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2006, pp. 2161–2168.