

Jornadas de Automática

Estimación de pose basada en marcadores ArUco usando cámaras de eventos

J. Luna-Santamaría^a, R. Tapia^{a,*}, J.R. Martínez-de Dios^a, A. Ollero^a

^aGRVC Robotics Laboratory, Universidad de Sevilla.

To cite this article: Luna-Santamaría, J., Tapia, R., Martínez-de Dios, J.R., Ollero, A. 2026. Pose Estimation Based on ArUco Markers Using Event Cameras. *Jornadas de Automática*, 47.
<https://doi.org/10.17979/ja-cea.2026.47.13871>

Resumen

La estimación de pose es una técnica fundamental en robótica y visión por computador. Sin embargo, los sistemas basados en cámaras convencionales presentan limitaciones ante movimientos rápidos, desenfoque y condiciones de iluminación adversas. Las cámaras de eventos surgen como una alternativa, capturando cambios de intensidad de forma asíncrona con alta resolución temporal y amplio rango dinámico, lo que las hace especialmente adecuadas para escenarios dinámicos. Este trabajo presenta un método de estimación de pose basado en marcadores ArUco utilizando únicamente eventos. El método emplea un detector de marcadores diseñado para este tipo de sensores, permitiendo su identificación sin necesidad de reconstrucción de imagen. A partir de los marcadores detectados, la pose de la cámara se estima mediante un método PnP sobre correspondencias 2D-3D. Los resultados experimentales a bordo de un robot aéreo tipo multirrotor volando en interior y exterior muestran que el sistema es capaz de proporcionar estimaciones precisas a más de 100 Hz en CPU.

Palabras clave: Cámaras de eventos, Estimación de pose, Visión por computador, Marcadores fiduciales, Localización visual

Pose Estimation Based on ArUco Markers Using Event Cameras

Abstract

Pose estimation is a fundamental technique in robotics and computer vision. However, systems based on conventional cameras suffer from limitations under fast motion, motion blur, and challenging lighting conditions. Event cameras have emerged as an alternative, capturing intensity changes asynchronously with high temporal resolution and a wide dynamic range, making them particularly suitable for dynamic scenes. This work presents a pose estimation method based on ArUco markers using only events. The method employs a marker detector specifically designed for event-based sensors, enabling marker identification without the need for image reconstruction. From the detected markers, the camera pose is estimated using a PnP approach based on 2D-3D correspondences. Experimental results on board a multirotor aerial robot flying indoors and outdoors show that the system is capable of providing accurate pose estimates at over 100 Hz on CPU.

Keywords: Event cameras, Pose estimation, Computer vision, Fiducial markers, Visual localization

1. Introducción

La estimación de pose mediante cámaras es un componente fundamental en arquitecturas de visión por computador, robótica, vehículos autónomos y sistemas de realidad virtual/aumentada, campos en los que suele ser necesario conocer con precisión la posición y orientación de una cámara en el entorno. Los métodos basados en cámaras convencionales han alcanzado un alto grado de madurez, especialmente aque-

llos que usan marcadores fiduciales (Kalaitzakis et al. (2021); Jurado-Rodríguez et al. (2023)) como los marcadores ArUco (Garrido-Jurado et al. (2014, 2016, 2026); Romero-Ramírez et al. (2018)). El uso de marcadores para estimación de pose es una estrategia eficaz y ampliamente utilizada para obtener una localización precisa y facilitar tareas de navegación autónoma Muñoz-Salinas et al. (2019); Muñoz-Salinas and Medina-Carnicer (2020). Sin embargo, estos enfoques presentan limi-

*Autor para correspondencia: raultapia@us.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

taciones en escenarios dinámicos, donde el movimiento rápido, el desenfoque y las condiciones de iluminación adversas degradan significativamente el rendimiento.

Las cámaras de eventos han emergido como una alternativa prometedora a los sensores tradicionales. A diferencia de las cámaras convencionales, que capturan imágenes de forma síncrona, las cámaras de eventos registran cambios de intensidad de manera asíncrona a nivel de píxel, proporcionando alta resolución temporal, bajo consumo y un amplio rango dinámico. Estas propiedades las hacen especialmente adecuadas para entornos con alta dinámica y variaciones de iluminación. No obstante, su naturaleza asíncrona y la ausencia de información de intensidad absoluta dificultan la aplicación directa de algoritmos de visión clásicos (Tapia et al. (2024, 2026)).

En el contexto de la detección de marcadores con cámaras de eventos, pueden distinguirse tres enfoques principales (Tofighi et al. (2025)). En primer lugar, los métodos basados en marcadores activos (como Huang et al. (2023), Zhang et al. (2023) o Kitade et al. (2024)) emplean patrones luminosos o modulaciones de intensidad para generar eventos de forma controlada, lo que facilita su detección, pero requiere hardware específico y aumenta la complejidad del sistema. En segundo lugar, algunos trabajos recurren a la reconstrucción de imágenes a partir del flujo de eventos (como Mai et al. (2023) o Loch et al. (2023)), permitiendo aplicar técnicas de visión por computador convencionales; sin embargo, este proceso introduce latencia adicional y puede degradar el rendimiento en escenarios complejos. Finalmente, existen algunos enfoques que operan directamente sobre los datos de eventos, como el propuesto en Sarmadi et al. (2021), aunque presentan limitaciones bajo condiciones de observabilidad parcial.

En este trabajo se presenta un método de estimación de pose basado en marcadores ArUco utilizando cámaras de eventos. El enfoque propuesto emplea un detector de marcadores diseñado específicamente para datos de eventos (Figura 1), permitiendo su identificación y localización sin necesidad de reconstrucción de imagen. A partir de los marcadores detectados, la pose de la cámara se estima mediante un método de Perspective-n-Point (PnP) sobre correspondencias 2D-3D. El método ha sido validado a bordo de un robot aéreo tipo multirrotor en distintos vuelos reales en escenarios interiores y exteriores. Los resultados experimentales muestran que el sistema es capaz de proporcionar estimaciones de pose precisas a frecuencias superiores a 100 Hz en CPU, manteniendo un rendimiento adecuado incluso ante movimientos complejos.

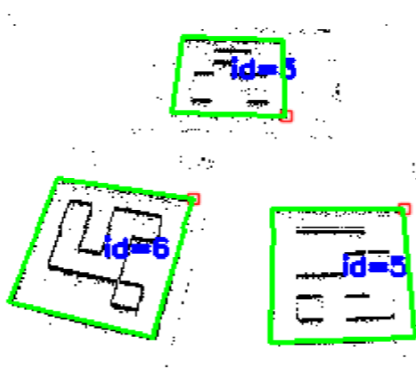


Figura 1: Detección de marcadores ArUco a partir del flujo de eventos capturado por una cámara de eventos.

El resto del trabajo se estructura de la siguiente forma. La Sección 2 describe el método propuesto para estimación de pose. La Sección 3 presenta la validación experimental y los resultados obtenidos. Finalmente, la Sección 4 recoge las principales conclusiones del trabajo.

2. Método

El método propuesto permite estimar la pose de una cámara de eventos con respecto a un sistema de referencia global definido por un conjunto de marcadores ArUco de tamaño conocido. Para ello se emplean únicamente los eventos generados por el sensor, sin reconstruir imágenes de intensidad y sin modificar físicamente los marcadores. Cada evento se representa como $e = (x, y, t, p)$, donde (x, y) son las coordenadas del píxel en el que se produce el evento, t es la marca temporal (*timestamp*) y $p \in \{+1, -1\}$ indica la polaridad del cambio de brillo (positiva si se trata de un incremento de brillo y negativa en caso contrario). El método transforma este flujo asíncrono en detecciones de marcadores $(ID_k, \mathbf{c}_k)$, donde ID_k es el identificador del marcador y $\mathbf{c}_k$ son sus cuatro esquinas en el plano de imagen. La pose se calcula a partir de esas detecciones, utilizando las correspondencias entre esquinas 3D conocidas y esquinas 2D detectadas.

Una cámara de eventos no mide intensidad absoluta, sino variaciones de brillo. Bajo la hipótesis de brillo constante, la generación de eventos estará relacionada con el movimiento relativo cámara-escena. En un marcador ArUco, las regiones interiores homogéneas generan pocos eventos, mientras que las fronteras entre celdas blancas y negras producen un número de eventos significativamente mayor. El detector empleado no decodifica intensidades de celdas, como en una cámara convencional, sino la presencia de eventos en los bordes. El flujo de eventos se convierte primero en una representación espacial sobre la que se detectan candidatos a marcador. Cada candidato se identifica mediante una representación basada en bordes. El identificador recuperado permite asociar las esquinas 2D observadas con las esquinas 3D del mapa. Finalmente, la pose de la cámara se estima resolviendo un problema PnP. La Figura 2 resume la arquitectura general del método.

2.1. Detección de marcadores

Los eventos generados por la cámara se acumulan en imágenes de eventos de tamaño fijo, de modo que cada imagen contiene siempre el mismo número de eventos recientes. Esta representación permite transformar temporalmente el flujo asíncrono de eventos en una estructura bidimensional sobre la que pueden aplicarse operaciones geométricas convencionales. En este trabajo se utilizan 5000 eventos por imagen. La elección de una acumulación por número de eventos, en lugar de por intervalo temporal fijo, mantiene una densidad de información más homogénea entre imágenes consecutivas, incluso cuando la velocidad relativa entre la cámara y los marcadores cambia durante el movimiento. Una vez construida cada imagen de eventos, se aplica una etapa de preprocesado destinada a mejorar la calidad de los bordes. En primer lugar, se eliminan activaciones aisladas que no forman parte de estructuras coherentes utilizando un filtro espacial basado en la vecindad. Después se aplica un suavizado gaussiano con kernel de 5×5 .

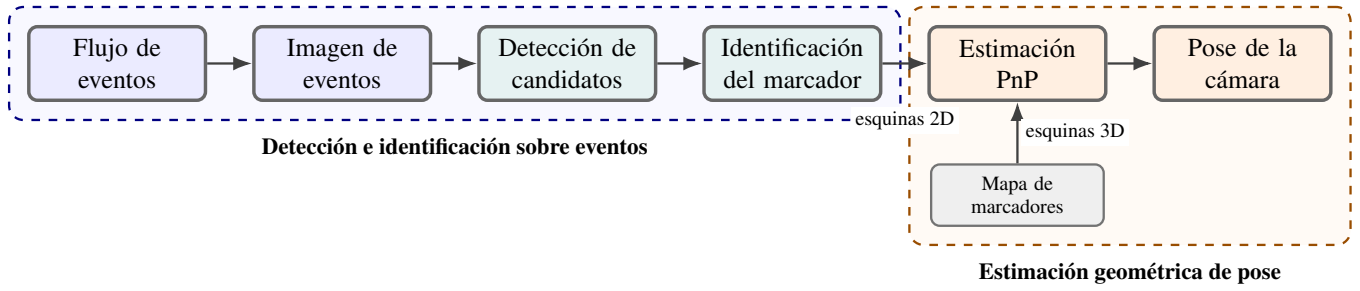


Figura 2: Arquitectura general del método propuesto. El flujo de eventos se acumula en imágenes de eventos, sobre las que se detectan candidatos a marcador y se recupera su identificador. Las esquinas 2D detectadas se asocian con las esquinas 3D del mapa de marcadores y la pose de la cámara se obtiene resolviendo un problema PnP.

píxeles y $\sigma = 1$, que atenúa ruido de alta frecuencia y favorece la continuidad espacial de los contornos. A continuación, la imagen se binariza para obtener una representación discreta de las regiones activadas por eventos.

Sobre la imagen filtrada se extraen contornos mediante el algoritmo de detección de bordes de Suzuki and Abe (1985). Los contornos obtenidos se aproximan mediante segmentos rectilíneos (Douglas and Peucker (1973)). Esta aproximación reduce la complejidad de la representación y permite describir los bordes observados del marcador mediante primitivas geométricas simples. A partir del conjunto de segmentos detectados, el método busca pares que puedan corresponder a lados opuestos de un marcador cuadrado. Para ello se aplican varias restricciones geométricas: los segmentos deben tener longitudes compatibles, ser aproximadamente paralelos y formar, con sus cuatro extremos, un cuadrilátero convexo válido. Los pares de segmentos que satisfacen estas condiciones generan candidatos cuadriláteros. Cada candidato se reordena en sentido horario para obtener una convención consistente en la posición de sus esquinas. A continuación, se calcula una homografía que transforma el cuadrilátero detectado en la imagen en una plantilla cuadrada canónica. Esta rectificación normaliza el marcador respecto a escala, orientación y punto de vista. Todos los candidatos pueden analizarse sobre una misma malla regular, lo que permite aplicar un procedimiento común de identificación. La Figura 3 ilustra el proceso de detección sobre un marcador de ejemplo.

2.2. Identificación de marcadores

Una vez el marcador ha sido rectificado, se define sobre este una malla regular. A diferencia de los detectores basados en imágenes convencionales, no se intenta clasificar cada celda como blanca o negra a partir de su intensidad, ya que la cámara de eventos no proporciona una imagen de intensidad absoluta. En su lugar, el marcador se describe mediante las transiciones espaciales de su patrón binario (es decir, los bordes internos del marcador), que son precisamente las regiones donde se generan eventos cuando existe movimiento relativo entre la cámara y el marcador. Para un marcador con una disposición interna de $N \times N$ celdas, la malla contiene $N(N + 1)$ bordes horizontales y $N(N + 1)$ bordes verticales. Cada uno de estos bordes se evalúa en la imagen rectificada comprobando si existe suficiente actividad de eventos en la región correspondiente. El resultado es un valor binario por borde: un borde se marca como activo si la cantidad de eventos acumulados en su región supera un umbral y como inactivo en caso contra-

rio. En este trabajo se ha seleccionado un umbral adaptativo igual al 90 % de la mayor cantidad de eventos acumulados por borde. Concatenando todos estos valores se obtiene un código binario asociado al candidato. Este código no representa directamente los colores de las celdas, sino los bordes entre celdas adyacentes. La identificación se realiza comparando el código extraído con un diccionario construido previamente a partir del diccionario ArUco estándar. Para generar este diccionario, cada marcador ArUco se transforma en su representación equivalente basada en bordes. Esta conversión permite reutilizar diccionarios ArUco ya existentes.

El proceso de identificación sigue una estrategia jerárquica. En primer lugar, se extrae del candidato su código binario de bordes y se intenta localizar directamente en el diccionario mediante una búsqueda exacta. Si el código coincide con una entrada almacenada, el candidato se acepta con el identificador asociado. Si no existe una coincidencia exacta, se aplica una búsqueda por distancia de Hamming entre el código observado y los códigos almacenados. Esta segunda etapa introduce tolerancia frente a errores locales producidos por bordes espurios, ausentes o con ruido. El identificador asignado corresponde al marcador cuyo código presenta la menor distancia respecto al código observado. No obstante, el candidato se descarta si esa distancia mínima supera un umbral (en este trabajo se ha seleccionado mediante prueba y error una distancia de 4), evitando aceptar correspondencias poco fiables.

2.3. Estimación de pose

La estimación de pose se formula como un problema de correspondencias entre puntos 3D conocidos del entorno y puntos 2D observados en la imagen. El entorno contiene un conjunto de marcadores ArUco colocados en posiciones conocidas (mapa). Para cada marcador se conoce su identificador, su tamaño físico y su pose en un sistema de referencia global (mundo o *world*). Cuando el detector reconoce un marcador en la imagen, no solo proporciona las coordenadas 2D de sus esquinas, sino que también permite recuperar las coordenadas 3D correspondientes a esas mismas esquinas en el mapa.

Sea $\mathcal{M}$ el conjunto de marcadores detectados en una ventana de eventos. Cada marcador $m_i \in \mathcal{M}$ aporta cuatro esquinas $(\mathbf{u}_{i,1}, \mathbf{u}_{i,2}, \mathbf{u}_{i,3}, \mathbf{u}_{i,4})$ observadas en el plano de la imagen. Si el marcador ha sido identificado correctamente, sus esquinas 2D pueden asociarse con las cuatro esquinas 3D almacenadas en el mapa. De este modo se construye el conjunto de correspondencias $\mathcal{C} = \{(\mathbf{u}_{i,k}, \mathbf{X}_{i,k}) : m_i \in \mathcal{M}, k = 1, \dots, 4\}$, donde $\mathbf{X}_{i,k} \in \mathbb{R}^3$ representa la posición 3D de la esquina k del marca-

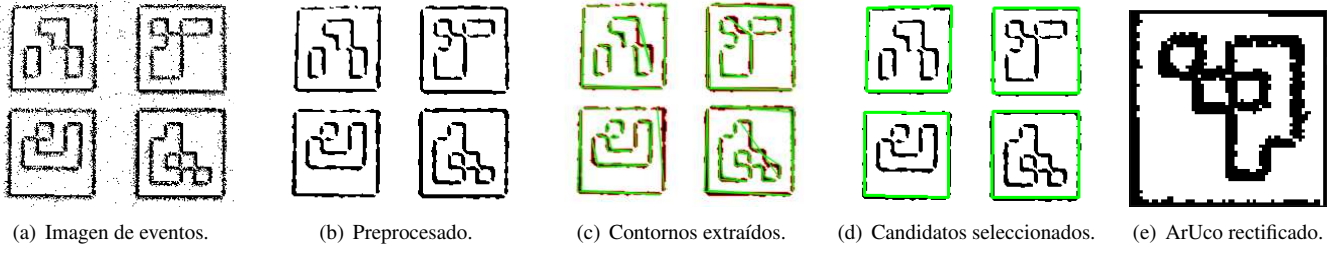


Figura 3: Ejemplo del proceso de detección de marcadores a partir de una imagen de eventos. A partir de la acumulación de eventos se obtiene una imagen sobre la que se aplica un preprocesado. Después se extraen contornos y se aproximan mediante segmentos rectilíneos. Los segmentos compatibles generan candidatos cuadriláteros y, finalmente, el candidato seleccionado se rectifica mediante una homografía a una plantilla cuadrada canónica.

dor m_i en el sistema de referencia global, y $\mathbf{u}_{i,k} \in \mathbb{R}^2$ representa la posición de esa misma esquina detectada en la imagen. El papel del identificador del marcador es esencial, ya que determina qué cuatro puntos 3D del mapa deben asociarse con las cuatro esquinas 2D detectadas. A partir de las correspondencias, la pose de la cámara se obtiene resolviendo un problema PnP. Este problema estima la transformación que relaciona el sistema de referencia global con el sistema de referencia de la cámara. Se busca la pose ${}^W\mathbf{T}_C \in \text{SE}(3)$, que representa la posición y orientación de la cámara respecto al mundo. La Figura 4 resume los sistemas de referencia utilizados.

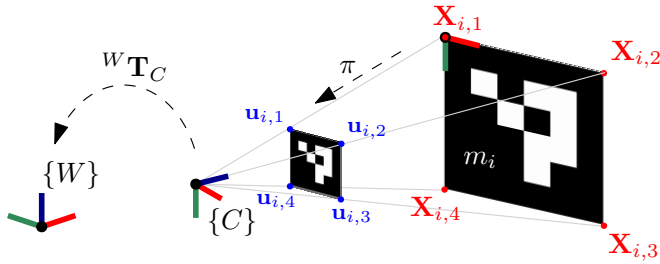


Figura 4: Relación entre los sistemas de referencia del mundo y la cámara, y las correspondencias 2D-3D empleadas por PnP.

La estimación de pose se obtiene minimizando el error de reproyección:

$${}^W\mathbf{T}_C^* = \arg \min_{{}^W\mathbf{T}_C} \sum_{(\mathbf{X}_{i,k}, \mathbf{u}_{i,k}) \in C} \left\| \mathbf{u}_{i,k} - \pi({}^C\mathbf{T}_W \mathbf{X}_{i,k}) \right\|^2, \quad (1)$$

donde ${}^C\mathbf{T}_W = ({}^W\mathbf{T}_C)^{-1}$ y $\pi(\cdot)$ es el modelo de proyección de la cámara, definido a partir de la calibración intrínseca. Para cada esquina 3D $\mathbf{X}_{i,k}$ del mapa, la transformación ${}^C\mathbf{T}_W$ expresa dicho punto en el sistema de referencia de la cámara. A continuación, la función $\pi(\cdot)$ proyecta el punto transformado sobre el plano de imagen. El residuo mide la diferencia entre la posición proyectada y la esquina realmente detectada, $\mathbf{u}_{i,k}$. La pose óptima ${}^W\mathbf{T}_C^*$ es aquella que hace que las proyecciones de las esquinas 3D del mapa coincidan lo mejor posible con las esquinas observadas en la imagen.

La pose obtenida corresponde a la pose de la cámara. En los experimentos presentados en la Sección 3, la transformación rígida entre la cámara y el robot se considera identidad, por lo que la pose estimada de la cámara se usa directamente como pose del robot. No obstante, esta simplificación no es una limitación del método. Si se dispone de una calibración extrínseca entre la cámara y el cuerpo del robot, basta con componer la pose estimada con dicha transformación para obtener la pose del robot en el sistema de referencia global.

3. Resultados experimentales

La evaluación experimental se estructura en tres bloques. En primer lugar, se realizan experimentos preliminares para analizar la calidad de la detección geométrica de los marcadores, es decir, la precisión con la que se localizan sus esquinas en la imagen de eventos. En segundo lugar, se evalúa la identificación de marcadores, midiendo si el código extraído de cada candidato permite recuperar correctamente su identificador. Finalmente, se analiza la estimación de pose de la cámara en secuencias de vuelo real con un robot aéreo de tipo multirrotor, tanto en interior como en exterior.

Los experimentos preliminares de detección e identificación se realizan con secuencias grabadas con una cámara de eventos DAVIS346 (resolución 346×260). Se emplean dos secuencias: *Preliminar 1* y *Preliminar 2*. Ambas corresponden a una cámara movida manualmente frente a un conjunto de marcadores ArUco, pero con distinta velocidad de movimiento. La secuencia *Preliminar 1* presenta un movimiento más lento (0.2 millones de eventos generados por segundo de media) y, por tanto, condiciones más favorables para la acumulación de eventos. La secuencia *Preliminar 2* introduce un movimiento más rápido (~ 1 millón de eventos por segundo).

La estimación de pose se evalúa con secuencias de vuelo de un multirrotor. La plataforma aérea utilizada en los experimentos es un multirrotor basado en un DJI Flamewheel F450. El vehículo incorpora un autopiloto PixRacer y un ordenador embarcado Khadas VIM3. Se consideran dos escenarios. El primero es una zona de vuelo interior con un sistema de captura de movimiento OptiTrack que proporciona *ground truth* de la pose del robot aéreo (véase Figura 5-izquierda). En este escenario se evalúan tres secuencias de vuelo: *Vuelo Interior 1*, *Vuelo Interior 2* y *Vuelo Interior 3*. El segundo escenario corresponde a vuelos exteriores, donde no se dispone de sistema de captura de movimiento (véase Figura 5-derecha). En este caso se emplean las secuencias *Vuelo Exterior 1*, *Vuelo Exterior 2* y *Vuelo Exterior 3*.

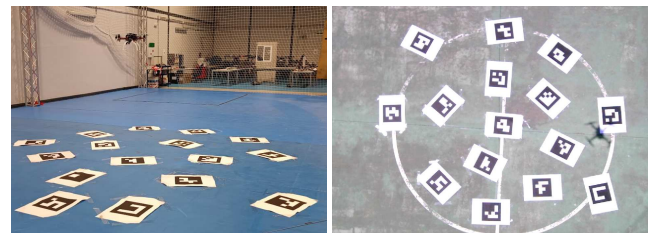


Figura 5: Escenarios experimentales utilizados para la evaluación: interior (izquierda; vista externa) y exterior (derecha; vista desde el robot aéreo).

3.1. Evaluación preliminar de la detección

La primera evaluación analiza la detección de los marcadores. Esta es una evaluación crítica porque la estimación de pose depende directamente de la exactitud en la localización de las esquinas de los marcadores. Aunque un marcador sea identificado correctamente, un error en sus esquinas introduce un error en las correspondencias 2D-3D utilizadas por el estimador PnP. Para evaluar la detección se emplean las secuencias *Preliminar 1* y *Preliminar 2*. Se consideran dos métricas: el error de reproyección medio (RepE), expresado en píxeles, y la intersección sobre unión (IoU) entre el cuadrilátero detectado y el cuadrilátero proyectado desde el mapa.

Secuencia	RepE (px)	IoU (%)
<i>Preliminar 1</i>	0.797	96.01
<i>Preliminar 2</i>	0.796	96.08

Tabla 1: Evaluación de la detección de contornos.

Los resultados de la Tabla 1 muestran que la detección de esquinas es precisa en ambas secuencias. El error medio de reproyección se mantiene por debajo de un píxel. Además, la IoU supera el 95 % en los dos casos, lo que indica que los cuadriláteros detectados se ajustan de forma consistente a la geometría real de los marcadores. Estos resultados validan la primera condición necesaria para la estimación de pose: las esquinas detectadas son suficientemente precisas para ser utilizadas como observaciones 2D en un problema PnP.

3.2. Evaluación preliminar de la identificación

La segunda evaluación analiza la capacidad del método para asignar correctamente el identificador de cada marcador detectado. Esta etapa es crítica para la estimación de pose, ya que el identificador determina qué puntos 3D del mapa deben asociarse con las esquinas 2D observadas. Una detección geoméricamente precisa pero con un identificador incorrecto produciría correspondencias erróneas y, por tanto, una estimación de pose inconsistente. La identificación se evalúa también sobre las secuencias *Preliminar 1* y *Preliminar 2*. Se emplean dos métricas: la exactitud, definida como el porcentaje de marcadores visibles que se identifican correctamente, y la precisión, definida como el porcentaje de identificaciones aceptadas que son correctas.

Secuencia	Exactitud (%)	Precisión (%)
<i>Preliminar 1</i>	96.62	99.93
<i>Preliminar 2</i>	98.63	99.97

Tabla 2: Evaluación de la identificación de marcadores.

Como se observa en la Tabla 2, la identificación es altamente fiable. Los valores de precisión indican que cuando el sistema acepta un marcador como identificado, la probabilidad de que su identificador sea correcto es muy alta.

3.3. Evaluación de la estimación de pose

Una vez evaluadas por separado la detección geométrica y la identificación, se analiza el rendimiento del sistema completo en estimación de pose. En estos experimentos, los marcadores detectados e identificados se usan como referencias conocidas del entorno. Cada marcador aceptado proporciona

cuatro correspondencias entre esquinas 3D del mapa y esquinas 2D observadas en la imagen.

La evaluación interior se realiza con un sistema de captura de movimiento OptiTrack. Se consideran tres secuencias de vuelo del multirrotor. Las trayectorias estimadas se alinean con el *ground truth* mediante una transformación SE(3) obtenida con alineamiento Umeyama antes de calcular el error absoluto de trayectoria (ATE; medio y máximo) y la frecuencia media de salida. La Figura 6 muestra las trayectorias estimadas para la secuencia *Vuelo Interior 1*.

Secuencia	ATE medio/máx. (m)	Frec. (Hz)
<i>Vuelo Interior 1</i>	0.052 / 0.246	142.33
<i>Vuelo Interior 2</i>	0.042 / 0.267	174.77
<i>Vuelo Interior 3</i>	0.041 / 0.293	150.36

Tabla 3: Estimación de pose en vuelos interiores.

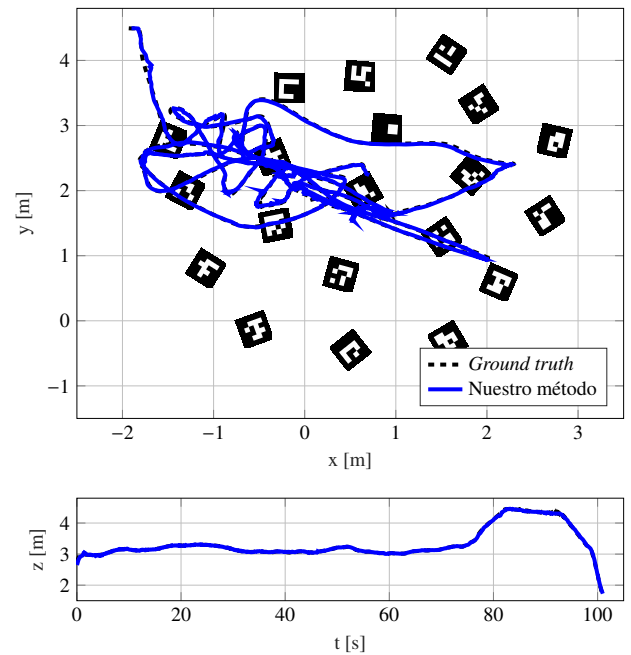


Figura 6: Trayectorias estimadas para la secuencia *Vuelo Interior 1*.

Para contextualizar estos resultados, la Tabla 4 compara el método propuesto con un estimador de pose basado en ArUco convencional y con ORB-SLAM3 (Campos et al. (2021)) en las mismas secuencias de multirrotor del escenario interior utilizando una cámara convencional ELP de 720p (rígidamente acoplada y con un campo de visión similar).

Método	Secuencia	ATE medio/máx. (m)	Frec. (Hz)
Nuestro	<i>V.I.1</i>	0.052 / 0.246	142.33
	<i>V.I.2</i>	0.042 / 0.267	174.77
	<i>V.I.3</i>	0.041 / 0.293	150.36
ArUco+PnP	<i>V.I.1</i>	0.067 / 0.702	43.60
	<i>V.I.2</i>	0.096 / 0.731	31.60
	<i>V.I.3</i>	0.056 / 0.850	30.64
ORB-SLAM3	<i>V.I.1</i>	0.669 / 2.441	47.77
	<i>V.I.2</i>	0.126 / 1.232	42.40
	<i>V.I.3</i>	0.164 / 0.816	48.57

Tabla 4: Comparación de estimación de pose en vuelos interiores. *V.I.1*, *V.I.2* y *V.I.3* corresponden a *Vuelo Interior 1*, *Vuelo Interior 2* y *Vuelo Interior 3*, respectivamente.

La evaluación exterior se realiza con tres vuelos del multirrotor en un entorno no controlado. En este escenario no se dispone de sistema de captura de movimiento, por lo que no existe *ground truth*. La trayectoria basada en eventos se compara con trayectorias de referencia obtenidas mediante ArUco convencional y ORB-SLAM3, alineadas mediante SE(3).

Referencia	Métrica	V.E.1	V.E.2	V.E.3
ArUco+PnP	ATE medio (m)	0.137	0.728	0.674
	ATE máx. (m)	1.309	3.967	2.494
ORB-SLAM3	ATE medio (m)	0.111	0.274	0.789
	ATE máx. (m)	1.357	2.502	3.810
	Frec. (Hz)	105.76	140.47	99.37

Tabla 5: Estimación de pose en vuelos exteriores. V.E.1, V.E.2 y V.E.3 corresponden a *Vuelo Exterior 1*, *Vuelo Exterior 2* y *Vuelo Exterior 3*, respectivamente. Al no existir *ground truth*, el ATE se define aquí como diferencia respecto a la trayectoria de referencia.

Los experimentos en interior y exterior demuestran que el método mantiene un comportamiento consistente en condiciones reales de vuelo. En interior, la comparación con el sistema de captura de movimiento confirma que las trayectorias estimadas siguen el movimiento del multirrotor, mientras que en exterior se observa una estimación de pose estable en condiciones no ideales. Cabe destacar la alta frecuencia de salida del método propuesto, que genera muchas más estimaciones de pose que un método convencional, proporcionando una descripción temporal más densa del movimiento, lo que resulta de gran interés en maniobras rápidas o dinámicas. Todo ello se logra con procesamiento en CPU, sin necesidad de GPU, lo que facilita su integración en plataformas robóticas con recursos limitados.

4. Conclusiones

Este trabajo ha presentado un método de estimación de pose basado en marcadores ArUco y cámaras de eventos. El método utiliza una representación basada en actividad de bordes para estimar la pose mediante un problema PnP a partir de correspondencias 2D-3D. El enfoque permite emplear marcadores convencionales sin reconstruir imágenes de intensidad ni recurrir a marcadores activos. Los experimentos demuestran una localización precisa, una identificación fiable y una estimación de pose consistente. En robots aéreos reales, el método alcanza errores medios del orden de centímetros y frecuencias de salida superiores a 100 Hz. Los resultados confirman que la combinación de cámaras de eventos y marcadores fiduciales pasivos constituye una solución adecuada para estimación de pose en robótica. Como trabajo futuro, se propone integrar esta estimación en sistemas de odometría visual-inercial y navegación autónoma.

Agradecimientos

Este trabajo ha sido realizado con financiación del proyecto ORIGAMI del European Defence Fund (EDF)

Referencias

Campos, C., Elvira, R., Rodríguez, J. J. G., M. Montiel, J. M., D. Tardós, J., 2021. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Trans. Robot.* 37 (6), 1874–1890. DOI: 10.1109/TR0.2021.3075644

Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica* 10 (2), 112–122. DOI: 10.3138/FM57-6770-U75U-7727

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R., 2016. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recognit.* 51, 481–491. DOI: 10.1016/j.patcog.2015.09.023

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., Marín-Jiménez, M. J., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* 47 (6), 2280–2292. DOI: 10.1016/j.patcog.2014.01.005

Garrido-Jurado, S., Romero-Ramírez, F. J., Muñoz-Salinas, R., 2026. ArUco nano: A simpler, faster, and more reliable fiducial marker detector. *SoftwareX* 34, 102690. DOI: 10.1016/j.softx.2026.102690

Huang, Y., Hou, T., Zhang, S., Pei, X., Yu, R., Liu, H., 2023. A novel ArUco marker for event cameras. In: *Int. Conf. Front. Robot. Softw. Eng.* pp. 271–277. DOI: 10.1109/FRSE58934.2023.00044

Jurado-Rodríguez, D., Muñoz-Salinas, R., Garrido-Jurado, S., Medina-Carnicer, R., 2023. Planar fiducial markers: A comparative study. *Virtual Real.* 27 (3), 1733–1749. DOI: 10.1007/s10055-023-00772-5

Kalaitzakis, M., Cain, B., Carroll, S., Ambros, A., Whitehead, C., Vitzilaios, N., 2021. Fiducial markers for pose estimation: Overview, applications and experimental comparison of the artag, AprilTag, ArUco and stag markers. *J. Intell. Robot. Syst.* 101 (4), 71. DOI: 10.1007/s10846-020-01307-9

Kitade, T., Yamada, W., Ochiai, K., Imai, M., 2024. Bicode: A hybrid binocular marker system for event cameras. In: *IEEE Int. Conf. Robot. Autom.* pp. 2939–2945. DOI: 10.1109/ICRA57147.2024.10611033

Loch, A., Haessig, G., Vincze, M., 2023. Event-based high-speed low-latency fiducial marker tracking. In: *Int. Conf. Mach. Vis.* p. 30. DOI: 10.1117/12.2679683

Mai, N. T., Komatsu, R., Asama, H., Yamashita, A., 2023. Pose estimation for event camera using charuco board based on image reconstruction. In: *IEEE SICE Int. Symp. Syst. Integr.* pp. 1–6. DOI: 10.1109/SII55687.2023.10039168

Muñoz-Salinas, R., Marín-Jiménez, M. J., Medina-Carnicer, R., 2019. SPM-SLAM: Simultaneous localization and mapping with squared planar markers. *Pattern Recognit.* 86, 156–171. DOI: 10.1016/j.patcog.2018.09.003

Muñoz-Salinas, R., Medina-Carnicer, R., 2020. UcoSLAM: Simultaneous localization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognit.* 101, 107193. DOI: 10.1016/j.patcog.2019.107193

Romero-Ramírez, F. J., Muñoz-Salinas, R., Medina-Carnicer, R., 2018. Speeded up detection of squared fiducial markers. *Image Vis. Comput.* 76, 38–47. DOI: 10.1016/j.imavis.2018.05.004

Sarmadi, H., Munoz-Salinas, R., Olivares-Mendez, M. A., Medina-Carnicer, R., 2021. Detection of binary square fiducial markers using an event camera. *IEEE Access* 9, 27813–27826. DOI: 10.1109/ACCESS.2021.3058423

Suzuki, S., Abe, K., 1985. Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* 30 (1), 32–46. DOI: 10.1016/0734-189X(85)90016-7

Tapia, R., Martínez-de Dios, J. R., Ollero, A., 2024. eFFT: An event-based method for the efficient computation of exact Fourier transforms. *IEEE Trans. Pattern Anal. Mach. Intell.* 46 (12), 9630–9647. DOI: 10.1109/TPAMI.2024.3422209

Tapia, R., Martínez-de Dios, J. R., Ollero, A., 2026. eFFT-C++: An open-source implementation of the event-based fast Fourier transform. *J. Open Res. Softw.* 14, 27. DOI: 10.5334/jors.642

Tofghi, N. J., Robic, M., Morbidi, F., Vasseur, P., 2025. A survey on event-based optical marker systems. DOI: 10.48550/ARXIV.2504.20736

Zhang, S., Huang, Y., Pei, X., Lin, H., Zheng, W., Wang, W., Hou, T., 2023. An improved LED ArUco-marker detection method for event camera. In: *Commun. Netw. Inf. Syst. Vol. 1839*. pp. 47–57. DOI: 10.1007/978-981-99-3581-9_3