# Hacia una Fusión Visual-Inercial-GNSS con Restricciones de Distancia


1st Javier Cremona
*CIFASIS*
*CONICET-UNR*
Rosario, Argentina
cremona@cifasis-conicet.gov.ar

2nd Javier Civera
*I3A*
*Universidad de Zaragoza*
Zaragoza, España
jcivera@unizar.es

3rd Taihú Pire
*CIFASIS*
*CONICET-UNR*
Rosario, Argentina
pire@cifasis-conicet.gov.ar



***Resumen*—La fusión de datos de cámaras, sensores inerciales y GNSS es una estrategia empleada para la localización y mapeo de robots móviles, especialmente en entornos abiertos, no estructurados y visualmente complejos. Este trabajo presenta un enfoque novedoso que incorpora las mediciones de GNSS como restricciones de distancia en la optimización conjunta visual-inercial, eliminando la necesidad de alinear los sistemas de coordenadas global y visual-inercial, y simplificando así el proceso de fusión. La evaluación se realiza con datos reales recolectados en un robot agrícola. Los resultados comparativos con ORB-SLAM3 y otros sistemas de fusión visual-inercial-GNSS demuestran la eficacia de la propuesta, con mejoras en la robustez de la estimación.**



***Keywords*—Fusión de Sensores, GNSS, Localización, SLAM, Agricultura de Precisión**


## I. Introducción

La localización precisa y robusta de robots móviles es fundamental para su navegación y operación segura en diversos entornos. Los robots modernos integran una variedad de sensores para obtener información del entorno y estimar su posición y orientación. La fusión de datos de diferentes sensores, como cámaras, unidades inerciales (*Inertial Measurement Unit*, IMU) y receptores de mediciones del sistema global de navegación por satélite (*Global Navigation Satellite System*, GNSS), ha demostrado ser un enfoque efectivo para mejorar la precisión y robustez de la estimación de la pose, especialmente en entornos desafiantes.

La fusión de información visual e inercial ha sido ampliamente estudiada en la última década [1]–[5]. Esta combinación ofrece una solución atractiva al combinar información exteroceptiva (cámara) y propioceptiva (IMU). Sin embargo, la estimación visual-inercial acumula error con el tiempo, lo que limita su precisión a largo plazo. Múltiples técnicas se han desarrollado para lidiar con este problema, entre ellas, detección y cierre de ciclos [6], [7] y la incorporación de sensores con información global como los receptores GNSS [8]–[10]. En particular, la detección de ciclos suele estar basada en aspectos puramente visuales. Esta característica hace que dichos métodos funcionen bien en ambientes interiores y estructurados, a la vez que fallan en ambientes visualmente desafiantes, como un campo agrícola, lugar en el que no abundan elementos distintivos que permitan un correcto reconocimiento de lugares previamente visitados [11].

Los receptores GNSS proveen mediciones de posición global, complementando la información visual e inercial. Sin embargo, la integración de GNSS presenta desafíos debido a la diferencia de referencia entre su sistema de coordenadas y el sistema visual-inercial. La alineación de estos dos sistemas de coordenadas ha sido abordada en trabajos previos [9], [10], [12], pero introduce complejidad y potenciales errores adicionales.

En este trabajo, presentamos un sistema de SLAM que fusiona cámara estéreo, IMU y GNSS de forma conjunta sin necesidad de alinear los sistemas de coordenadas. Nuestro enfoque, implementado como una mejora a [10], utiliza las mediciones de GNSS como restricciones de distancia euclídea en la optimización conjunta. El sistema presentado es evaluado en un conjunto de datos agrícolas en comparación con el sistema visual-inercial ORB-SLAM3 [1] y dos sistemas de fusión de cámaras, IMU y GNSS, VINS-Fusion [13] y GNSS-SI [10]. Puntualmente, las contribuciones de este artículo son las siguientes:

- Se presenta un sistema de SLAM que fusiona cámara estéreo, IMU y GNSS sin necesidad de alineación entre los sistemas de coordenadas global y visual-inercial.
- Se valida la performance del sistema mediante experimentación en datos reales recolectados en un robot operando en un campo agrícola.

El resto de este artículo se estructura de la siguiente manera: en la Sección II se describen trabajos relacionados. En la Sección III se describe el sistema propuesto, resaltando las diferencias respecto a [10]. Los experimentos realizados se detallan en la Sección IV. Finalmente, las conclusiones obtenidas de este trabajo se detallan en la Sección V.

## II. Trabajo Relacionado

Los métodos de fusión de sensores generalmente se clasifican en dos grandes grupos: métodos débilmente acoplados y métodos fuertemente acoplados. Los métodos débilmente acoplados no consideran las correlaciones entre mediciones de diferentes sensores. A partir de cada sensor se calcula la pose de forma independiente y posteriormente se fusionan estas estimaciones. En la mayoría de los casos, estos métodos se basan en el Filtro de Kalman Extendido (*Extended Kalman Filter*, EKF), aunque existen excepciones basadas

en optimización [13], [14]. En contraposición, los métodos fuertemente acoplados modelan la relación entre las variables de estado y las mediciones de los sensores. De esta manera, esta información se optimiza de forma conjunta, produciendo en general mejores resultados [15].

Lynen et al. [16] presentan MSF, un sistema de fusión de sensores modular basado en EKF. Las mediciones de IMU se emplean en el paso de predicción del filtro (*prediction step*). Las mediciones de otros sensores, tales como cámara, LiDAR o GNSS, se incorporan en el paso de actualización (*update step*). Similarmente, Shen et al. [17] proponen un método modular de fusión de sensores que proveen mediciones absolutas o relativas. Particularmente, el filtro empleado se basa en UKF.

Yu et al. [18] presentan un *framework* de estimación visual-inercial que extiende VINS-Mono [19] para soportar múltiples cámaras y fusionar, de forma débilmente acoplada, mediciones de GNSS. Adopta una estimación de estado basada en optimización. A su vez, Mascaro et al. [14] proponen GOMSF, un *framework* que estima la pose del robot a partir de la fusión entre estimaciones de MSF en coordenadas locales y mediciones de GNSS expresadas en coordenadas globales. Un enfoque similar es el que proponen Qin et al. [13], con la salvedad de ser un método más general y que soporta múltiples sensores globales.

Respecto a los métodos fuertemente acoplados, se destaca el trabajo de Lee et al. [12]. Los autores proponen un método visual-inercial que fusiona GNSS con la capacidad de estimar adicionalmente una calibración espacial y temporal entre GNSS e IMU de forma *online*. Por su parte, Cioffi et al. [8] desarrollan un método visual-inercial-GNSS fuertemente acoplado basado en optimización, el cual define factores globales a partir de mediciones de GNSS e información de la preintegración la IMU. Boche et al. [9] presentan una extensión de OKVIS2 [20] que combina cámaras, IMU y GNSS. A diferencia de [8], no asume que las mediciones globales vienen expresadas en el sistema de coordenadas visual-inercial, sino que introduce un sistema de coordenadas global. Como consecuencia, una estimación de la alineación entre ambos sistemas de coordenadas debe ser realizada. Los trabajos anteriormente mencionados emplean mediciones estándar de GNSS. Un enfoque alternativo es fusionar mediciones de GNSS *raw*. Tanto [21] como [22] plantean la fusión de pseudorango y efecto Doppler junto a la información visual-inercial.

Por último, Cremona et al. [10] presentan un sistema de fusión visual-inercial-GNSS fuertemente acoplado basado en ORB-SLAM3 [1]. Las mediciones globales se relativizan a la primera medición de GNSS en un sistema de coordenadas cartesiano local. Para poder incorporar esta información a la optimización conjunta es necesario estimar la rotación entre los sistemas de coordenadas visual-inercial y el sistema de coordenadas de las mediciones de GNSS. Dicha rotación, calculada durante la inicialización, permanece constante durante toda la vida del sistema. Por lo tanto, la fusión de mediciones GNSS resulta altamente dependiente de la precisión de la estimación de dicha rotación. En este trabajo se propone una modificación a [10], de forma tal que no es necesario estimar la mencionada rotación. Específicamente, las mediciones de GNSS se convierten en restricciones de distancia euclídea. El uso de mediciones de distancia a ubicaciones conocidas (usualmente denominado rango) se emplea en el contexto de sensores de tecnología de banda ultraancha (*Ultra-WideBand*, UWB) [23], [24]. Finalmente, el resultado de nuestro trabajo es un sistema que simplifica la etapa de inicialización y la fusión de mediciones GNSS.

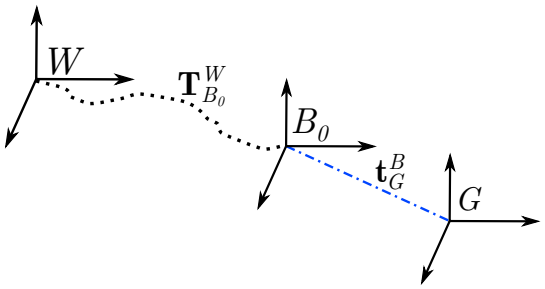



Figura 1. Sistemas de coordenadas empleados en este artículo. Se estima la trayectoria del robot respecto a un sistema de referencia fijo $W$. $B$ representa el sistema de coordenadas del cuerpo, particularmente aquí situado en la IMU. $B_0$ hace referencia al sistema de coordenadas $B$ en el instante de tiempo en que se toma la primera medición de GNSS. $W$ y $B_0$ se relacionan a través de la transformación o pose $\mathbf{T}_{B_0}^{W}$. La primera medición de GNSS es la que determina el origen de $G$, el cual se define con orientación global ENU (*East, North, Up*). En azul, $\mathbf{t}_G^B$ representa la posición de la antena de GNSS en $B$, la cual es fija y conocida de antemano en una etapa de calibración.

## III. Método propuesto

En esta sección se presenta el sistema propuesto. Dicho sistema está basado en [10]. El presente trabajo propone modificar la forma de incorporar la información provista por las mediciones de GNSS a la optimización conjunta. En primer lugar, se describe la notación empleada en este artículo. Posteriormente, se introduce el enfoque presentado en [10]. Finalmente, se desarrolla el nuevo enfoque propuesto en este artículo, enfatizando las diferencias con el trabajo previo.

### *III-A. Notación*

En este apartado brevemente se detallan los sistemas de coordenadas y la notación empleada en este artículo. Dichos sistemas de coordenadas se muestran gráficamente en la Figura 1. Un robot en movimiento se rastrea con respecto a un sistema de referencia fijo $W$. $B$ representa el sistema de coordenadas del cuerpo, que colocamos en el sensor IMU. Todas las mediciones de GNSS expresadas en latitud, longitud y altitud se transforman al sistema cartesiano local que denotamos como $G$, el cual se define de acuerdo a lo detallado en la subsección III-B. Mediante $\mathbf{a}^S$ se representan las coordenadas de una entidad geométrica $\mathbf{a}$ con respecto al sistema de coordenadas $S$. $\mathbf{R}_B^W \in SO(3)$ se refiere a la rotación de $B$ con respecto a $W$, y $\mathbf{t}_B^W \in \mathbb{R}^3$ representa la traslación del sistema de referencia $B$ expresada en el sistema $W$. La transformación de cuerpo rígido formada por la rotación $\mathbf{R}_B^W$ y la traslación $\mathbf{t}_B^W$ se denota como $\mathbf{T}_B^W \in SE(3)$. Para las mediciones de

GNSS, $\mathbf{t}_G^B \in \mathbb{R}^3$ es la posición de la antena GNSS en $B$, y se asume que se conoce de antemano a partir de la calibración.

### *III-B. GNSS-SI*

El sistema presentado en [10] fusiona de manera fuertemente acoplada los datos de GNSS, cámara estéreo e IMU. Dicho sistema está basado en ORB-SLAM3 [1], un sistema visual-inercial del estado del arte. Su versión estéreo-inercial fue usada como base para implementar la fusión con mediciones GNSS.

Inicialmente, las mediciones de GNSS son asociadas al keyframe más cercano temporalmente siempre y cuando la diferencia temporal sea menor a un umbral fijo. Las mediciones de GNSS que no cumplan con esta condición son descartadas. La primera medición de GNSS asociada a un keyframe es considerada como el origen del sistema de coordenadas $G$. Luego, todas las mediciones de GNSS, originalmente expresadas en coordenadas geográficas (latitud, longitud y altitud), son transformadas al sistema de coordenadas cartesiano $G$, el cual es definido con orientación ENU (*East*, *North*, *Up*)

La fusión se implementó modificando el grafo de factores correspondiente al *bundle adjustment* local de ORB-SLAM3 [1], agregando un factor para las mediciones de GNSS. La Figura 2 muestra el grafo de factores modificado. Tal como ocurre con otros sistemas que fusionan cámaras, IMU y GNSS [9], para definir la función de costo es necesario alinear el sistema de coordenadas visual-inercial y el sistema de coordenadas de las mediciones de GNSS, en este caso $G$. De este modo, podemos establecer una relación entre las mediciones GNSS expresadas en $G$ y la pose estimada por el sistema expresada en $W$. La alineación se realiza empleando el método de Umeyama [25]. Dicho método se aplica entre las primeras $K$ mediciones de GNSS y las poses estimadas por el sistema visual-inercial subyacente en el mismo período de tiempo. Como resultado se obtiene una matriz de rotación $\mathbf{R}_W^G$. Al contar con $\mathbf{R}_W^G$ queda definida la función de costo asociada al factor de GNSS y por lo tanto es posible fusionar las mediciones globales en el sistema. Derivamos al lector a nuestro trabajo previo [10] para más detalle acerca de la definición de la función de costo. Dado que $\mathbf{R}_W^G$ queda fija durante toda la trayectoria del robot, resulta evidente que la alineación debe realizarse de forma muy precisa, ya que de lo contrario, los errores cometidos en dicha estimación impactarán negativamente en la posterior fusión de las mediciones de GNSS.

### *III-C. Fusión GNSS-Stereo-Inertial empleando restricciones de distancia*

El presente trabajo propone una alternativa a [10] que permite fusionar cámara estéreo, IMU y mediciones de GNSS sin tener que alinear los sistemas de coordenadas $G$ y $W$. Al igual que en [10], la fusión se lleva a cabo en el *bundle adjustment* local. El grafo de factores correspondiente se muestra en la Figura 2. Las variables de estado a optimizar son $\mathcal{X} = \{\mathcal{X}_B, \mathcal{L}\}$, donde $\mathcal{X}_B = [\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_N]$ es el conjunto de estados para una ventana deslizante con los

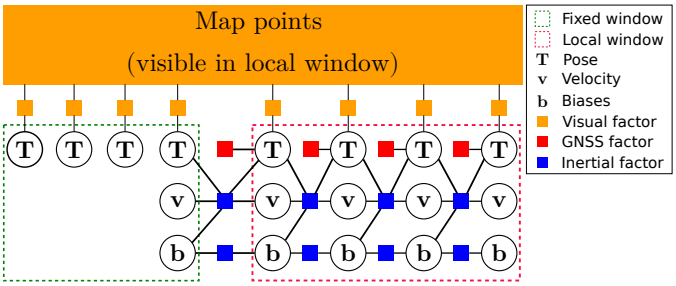



Figura 2. Grafo de factores correspondiente al *Bundle Adjustment* local del sistema propuesto. La ventana local (*Local window*) se compone de los últimos $N$ *keyframes*. La ventana fija (*Fixed window*) contiene *keyframes* que no pertenecen a la ventana local que están conectados en el grafo de covisibilidad a algún *keyframe* local. Estos últimos permanecen fijos durante la optimización. Además, el *keyframe* $N+1$ se incluye en la ventana fija ya que agrega restricción a las variables inerciales.

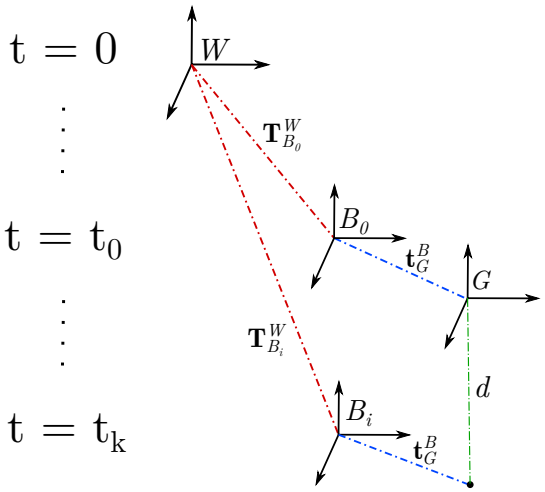



Figura 3. $d \in \mathbb{R}$ (en verde) es la distancia a comparar en la definición de la función de costo. Por un lado se obtiene mediante la medición de GNSS correspondiente al instante de tiempo $t_k$ y representa la distancia al origen de $G$. Por otro lado, se puede estimar la misma distancia en función de las poses expresadas en $W$. En particular, corresponde a la distancia entre las antenas de GNSS estimadas en $t_0$ y $t_k$. La misma está definida en función de $\mathbf{T}_{B_0}^W$, $\mathbf{T}_{B_i}^W$ (en rojo) y $\mathbf{t}_G^B$ (en azul).

últimos $N$ *keyframes* y $\mathcal{L} = [\mathbf{y}_1, \ldots, \mathbf{y}_j, \ldots, \mathbf{y}_M]$ es el conjunto de estados de los *landmarks* que se midieron durante esos últimos $N$ *keyframes*. El estado para el *keyframe* $i$ es:

$$\mathbf{x}_i = [\mathbf{T}_{B_i}^W, \mathbf{v}_i^\top, \mathbf{b}_{a_i}^\top, \mathbf{b}_{g_i}^\top], \tag{1}$$

y contiene la pose $\mathbf{T}_{B_i}^W \in SE(3)$, su velocidad local $\mathbf{v}_i \in \mathbb{R}^3$ y el bias del acelerómetro y el giroscopio $\mathbf{b}_{a_i} \in \mathbb{R}^3$ y $\mathbf{b}_{g_i} \in \mathbb{R}^3$. En cuanto a los *landmarks*, están representados por sus coordenadas en $W$, es decir, $\mathbf{y}_j = [X^W, Y^W, Z^W]^\top \in \mathbb{R}^3$.

Luego, la optimización viene dada por:

$$\hat{\mathcal{X}} = \arg\min_{\mathcal{X}} \left( \sum_{i=1}^{N} \|\mathbf{r}_{\mathcal{I}_{i-1,i}}\|^2_{\Sigma^{-1}_{\mathcal{I}_{i-1,i}}} + \right. \\ + \sum_{j=1}^{M} \sum_{i \in \mathcal{K}_j} \rho \left( \|\mathbf{r}_{\mathcal{V}_{ij}}\|_{\Sigma^{-1}_{V_{ij}}} \right) \\ \left. + \sum_{j=1}^{N} \sum_{i \in \mathcal{N}_j} \|\mathbf{r}_{\mathcal{G}_{ij}}\|^2_{\Sigma^{-1}_{\mathcal{G}_i}} \right), \tag{2}$$

donde $\rho$ es la función de Huber, $\mathcal{K}_j$ es el conjunto de *keyframes* que observan al *landmark* $j$ y $\mathcal{N}_j$ es el conjunto de mediciones de GNSS asociadas al *keyframe* $j$. Cada uno de los sumandos corresponde a las restricciones inerciales, visual y de GNSS, respectivamente. La modificación respecto a [10] es la definición de $\mathbf{r}_{\mathcal{G}_{ij}}$. Por cuestiones de espacio no se definirán tanto $\mathbf{r}_{\mathcal{I}_{i-1,i}}$ como $\mathbf{r}_{\mathcal{V}_{ij}}$, las cuales pueden consultarse en su artículo original [10].

Finalmente, se define $\mathbf{r}_{\mathcal{G}_{ij}}$ como:

$$\mathbf{r}_{\mathcal{G}_{ij}} = \left\| \hat{\mathbf{z}}_i^G \right\| - \left\| \mathbf{R}_{B_i}^W \mathbf{t}_G^B + \mathbf{t}_{B_i}^W - \left( \mathbf{R}_{B_0}^W \mathbf{t}_G^B + \mathbf{t}_{B_0}^W \right) \right\|, \tag{3}$$

donde $\hat{\mathbf{z}}_i^G \in \mathbb{R}^3$ es una medición de GNSS asociada al *keyframe* $j$ expresada en $G$, $\mathbf{R}_{B_0}^W$ y $\mathbf{t}_{B_0}^W$ es la pose del *keyframe* cuya medición definió el origen de $G$, la cual queda fija en esta optimización y $\mathbf{R}_{B_i}^W$ y $\mathbf{t}_{B_i}^W$ conforman la pose del *keyframe*. Una visualización gráfica puede observarse en la Figura 3.

Notar que no fue necesario alinear los sistemas de coordenadas $G$ y $W$. En este caso, se emplean restricciones de distancia computadas a partir de las mediciones de GNSS. En particular, $\hat{\mathbf{z}}_i^G$ es la distancia euclídea entre la medición de GNSS que seteó el origen de $G$ y la medición de GNSS $i$ asociada al *keyframe* $j$. En $\mathbf{r}_{\mathcal{G}_{ij}}$ dicha distancia se compara contra la distancia entre las posiciones de las antenas en los tiempos correspondientes obtenidas en función de las poses estimadas.

## IV. Experimentos

En esta sección se detallan los experimentos realizados para evaluar el desempeño del sistema propuesto. Los experimentos se llevaron a cabo utilizando el Rosario Dataset [26], una colección de datos sensoriales obtenidos de un robot en funcionamiento en un campo agrícola. Este conjunto de datos abarca imágenes estéreo capturadas a una frecuencia de 15 Hz, mediciones de una IMU a 142 Hz (compuesta por giroscopio y acelerómetro), odometría de ruedas obtenida a 10 Hz, y mediciones de RTK-GNSS a 5 Hz. Este último sensor se emplea como referencia real para la posición del robot (*ground-truth*). Dado que en el Rosario Dataset las mediciones de GNSS se utilizan como *ground-truth* y no se dispone de otro receptor GNSS a bordo ni de mediciones de GNSS crudas sin corrección, se opta por corromper el RTK-GNSS con ruido blanco Gaussiano aditivo [8], [10]. Esta técnica permite además simular el comportamiento errático de un GNSS de bajo costo o de un receptor operando en un entorno con poca visibilidad de satélites. En este experimento se emplea un error

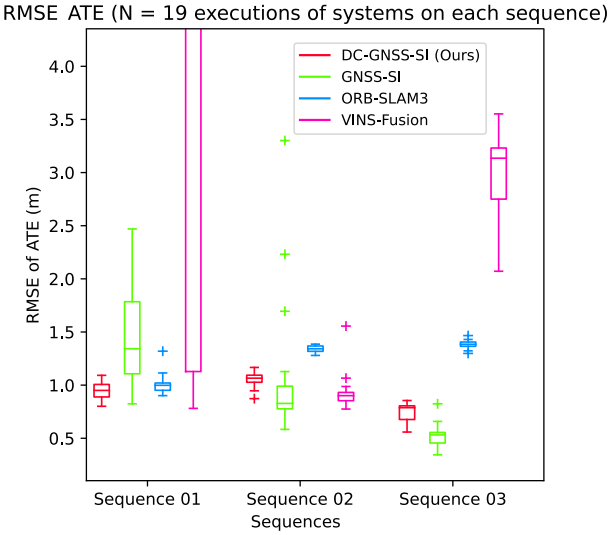
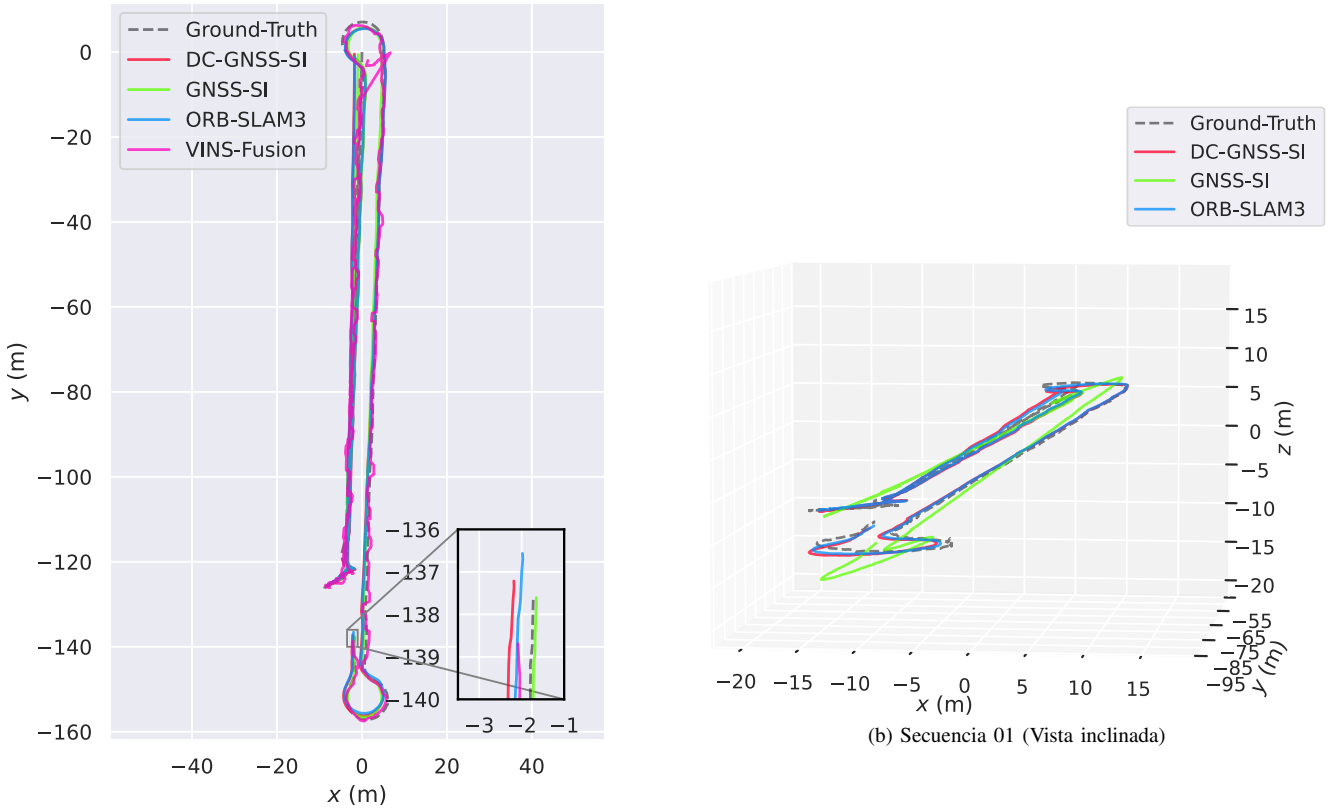



Figura 4. Gráfico de cajas representando el RMSE de ATE (*Absolute Trajectory Error*) para $N = 19$ ejecuciones de cada sistema sobre cada una de las secuencias 01, 02 y 03.

Gaussiano aditivo parametrizado por $\mathbf{n}_p \sim \mathcal{N}\left(\mathbf{0}, \sigma_p^2 \cdot \mathbf{I}\right)$, con desviación estándar $\sigma_p = 50\,\text{cm}$, un valor razonable para poner a prueba la robustez del método ante situaciones desfavorables.

Para evaluar la precisión y robustez de nuestro enfoque, se compara nuestro sistema, denominado DC-GNSS-SI, con su versión previa de fusión fuertemente acoplada, GNSS-SI [10], el sistema visual-inercial del estado del arte, ORB-SLAM3 [1], y un sistema de fusión débilmente acoplada, VINS-Fusion [13]. Para establecer una comparación más justa, se deshabilita la detección y cierre de ciclos en ORB-SLAM3. Se adopta el Error de Trayectoria Absoluto (*Absolute Trajectory Error*, ATE) [27] como métrica para comparar la precisión de los sistemas. Dado que el ruido se añade a las mediciones de GNSS de forma online y es distinto para cada ejecución de los sistemas, se propone ejecutar los sistemas un número $N$ de veces sobre cada una de las seis secuencias del Rosario Dataset. Esto permite analizar la robustez del enfoque propuesto. Como suele hacerse generalmente, cada una de las trayectorias estimadas se alinea con la trayectoria *ground-truth* mediante el método de Umeyama [25]. A diferencia de lo habitual, donde se presenta un único resultado, se opta por mostrar el RMSE del ATE para las $N$ ejecuciones para todas las secuencias. Esto se resume en la Figura 4. Cabe mencionar que el gráfico está truncado debido a que los resultados de RMSE de ATE para VINS-Fusion en la secuencia 01 difieren en orden de magnitud. Por este motivo, se presenta además la Tabla I, en la cual se muestran los valores mínimos, máximos y mediana del RMSE del ATE en las $N$ ejecuciones de los sistemas. Si observamos la secuencia 01, puede observarse que el mínimo RMSE de ATE es obtenido por VINS-Fusion. Sin embargo, tanto en el gráfico como en la tabla queda claro que VINS-Fusion no es consistente entre ejecuciones, teniendo valores de RMSE de ATE mucho mayores en general. Este es la principal razón por la que se presentan los datos en forma de gráfico




(a) Secuencia 01 (Vista superior)

(b) Secuencia 01 (Vista inclinada)

Figura 5. Vista superior (izquierda) y vista inclinada (derecha) para las estimaciones de los sistemas sobre la Secuencia 01. De las $N$ ejecuciones de los sistemas se escoge la mediana de acuerdo al RMSE del ATE. Se quita VINS-Fusion en el gráfico de la derecha para una mejor visualización.

de cajas.

Como puede verse a simple vista en el gráfico, los resultados revelan que DC-GNSS-SI, nuestro sistema propuesto, claramente supera a su sistema visual-inercial subyacente ORB-SLAM3, dando validez al enfoque presentado y a la fusión de mediciones de GNSS aún en presencia de ruido. Si bien en general GNSS-SI presenta un menor RMSE de ATE que DC-GNSS-SI, puede observarse una menor dispersión en los valores obtenidos por nuestro nuevo enfoque, especialmente en las secuencias 01 y 02. Esto es un indicador de que DC-GNSS-SI presenta mayor robustez al ruido de las mediciones de GNSS. Adjudicamos esta problemática en GNSS-SI a la dificultad para inicializar y alinear los sistemas de coordenadas $G$ y $W$. Tal como se mencionó en la Sección III-B, $\mathbf{R}_W^G$ queda fijo para toda la estimación de la trayectoria, por lo tanto, es imprescindible que su cálculo sea preciso. Aún mas, la trayectoria en línea recta característica de los ambientes agrícolas debido a la disposición de los surcos, hace que el problema de la alineación quede mal condicionado en estos casos. Por este motivo, creemos que la alternativa propuesta, en la cual nos deshacemos de $\mathbf{R}_W^G$, resulta atractiva, presentando una mayor robustez en los datos analizados. Finalmente, respecto a VINS-Fusion, si bien en algunas secuencias presenta un bajo RMSE de ATE, en algunas secuencias el sistema pierde el *tracking* visual. La Figura 5 muestra las trayectorias correspondientes a la mediana en cada caso para la secuencia 01. El RMSE ATE para estas trayectorias se presenta en la Tabla I. Puede verse que la trayectoria estimada por VINS-Fusion tiene muchos saltos, volviéndola inusable en métodos de control y navegación. Esto es producto de la fusión débilmente acoplada, en la que no se tienen en cuenta la relación de las mediciones de los distintos sensores en forma conjunta. Por último, en la Figura 5b puede observarse que la estimación de GNSS-SI no retorna una trayectoria en la que el robot se mueve sobre un plano, si no que acumula error. Esto se lo adjudicamos a una mala estimación de $\mathbf{R}_W^G$, lo que contamina la fusión de todas las posteriores mediciones de GNSS.

## V. Conclusiones y Trabajo Futuro

En este trabajo se presenta un sistema de SLAM que fusiona cámara estéreo, IMU y mediciones de GNSS. A diferencia de trabajos previos de fusión de mediciones globales [9], [10], se propone un método en el que no es necesario estimar la alineación entre los sistemas de coordenadas visual-inercial y global. En cambio, se emplean restricciones de distancias obtenidas a partir de las mediciones de GNSS. La experimentación se realiza sobre un conjunto de datos agrícolas reales. Los resultados muestran que el sistema propuesto resulta más

Tabla I
MÍNIMOS, MÁXIMOS Y MEDIANAS DEL RMSE DE ATE (*Absolute Trajectory Error*) HABIENDO EJECUTADO $N = 19$ VECES CADA SISTEMA SOBRE CADA UNA DE LAS SECUENCIAS.

| System | Min. RMSE ATE (m) | | | Max. RMSE ATE (m) | | | Median RMSE ATE (m) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Seq. 01 | Seq. 02 | Seq. 03 | Seq. 01 | Seq. 02 | Seq. 03 | Seq. 01 | Seq. 02 | Seq. 03 |
| DC-GNSS-SI (ours) | 0.801 | 0.873 | 0.559 | 1.092 | 1.166 | 0.855 | 0.950 | 1.065 | 0.788 |
| GNSS-SI [10] | 0.824 | 0.584 | 0.344 | 2.469 | 3.301 | 0.824 | 1.342 | 0.828 | 0.531 |
| ORB-SLAM3 [1] | 0.901 | 1.280 | 1.298 | 1.319 | 1.386 | 1.466 | 0.998 | 1.342 | 1.384 |
| VINS-Fusion [13] | 0.781 | 0.775 | 2.072 | 2895.625 | 1.556 | 3.552 | 223.292 | 0.901 | 3.135 |

robusto mientras que sigue siendo competitivo respecto a la precisión en comparación con nuestro trabajo previo [10].

Como trabajo futuro, se propone la incorporación de un sensor que provea información acerca de la orientación global, tal como el magnetómetro. De esta manera, podremos corregir no solo la posición, si no también la orientación, logrando así una estimación más robusta de la pose del robot.

## AGRADECIMIENTOS

Este trabajo fue apoyado por CONICET (PIBAA Nº 0042), AGENCIA I+D+i (PICT 2021-570) y Universidad Nacional de Rosario (PCCT-UNR 80020220600072UR).

## REFERENCIAS

[1] 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.

[2] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *Intl. J. of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[3] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight," *(IEEE) Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, 2018.

[4] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems," *IEEE Trans. Robotics*, vol. 33, no. 2, pp. 249–265, 2017.

[5] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2015, pp. 298–304.

[6] D. Galvez-López and J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Trans. Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[7] X. Gao, R. Wang, N. Demmel, and D. Cremers, "LDSO: Direct Sparse Odometry with Loop Closure," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE Press, 2018, pp. 2198—-2204. [Online]. Available: https://doi.org/10.1109/IROS.2018.8593376

[8] G. Cioffi and D. Scaramuzza, "Tightly-coupled Fusion of Global Positional Measurements in Optimization-based Visual-Inertial Odometry," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2020, pp. 5089–5095.

[9] S. Boche, X. Zuo, S. Schaefer, and S. Leutenegger, "Visual-Inertial SLAM with Tightly-Coupled Dropout-Tolerant GPS Fusion," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022, pp. 7020–7027.

[10] J. Cremona, J. Civera, E. Kofman, and T. Pire, "GNSS-stereo-inertial SLAM for arable farming," *Journal of Field Robotics*, vol. n/a, no. n/a. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22232

[11] J. Cremona, R. Comelli, and T. Pire, "Experimental evaluation of Visual-Inertial Odometry systems for arable farming," *Journal of Field Robotics*, vol. 39, no. 7, pp. 1123–1137, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.22099

[12] W. Lee, K. Eckenhoff, P. Geneva, and G. Huang, "Intermittent GPS-aided VIO: Online Initialization and Calibration," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020, pp. 5724–5731.

[13] T. Qin, S. Cao, J. Pan, and S. Shen, "A General Optimization-based Framework for Global Pose Estimation with Multiple Sensors," 2019.

[14] R. Mascaro, L. Teixeira, T. Hinzmann, R. Siegwart, and M. Chli, "GOMSF: Graph-Optimization Based Multi-Sensor Fusion for robust UAV Pose estimation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2018, pp. 1421–1428.

[15] H. Strasdat, J. Montiel, and A. J. Davison, "Visual SLAM: Why filter?" *Image and Vision Computing*, vol. 30, no. 2, pp. 65–77, 2012.

[16] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to MAV navigation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 3923–3929.

[17] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar, "Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2014, pp. 4974–4981.

[18] Y. Yu, W. Gao, C. Liu, S. Shen, and M. Liu, "A GPS-aided Omnidirectional Visual-Inertial State Estimator in Ubiquitous Environments," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019, pp. 7750–7755.

[19] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Trans. Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[20] S. Leutenegger, "OKVIS2: Realtime Scalable Visual-Inertial SLAM with Loop Closure," 2022. [Online]. Available: https://arxiv.org/abs/2202.09199

[21] S. Cao, X. Lu, and S. Shen, "GVINS: Tightly Coupled GNSS-Visual-Inertial Fusion for Smooth and Consistent State Estimation," *IEEE Transactions on Robotics*, vol. 38, no. 4, pp. 2004–2021, 2022.

[22] J. Liu, W. Gao, and Z. Hu, "Optimization-Based Visual-Inertial SLAM Tightly Coupled with Raw GNSS Measurements," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2021, pp. 11 612–11 618.

[23] M. Pacholska, F. Dümbgen, and A. Scholefield, "Relax and Recover: Guaranteed Range-Only Continuous Localization," *(IEEE) Robotics and Automation Letters*, vol. 5, no. 2, pp. 2248–2255, 2020.

[24] F. Dümbgen, C. Holmes, and T. D. Barfoot, "Safe and Smooth: Certified Continuous-Time Range-Only Localization," *(IEEE) Robotics and Automation Letters*, vol. 8, no. 2, pp. 1117–1124, 2023.

[25] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 4, pp. 376–380, 1991.

[26] T. Pire, M. Mujica, J. Civera, and E. Kofman, "The Rosario Dataset: Multisensor Data for Localization and Mapping in Agricultural Environments," *Intl. J. of Robotics Research*, vol. 38, no. 6, pp. 633–641, 2019.

[27] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580.