

Jornadas de Automática

Estimación de pose por filtrado complementario de marcadores y odometría

Victor Quesada Conejero^{a,b,*}, Fran Real^b, Sergio Garrido-Jurado^b

^aDpto. de Ingeniería Eléctrica y Automática, Universidad de Córdoba, Avda. de Medina Azahara, 5, 14071, Córdoba, España.

^bSeabery R&D, Doctor Emilio Haya Prats 15, Huelva, 21005, Spain.

To cite this article: Quesada, V., Real, F., Garrido-Jurado, S. 2026. Pose Estimation via Complementary Filtering of Markers and Odometry. Jornadas de Automática, 47. <https://doi.org/10.17979/ja-cea.2026.47.13724>

Resumen

La estimación precisa de la pose de objetos rígidos constituye un requisito fundamental en robótica guiada por visión. En configuraciones *eye-in-hand* las estimaciones obtenidas mediante *Perspective-n-Point* (PnP) a partir de marcadores visuales presentan ruido entre fotogramas. Se propone un método recursivo que combina observaciones de marcadores ArUco con la odometría del manipulador. La odometría reexpresa las observaciones en el sistema de coordenadas de la base del robot, donde el objeto permanece estático, y sobre ellas se aplica una media móvil exponencial (EWMA) robusta formulada en el grupo de Lie SE(3). El método incorpora un peso de Huber dependiente del estado y un peso derivado del error de reproyección del PnP para modelar la fiabilidad de cada observación. La validación experimental emplea un brazo robótico con cámara monocular que observa un objeto estático con marcadores ArUco. Los resultados muestran una reducción significativa del error relativo de pose y del *jitter* temporal frente a PnP directo y variantes EWMA sin ponderación robusta.

Palabras clave: Filtrado y suavizado, Estimación y filtrado, Percepción y sensores en robótica, Estimación robusta, Cooperación e integración humano-máquina.

Pose Estimation via Complementary Filtering of Markers and Odometry

Abstract

Accurate estimation of rigid object pose is a fundamental requirement in vision-based robotics. In *eye-in-hand* configurations, estimates obtained through *Perspective-n-Point* (PnP) from visual markers exhibit frame-to-frame noise. We propose a recursive method that combines visual observations of ArUco markers with the robot odometry. The odometry transforms the observations into the robot base coordinate frame, where the object remains static, and a robust exponentially weighted moving average (EWMA) formulated on the Lie group SE(3) is applied to them. The method incorporates a state-dependent Huber weighting term and a reprojection error-based weight from PnP to model the reliability of each observation. Experimental validation is carried out using a robotic arm with a monocular camera observing a static object with ArUco markers. The results show a significant reduction in relative pose error and temporal jitter compared to direct PnP and EWMA variants without robust weighting.

Keywords: Filtering and smoothing, Estimation and filtering, Robot perception and sensing, Robust estimation, Human machine cooperation & integration.

1. Introducción

La estimación precisa de la pose de objetos rígidos constituye un problema central en robótica guiada por visión (Chau-

mette and Hutchinson, 2006), con aplicaciones en manipulación (Labbé et al., 2022), realidad aumentada (Marchand et al., 2016) y sistemas de interacción humano-máquina (Goochrich and Schultz, 2007). En estos escenarios, la calidad de

*Autor para correspondencia: z32qucov@uco.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

la pose estimada repercute directamente en la estabilidad del control y en la consistencia perceptual del sistema.

Entre los enfoques más empleados para la estimación de pose destacan los métodos basados en Perspective-n-Point (PnP) (Lepetit et al., 2009; Terzakis and Lourakis, 2020), que permiten recuperar la transformación cámara-objeto a partir de correspondencias 2D-3D. Sin embargo, en aplicaciones reales, dichas estimaciones suelen verse afectadas por errores de detección, oclusiones parciales y configuraciones geométricas degeneradas. Como consecuencia, aparecen variaciones fotograma a fotograma que degradan la estabilidad temporal de la pose estimada, incluso cuando el error absoluto medio permanece bajo.

Para mitigar este problema, se han propuesto distintas estrategias de filtrado temporal, desde filtros exponenciales simples hasta enfoques probabilísticos más complejos, como filtros de Kalman o métodos basados en optimización (Kam et al., 2018; Kim et al., 2024). No obstante, estos métodos suelen requerir modelos dinámicos explícitos y un ajuste cuidadoso de sus parámetros, como matrices de covarianza o ruidos de proceso. Esto introduce una complejidad adicional que puede resultar innecesaria cuando las variaciones temporales de la pose estimada provienen principalmente del ruido de detección y no del movimiento real del objeto.

En este trabajo se propone un enfoque alternativo basado en una media móvil exponencial, o EWMA, formulada directamente en el espacio de Lie $\mathbf{SE}(3)$. El método incorpora dos mecanismos de ponderación complementarios, que permiten adaptar la contribución de cada observación en función tanto de su consistencia geométrica como de su fiabilidad intrínseca, estimada a partir del propio proceso de PnP.

El sistema se evalúa en un entorno robótico real, con una cámara montada en el efector final que observa un objeto estático cubierto con marcadores ArUco (Garrido-Jurado et al., 2014; García-Ruiz et al., 2023). Los resultados experimentales muestran una reducción significativa del error relativo de pose, lo que evidencia una mejora de la estabilidad temporal frente a enfoques basados en estimaciones PnP directas o en filtrado exponencial sin ponderación robusta.

Las contribuciones de este trabajo son las siguientes:

1. La formulación del problema de filtrado de pose en el sistema de referencia de la base del robot, donde el objeto permanece estático, desacoplando el movimiento de la cámara del ruido de estimación.
2. La propuesta de un filtro EWMA robusto formulado sobre $\mathbf{SE}(3)$ para la estimación de la pose.
3. La combinación de dos criterios complementarios de robustez sobre EWMA: un peso dependiente del estado del filtro y un peso derivado del error de reproyección dependiente de la fiabilidad de cada observación.
4. La validación experimental en un sistema robótico real.

El resto del artículo se organiza de la siguiente manera: la Sección 2 describe el planteamiento del problema, la Sección 3 presenta el método propuesto, la Sección 4 detalla la validación experimental y, finalmente, la Sección 5 recoge las conclusiones.

2. Planteamiento del problema

El sistema considerado, ilustrado en la Figura 1, está formado por un brazo robótico con una cámara montada en el efector final y un objeto estático cubierto con un mapa de marcadores ArUco. En cada instante k , la cámara proporciona un fotograma f_k , mientras que el controlador del robot entrega la pose del efector ${}^b\mathbf{T}_{g,k} \in \mathbf{SE}(3)$, expresada en el sistema de coordenadas de la base del robot. Además, mediante un proceso previo de calibración Hand-eye (Tsai and Lenz, 1989), se conoce la transformación fija ${}^c\mathbf{T}_g \in \mathbf{SE}(3)$ desde el efector a la cámara.

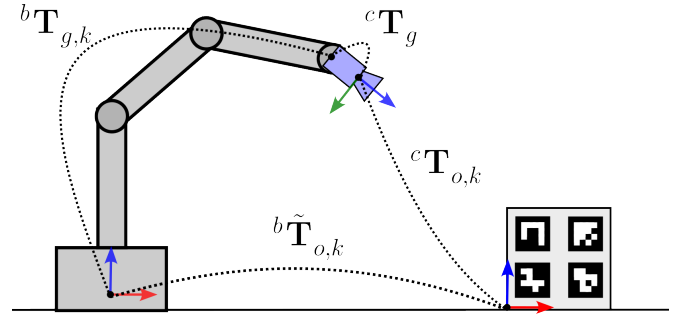


Figura 1: Esquema del sistema: brazo robótico con cámara en el efector final observando un objeto estático con un mapa de marcadores ArUco. Se indican los sistemas de coordenadas de la base del robot, el efector final, la cámara y el objeto, junto con las transformaciones ${}^b\mathbf{T}_{g,k}$ (odometría del robot), ${}^c\mathbf{T}_g$ (calibración hand-eye), ${}^c\mathbf{T}_{o,k}$ (estimación PnP) y ${}^b\tilde{\mathbf{T}}_{o,k}$ (relación entre la base del robot y el objeto).

La transformación entre la cámara y el objeto se obtiene mediante un proceso de estimación de pose basado en marcadores ArUco. En primer lugar, se detectan los marcadores visibles en el fotograma y se establecen las correspondencias 2D-3D entre las esquinas detectadas y sus posiciones conocidas en el modelo del objeto. A continuación, dichas esquinas se transforman a coordenadas de imagen normalizadas empleando los parámetros intrínsecos y el modelo de distorsión de la cámara. Con estas correspondencias se resuelve el problema PnP, obteniendo la transformación cámara-objeto ${}^c\mathbf{T}_{o,k}$ y la raíz del error cuadrático medio de reproyección (RMSE) $r_{n,k}$, expresado en coordenadas normalizadas.

Finalmente, la pose del objeto respecto a la base del robot se obtiene mediante la concatenación de transformaciones:

$${}^b\tilde{\mathbf{T}}_{o,k} = {}^b\mathbf{T}_{g,k} ({}^c\mathbf{T}_g)^{-1} {}^c\mathbf{T}_{o,k} \in \mathbf{SE}(3). \quad (1)$$

El objetivo del problema es obtener una estimación temporalmente estable de la pose cámara-objeto ${}^c\mathbf{T}_{o,k}$, reduciendo las fluctuaciones entre fotogramas introducidas por el proceso de detección y resolución PnP.

Dado que el objeto permanece fijo respecto a la base del robot, resulta más conveniente reformular el problema en dicho sistema de coordenadas, donde la pose objeto-base ideal corresponde a una única transformación constante $\mathbf{T}^* \in \mathbf{SE}(3)$. En consecuencia, cada observación ${}^b\tilde{\mathbf{T}}_{o,k}$ puede interpretarse como una medida ruidosa de una misma pose fija, permitiendo formular el problema como una estimación robusta sobre $\mathbf{SE}(3)$. Esta formulación evita filtrar directamente sobre ${}^c\mathbf{T}_{o,k}$, donde las observaciones incorporarían además

el movimiento propio de la cámara. La ecuación (1) permite incorporar explícitamente la cinemática del robot y expresar todas las observaciones en un sistema de referencia en el que la señal de interés permanece estática, simplificando así el diseño del filtro.

En las siguientes secciones, los superíndices y subíndices asociados a la transformación base–objeto se omitirán por brevedad. En particular, se utilizará $\tilde{\mathbf{T}}_k$ para denotar la muestra obtenida en (1), y $\hat{\mathbf{T}}_k$ para representar la estimación filtrada correspondiente.

Cada muestra $\tilde{\mathbf{T}}_k$, junto con el RMSE de reproyección normalizado $r_{n,k}$ y la profundidad cámara–objeto, definida como

$$Z_k = \|\mathbf{t}_{o,k}^c\|, \quad (2)$$

constituyen las entradas del filtro presentado en la Sección 3.

3. Método

3.1. Filtro EWMA en SE(3)

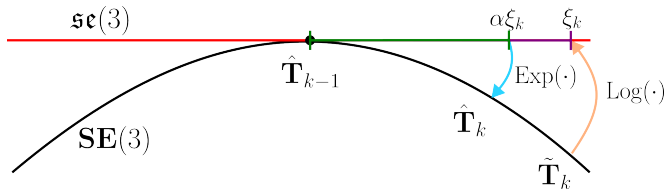


Figura 2: Interpretación geométrica de la actualización EWMA sobre $\mathbf{SE}(3)$. El residuo entre la estimación previa $\hat{\mathbf{T}}_{k-1}$ y la observación $\tilde{\mathbf{T}}_k$ se expresa en el espacio tangente mediante $\text{Log}(\cdot)$, se escala por el factor α , y se reinyecta sobre el grupo mediante $\text{Exp}(\cdot)$ para obtener la nueva estimación $\hat{\mathbf{T}}_k$.

Bajo la hipótesis de movimiento cuasi-estático, el filtro mantiene una única estimación de la pose $\hat{\mathbf{T}}_k$, que se actualiza recursivamente con cada nuevo fotograma. Dada la muestra observada $\tilde{\mathbf{T}}_k$ y la estimación previa $\hat{\mathbf{T}}_{k-1}$, se calcula el residuo en el espacio tangente (Solà et al., 2018), junto con su norma, como

$$\xi_k = \text{Log}(\hat{\mathbf{T}}_{k-1}^{-1} \tilde{\mathbf{T}}_k), \quad v_k = \|\xi_k\|_2, \quad (3)$$

donde $\text{Log}(\cdot)$ representa el mapa logarítmico desde $\mathbf{SE}(3)$ hacia su álgebra de Lie $\mathfrak{se}(3)$, y $\|\cdot\|_2$ denota la norma euclídea del vector tangente asociado, definido como la concatenación de los vectores de traslación y rotación.

La actualización del estado se define como una media móvil exponencial, o EWMA, sobre $\mathbf{SE}(3)$, aplicada en el espacio tangente:

$$\hat{\mathbf{T}}_k = \hat{\mathbf{T}}_{k-1} \text{Exp}(\alpha \xi_k), \quad (4)$$

donde $\text{Exp}(\cdot)$ representa el mapa exponencial desde el álgebra de Lie $\mathfrak{se}(3)$ al grupo de Lie $\mathbf{SE}(3)$, y $\alpha \in (0, 1]$ es la tasa de actualización de la EWMA. El parámetro α puede seleccionarse a partir de una longitud de ventana equivalente W mediante la relación $\alpha = 2/(W + 1)$.

La Figura 2 ilustra el principio geométrico de la actualización: el residuo entre la estimación previa y la observación se expresa en el espacio tangente mediante el mapa logarítmico,

se atenúa por el factor α , y posteriormente se reinyecta sobre $\mathbf{SE}(3)$ mediante el mapa exponencial.

Esta recursión asigna el mismo peso a todas las muestras, por lo que un único fotograma con un residuo elevado puede desplazar de forma apreciable la estimación $\hat{\mathbf{T}}$. Las subsecciones siguientes introducen dos modificaciones en la actualización del estado para limitar la influencia de este tipo de muestras: una basada en la distancia al estado actual del filtro, mediante un peso de Huber, descrita en la Sección 3.2, y otra basada en la calidad intrínseca de la estimación PnP, presentada en la Sección 3.3.

El método asume que la pose objeto–base permanece constante durante la secuencia, de modo que las variaciones entre observaciones se atribuyen principalmente al ruido del proceso PnP. Ante cambios reales de pose, el filtro sigue adaptándose de forma recursiva, aunque introduce un retardo temporal en función de α .

3.2. Peso de Huber

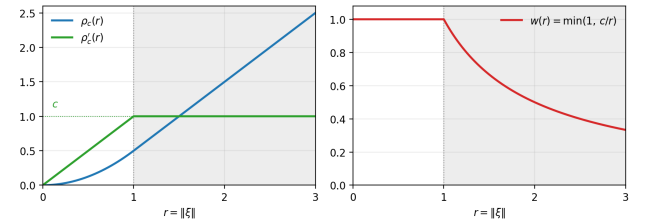


Figura 3: Función de Huber y peso asociado, representados con $c = 1$. Izquierda: pérdida $\rho_c(r)$ (azul) y su derivada $\rho'_c(r)$ (verde), con transición entre los regímenes cuadrático y lineal en $r = c$. Derecha: peso $w(r) = \min(1, c/r)$. La región sombreada ($r > c$) corresponde al régimen lineal, en el que el peso decrece proporcionalmente a c/r .

Para mejorar la robustez frente a observaciones inconsistentes, se ponderan las muestras según su coherencia con el estado actual del filtro mediante la función de Huber (Huber, 1964). Esta función combina un comportamiento cuadrático para residuos pequeños con un crecimiento lineal para residuos grandes:

$$\rho_c(r) = \begin{cases} \frac{1}{2} r^2 & |r| \leq c, \\ c|r| - \frac{1}{2} c^2 & |r| > c, \end{cases} \quad (5)$$

donde el parámetro c establece la transición entre ambos regímenes (Figura 3 izquierda). El tramo cuadrático reproduce el comportamiento del estimador de mínimos cuadrados bajo ruido gaussiano, mientras que el tramo lineal acota el gradiente, $|\rho'_c(r)| \leq c$, limitando así la influencia máxima que un outlier puede ejercer sobre la estimación.

En la media ponderada no se emplea directamente la función de Huber, sino su derivada ρ'_c . Al factorizarla como

$$\rho'_c(r) = w(r) r \quad (6)$$

se obtiene un peso adimensional por muestra,

$$w(r) := \rho'_c(r)/r \in (0, 1], \quad (7)$$

que modula el tamaño del paso de la EWMA sin alterar su estructura original. Aplicando esta expresión a la función de pérdida de Huber en (5), y evaluándola sobre el residuo tangente v_k definido en (3), se obtiene

$$w_k = \min(1, c/v_k), \quad (8)$$

por lo que la actualización del filtro queda definida como

$$\hat{\mathbf{T}}_k = \hat{\mathbf{T}}_{k-1} \text{Exp}(\alpha w_k \xi_k). \quad (9)$$

La Figura 3 (panel derecho) ilustra el comportamiento de w_k : las muestras con residuo tangente pequeño ($v_k \leq c$) no se penalizan, mientras que aquellas con $v_k > c$ se atenúan proporcionalmente a c/v_k . De este modo, la magnitud del incremento tangente queda limitada a un máximo de αc , con independencia del tamaño del residuo.

El umbral c se aplica sobre la norma $\|\xi_k\|$, que combina traslación (m) y rotación (rad) en una única magnitud. Esta simplificación renuncia a la invarianza dimensional en favor de un único parámetro robusto, calibrado empíricamente.

El peso w_k constituye, por tanto, un mecanismo dependiente del estado: resulta eficaz frente a observaciones que presentan grandes desviaciones respecto a la estimación actual, pero no distingue entre medidas intrínsecamente fiables y aquellas que, por azar, se encuentran próximas al estado del filtro.

3.3. Peso en función del error de reproyección

Una estrategia complementaria para mejorar la robustez consiste en aprovechar la calidad autoestimada por el propio algoritmo PnP. Junto con la pose estimada, el solver proporciona la raíz del error cuadrático medio de reproyección (RMSE) $r_{n,k}$, expresado en las coordenadas de imagen normalizadas empleadas durante el proceso de estimación, tal como se describe en la Sec. 2.

El valor $r_{n,k}$ es un residuo adimensional que mide la discrepancia angular entre el rayo óptico observado y el rayo obtenido al proyectar el modelo con la estimación actual. Al trabajar con coordenadas de imagen normalizadas, esta magnitud no depende explícitamente de los parámetros intrínsecos de la cámara.

En primera aproximación, dicho error angular se propaga a un error tridimensional proporcional a la profundidad del objeto detectado:

$$\sigma_{3D,k} = r_{n,k} Z_k, \quad (10)$$

donde Z_k representa la profundidad media del objeto respecto a la cámara. Esta relación constituye una aproximación de primer orden válida bajo la hipótesis de pequeños ángulos.

A partir de esta magnitud se define un peso basado en varianza inversa, referido a una escala de referencia σ_{ref} :

$$q_k = \frac{\sigma_{\text{ref}}^2}{\sigma_{\text{ref}}^2 + \sigma_{3D,k}^2}, \quad (11)$$

de modo que q_k queda acotado en el intervalo (0, 1].

Este peso actúa como un factor de fiabilidad asociado a la calidad de la observación. Las muestras con $\sigma_{3D,k} \ll \sigma_{\text{ref}}$ reciben un peso cercano a la unidad, $q_k \approx 1$, mientras que aquellas con $\sigma_{3D,k} \gg \sigma_{\text{ref}}$ ven reducida su influencia aproximadamente como $(\sigma_{\text{ref}}/\sigma_{3D,k})^2$.

La escala σ_{ref} es el único parámetro asociado a este peso que debe ajustarse, y representa la desviación estándar máxima considerada aceptable en el espacio tridimensional.

A diferencia del peso w_k , el peso q_k no depende del estado del filtro, sino únicamente de la calidad estimada por el solver PnP. Por tanto, una observación puede ser coherente con el estado actual del filtro ($w_k = 1$) y, aun así, ser intrínsecamente poco fiable, por ejemplo debido a una configuración degenerada del PnP ($q_k \rightarrow 0$). Del mismo modo, una observación puede presentar un error de reproyección bajo y, sin embargo, ser inconsistente con la estimación filtrada.

3.4. Combinación de pesos

Los pesos w_k y q_k capturan fuentes de robustez complementarias. El primero atenúa las observaciones que son inconsistentes con el estado estimado, mientras que el segundo evalúa la fiabilidad intrínseca de cada observación a partir de la calidad de la estimación PnP.

Ambos mecanismos pueden integrarse de forma natural en una única actualización:

$$\hat{\mathbf{T}}_k = \hat{\mathbf{T}}_{k-1} \text{Exp}(\alpha w_k q_k \xi_k), \quad (12)$$

donde el tamaño efectivo del paso queda modulado simultáneamente por la consistencia geométrica de la observación y por la calidad de la pose estimada mediante PnP.

Esta formulación unificada permite limitar la influencia tanto de outliers geométricos como de observaciones intrínsecamente poco fiables, manteniendo al mismo tiempo la estructura recursiva y sencilla del filtro EWMA.

4. Experimentación

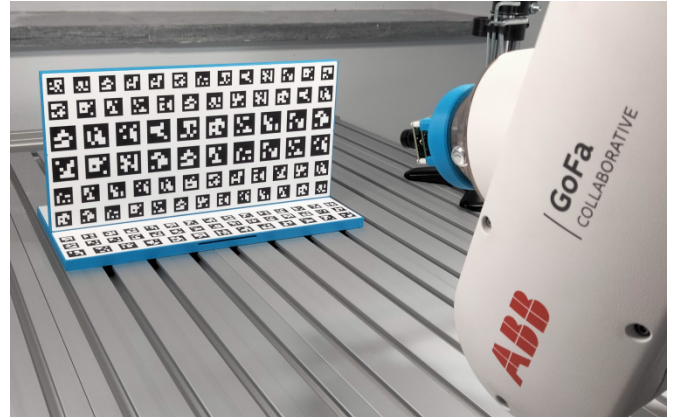


Figura 4: Imagen del equipamiento utilizado para la experimentación.

La Figura 4 muestra el sistema experimental empleado para validar el método propuesto. El montaje está formado por un brazo robótico ABB GoFa 5 CRB 15000, que proporciona información de odometría a una tasa de 150 Hz. En el efector final se instaló una cámara calibrada See3CAM_CU30, capaz de adquirir imágenes con resolución 1080p a 30 fotogramas por segundo, equipada con una óptica de 80° de campo de visión horizontal (HFOV). El objeto utilizado está compuesto por dos planos perpendiculares cubiertos por un total de 113 marcadores ArUco.

Se realizaron ocho grabaciones de movimiento sobre el mismo objeto, con profundidades medias cámara–objeto comprendidas entre 0,25 y 0,65 m. Cada grabación tiene una duración media de 450 fotogramas.

Para la evaluación se reportan el error absoluto de trayectoria (ATE) y el error relativo de pose (RPE) (Zhang and Scaramuzza, 2018), considerando tanto las componentes de traslación como de rotación. Para cada fotograma k , se definen los residuos en el espacio tangente como

$$\xi_k^{\text{ATE}} = \text{Log}(\mathbf{T}_{\text{GT}}^{-1} \hat{\mathbf{T}}_k), \quad \xi_k^{\text{RPE}} = \text{Log}(\hat{\mathbf{T}}_{k-1}^{-1} \hat{\mathbf{T}}_k), \quad (13)$$

y se separan en sus componentes de traslación y rotación, $\xi_k = (\rho_k, \phi_k)$, con $\rho_k \in \mathbb{R}^3$ y $\phi_k \in \mathbb{R}^3$. A partir de estas componentes, se definen cuatro métricas escalares por fotograma como las normas

$$e_{t,k}^{\text{ATE}} = \|\rho_k^{\text{ATE}}\|, \quad e_{r,k}^{\text{ATE}} = \|\phi_k^{\text{ATE}}\|, \quad e_{t,k}^{\text{RPE}} = \|\rho_k^{\text{RPE}}\|, \quad e_{r,k}^{\text{RPE}} = \|\phi_k^{\text{RPE}}\|. \quad (14)$$

El ATE compara la estimación con la referencia $\mathbf{T}_{\text{GT}}$ y mide la desviación absoluta respecto a la pose verdadera del objeto. Por su parte, el RPE mide el salto entre estimaciones consecutivas y permite cuantificar el jitter fotograma a fotograma, de forma independiente de la referencia.

Dado que el objeto permanece físicamente fijo durante toda la sesión, y que el único elemento en movimiento es la cámara montada en el efector del robot, se construye una única referencia común para las ocho grabaciones. Esta referencia se obtiene como una media robusta de Karcher (Karcher, 1977; Hartley et al., 2013) por lotes, calculada sobre las muestras brutas concatenadas de todos los registros. El procedimiento se resuelve de forma iterativa: en cada paso, las muestras se reponderan mediante su peso de Huber respecto al estimador actual y el centroide se actualiza como un promedio ponderado en el espacio tangente, hasta alcanzar la convergencia. La Tabla 1 resume los valores utilizados en la experimentación.

Tabla 1: Parámetros del filtro.

parámetro	símbolo	valor
Tasa EWMA	α	0,02
Umbral Huber	c	50 (mm $\cup$ mrad)
Escala de referencia	σ_{ref}	0,3 mm

4.1. Ablación de componentes del filtro

Para verificar la contribución de cada componente del filtro propuesto, se comparan cinco configuraciones: PnP directo, sin filtrado; EWMA simple, sin ponderación robusta; EWMA+Huber, que incorpora únicamente la ponderación basada en la distancia tangente; EWMA+RepError, que emplea solo la ponderación asociada a la calidad del PnP; y la propuesta completa, EWMA+Huber+RepError, que combina ambos pesos según (12).

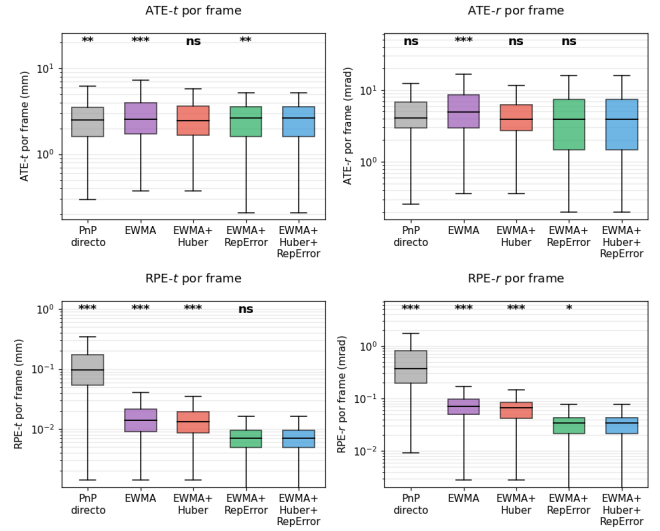


Figura 5: Distribución por fotograma del error de las cuatro métricas (traslación y rotación de ATE y RPE), agregando los fotogramas de las 8 grabaciones. Marcas superiores: significancia del test de Wilcoxon de rangos con signo (pares por fotograma) frente a la propuesta (EWMA+Huber+RepError).

La Figura 5 muestra la distribución por fotograma de los cuatro errores escalares definidos en (14), agregando los fotogramas de las ocho grabaciones en un único conjunto de datos por configuración. Los marcadores superiores indican la significancia de la prueba de Wilcoxon de rangos con signo, aplicada frame a frame entre cada baseline y la propuesta: *** $p < .001$; ** $p < .01$; * $p < .05$; y ns, no significativo.

En términos de RPE, la propuesta reduce claramente el jitter. Frente a PnP directo, EWMA y EWMA+Huber, el método propuesto obtiene mejoras significativas en ambas componentes, con valores estadísticos normalizados de Wilcoxon Z entre 46,4 y 49,2 ($p < .001$). En comparación con EWMA+RepError, la diferencia es más moderada: ambos métodos resultan indistinguibles en RPE-t ($Z = 0,89$, ns), mientras que la propuesta muestra una mejora marginal en RPE-r ($Z = 1,97$, $p < .05$). Estos resultados indican que el peso basado en el error de reproyección explica la mayor parte de la reducción del jitter, mientras que la incorporación adicional del peso de Huber aporta solo una mejora marginal en este régimen.

En cuanto al ATE, los resultados apuntan a un régimen saturado por la calibración. Los filtros que incorporan ponderación robusta, ya sea mediante Huber, error de reproyección o la combinación de ambos, producen errores ATE estadísticamente equivalentes, situados alrededor de $\sim 2,7$ mm en traslación según la Figura 5. La única diferencia estadísticamente detectable entre filtros robustos es marginal, correspondiente a EWMA+RepError frente a la propuesta en ATE-t ($Z = 3,08$, $p < .01$), mientras que en ATE-r los tres métodos resultan indistinguibles. Por el contrario, la EWMA simple sí presenta un ATE significativamente peor en traslación ($Z = 8,48$, $p < .001$).

Finalmente, aunque la diferencia entre EWMA+RepError y EWMA+Huber+RepError es marginal en todas las métricas evaluadas, la inclusión del peso de Huber aporta una propiedad adicional relevante: la influencia acotada frente a observa-

ciones atípicas. En particular, una detección con pocos marcadores puede producir una pose incorrecta y, aun así, presentar un error de reproyección bajo.

5. Conclusiones

Este trabajo aborda la inestabilidad temporal en la estimación de pose rígida basada en PnP en un sistema de visión montado en un brazo robótico, con el objetivo de obtener estimaciones de pose estables en el tiempo a partir de observaciones ruidosas, sin recurrir a modelos dinámicos explícitos ni a optimización por lotes. Para ello, se propone un método recursivo de filtrado robusto en $SE(3)$, basado en una media móvil exponencial (EWMA) formulada sobre el grupo de Lie mediante su espacio tangente.

El método incorpora dos mecanismos de ponderación complementarios: un peso robusto de Huber dependiente del estado, que limita la influencia de observaciones geométricamente inconsistentes, y un peso basado en el RMSE de reproyección normalizado, que modela la fiabilidad intrínseca de cada estimación PnP.

Los resultados experimentales muestran que el filtrado propuesto reduce significativamente el error relativo de pose (RPE), especialmente en presencia de ruido fotograma a fotograma. La mayor parte de la reducción del jitter se debe a la ponderación basada en el error de reproyección, mientras que la inclusión adicional del peso de Huber aporta una propiedad de influencia acotada frente a observaciones atípicas, sin degradar el rendimiento nominal.

En términos de error absoluto de trayectoria (ATE), los distintos esquemas robustos evaluados presentan un comportamiento estadísticamente equivalente, con errores de traslación estabilizados en torno a $\sim 2,7$ mm. Este resultado sugiere que el límite de precisión observado está dominado por errores sistemáticos externos al filtro, principalmente asociados a la calibración geométrica del sistema, más que por la estrategia de filtrado empleada.

El método conserva una estructura recursiva simple, computacionalmente eficiente y fácilmente integrable en sistemas existentes, sin requerir modelos dinámicos explícitos ni optimización por lotes. Esta característica lo hace adecuado para aplicaciones en línea en las que la estabilidad temporal de la estimación resulta prioritaria frente a la minimización de la latencia.

Como trabajo futuro, se plantea extender el método a escenarios dinámicos en los que el objeto presente un movimiento lento pero sostenido, así como incorporar modelos explícitos de incertidumbre que permitan calibrar automáticamente los parámetros del filtro. Otra línea de trabajo consiste en integrar estrategias de recalibración continua o de adaptación de la transformación Hand-eye, con el objetivo de mitigar errores sistemáticos.

Agradecimientos

Esta publicación forma parte de la subvención DIN2024-013345, financiada por el MICIU / AEI / 10.13039/501100011033.

Referencias

- Chaumette, F., Hutchinson, S., 2006. Visual servo control. i. basic approaches. *IEEE Robotics & Automation Magazine* 13 (4), 82–90. DOI: 10.1109/MRA.2006.250573
- García-Ruiz, P., Romero-Ramírez, F. J., Muñoz-Salinas, R., Marín-Jiménez, M. J., Medina-Carnicer, R., 2023. Fiducial objects: Custom design and evaluation. *Sensors* 23 (24). DOI: 10.3390/s23249649
- Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Marín-Jiménez, M., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47 (6), 2280–2292. DOI: 10.1016/j.patcog.2014.01.005
- Goodrich, M. A., Schultz, A. C., Jan. 2007. Human-robot interaction: a survey. *Found. Trends Hum.-Comput. Interact.* 1 (3), 203–275. DOI: 10.1561/11000000005
- Hartley, R., Trunf, J., Dai, Y., Li, H., 2013. Rotation averaging. *International Journal of Computer Vision* 103 (3), 267–305. DOI: 10.1007/s11263-012-0601-0
- Huber, P. J., 1964. Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35 (1), 73–101. DOI: 10.1214/aoms/1177703732
- Kam, H. C., Yu, Y. K., Wong, K. H., 2018. An improvement on ArUco marker for pose tracking using kalman filter. In: 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). pp. 65–69. DOI: 10.1109/SNPD.2018.8441049
- Karcher, H., 1977. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics* 30 (5), 509–541. DOI: 10.1002/cpa.3160300502
- Kim, D., Bong, J. H., Jeong, S., 2024. Enhancing pose estimation using multiple graphical markers with spatial and temporal outlier detection. *Applied Sciences* 14 (22). DOI: 10.3390/app142210225
- Labbé, Y., Manuelli, L., Mousavian, A., Tyree, S., Birchfield, S., Tremblay, J., Carpentier, J., Aubry, M., Fox, D., Sivic, J., 2022. Megapose: 6d pose estimation of novel objects via render & compare. In: *Proceedings of the 6th Conference on Robot Learning (CoRL)*.
- Lepetit, V., Moreno-Noguer, F., Fua, P., Feb. 2009. Epnp: An accurate o(n) solution to the pnp problem. *Int. J. Comput. Vision* 81 (2), 155–166. DOI: 10.1007/s11263-008-0152-6
- Marchand, E., Uchiyama, H., Spindler, F., 2016. Pose estimation for augmented reality: A hands-on survey. *IEEE Transactions on Visualization and Computer Graphics* 22 (12), 2633–2651. DOI: 10.1109/TVCG.2015.2513408
- Solà, J., Deray, J., Atchuthan, D., 2018. A micro Lie theory for state estimation in robotics. *ArXiv:1812.01537*.
- Terzakos, G., Lourakis, M., 2020. A consistently fast and globally optimal solution to the perspective-n-point problem. In: *ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I*. Springer-Verlag, pp. 478–494. DOI: 10.1007/978-3-030-58452-8_28
- Tsai, R., Lenz, R., 1989. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation* 5 (3), 345–358. DOI: 10.1109/70.34770
- Zhang, Z., Scaramuzza, D., 2018. A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 7244–7251. DOI: 10.1109/IROS.2018.8593941