

# Localización externa para robots móviles utilizando múltiples cámaras

Facundo Pessacg\*<sup>†</sup>, Thomas Fischer\*<sup>†</sup> y Matias Nitsche\*

\*CONICET-Universidad de Buenos Aires, Instituto de Investigación en Ciencias de la Computación (ICC)

<sup>†</sup>Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Computación

**Resumen**—En este trabajo se presenta un método de localización externa basado en visión. El mismo utiliza múltiples cámaras, lo que permite obtener la pose de un robot móvil mediante la fusión, utilizando el Filtro de Kalman Extendido, de la estimación obtenida mediante cada cámara. Para obtener cada estimación individual mediante una única cámara, se emplea el sistema de localización WhyCon. En este trabajo se muestran los resultados al utilizar el sistema propuesto en dos configuraciones particulares (maximizando la precisión del método y maximizando el área de cobertura) y en un caso más general. En todos los casos los errores obtenidos fueron comparables con errores de métodos similares del estado del arte.

## I. INTRODUCCIÓN

Uno de los problemas fundamentales en el campo de la robótica móvil es el de la localización, que consiste en determinar la pose del robot, es decir, posición y orientación, respecto de un sistema de coordenadas determinado. Cuando la localización es utilizada para resolver el control del robot, estos sistemas tienen que funcionar en tiempo real.

En el caso de ciertos entornos interiores (como fábricas o en un ambiente de experimentación controlado), donde utilizar localización basada en GPS no es posible, resulta de interés el obtener un sistema de localización externa alternativo. En particular, en este trabajo se propone el uso de cámaras de video convencionales como sensores para la localización, debido a su bajo costo y a que proveen mucha información en un rango virtualmente ilimitado. Por otro lado, son sensores pequeños y de bajo consumo.

En la actualidad existen diversos sistemas de localización externa basados en visión, entre los cuales se pueden mencionar [1], [2], [3], [4]. En general estos sistemas se basan en detectar un patrón en las imágenes, el cual varía en cada método.

Un sistema de localización externa muy conocido que utiliza cámaras es Apriltag [2]. Los patrones de Apriltag permiten codificar no solo la orientación sino también un identificador único. Este sistema es flexible y robusto, por este motivo muchos sistemas lo utilizan como base de su desarrollo con el fin de obtener la estimación de la pose [5], [6], y muchos otros para obtener información de *ground truth* [7], [8].

Siguiendo en esta línea, otro sistema de localización externa basado en sensado pasivo es TRIP [3], en donde el patrón consta de un conjunto de anillos concéntricos, divididos en varias regiones angulares, cada una de las cuales puede ser negra o blanca. El esquema de codificación, que incluye

la comprobación de paridad, permite que el método TRIP distinga entre  $3^9$  patrones distintos.

Entre los diversos métodos de localización externa basados en la detección de patrones visuales, podemos distinguir el método conocido como WhyCon [4], el cual se presenta como una alternativa de bajo costo a sistemas comerciales tales como ViCon [9]. El sistema WhyCon es capaz de detectar patrones circulares monocromáticos (ver figura 2). A partir del conocimiento de las dimensiones reales de dicho patrón, es posible recuperar la posición del mismo en tres dimensiones utilizando una sola cámara. Este tipo de patrón en particular permite emplear algoritmos de visión por computadora muy eficientes que resultan en tasas de detección de cientos a miles de cuadros por segundo, utilizando hardware convencional, valores incluso muy superiores a las capacidades de las cámaras de video convencionales actuales. Además, la precisión de dicho sistema se encuentra en el orden del 1 al 3 % de error relativo.

Los sistemas anteriormente mencionados, al contar con una única vista de la escena, poseen limitaciones tales como un área de cobertura definida directamente por el alcance de la cámara y una precisión mayormente limitada por la ubicación y características de la misma. Por esta razón, resulta interesante analizar sistemas de localización basados en el uso de múltiples cámaras. Por un lado, esto permite aumentar el área de trabajo posicionando las cámaras minimizando el solapamiento entre lo que ve cada una, y por otro, es posible mejorar la precisión de la localización maximizando el solapamiento utilizando distintas cámaras convergentes, logrando así que varias cámaras vean el mismo patrón a localizar desde diversos puntos de vista. De hecho, es posible combinar ambas estrategias según se requiera mayor área de cobertura o mayor precisión en cada sector del ambiente.

Un caso donde se utiliza un sistema multicámara es [10]. En este trabajo un *dataset* de un robot móvil desplazándose en diversos entornos con múltiples sensores es presentado. Para obtener la pose del robot se utiliza un conjunto de cámaras fijas que observan la escena, con el objetivo de detectar una serie de patrones adosados al robot. Cada cámara permite obtener un estimado de la pose del patrón y se utiliza el filtro *Extended Kalman Smoother* [11], (similar al Filtro de Kalman Extendido [12], pero que realiza correcciones en la estimación actual y también, en las estimaciones anteriores) para unificar la información y obtener un único estimado.

En [13] la pose del robot es obtenida detectando al robot directamente en las imágenes a partir de diferenciar las regiones en movimiento de un escenario que se asume estático (controlado). El aporte interesante de este trabajo es la generalización de una función objetivo, planteada originalmente para una sola cámara a varias. Sin embargo, en este trabajo no se aborda el problema de hallar la pose relativa entre las distintas cámaras.

En general, si bien se pueden identificar trabajos en los cuales se proponen métodos de localización externa basados en el uso de múltiples cámaras, no existe en la actualidad un sistema de propósito general que resuelva este problema dado que generalmente se proponen sistemas *ad-hoc*. Por otro lado, entre los sistemas de localización monoculares, se puede destacar WhyCon [4] por su precisión, eficiencia y robustez. Por estas razones, en este trabajo se presenta una generalización del sistema WhyCon al caso con múltiples cámaras, con el objetivo de satisfacer las limitaciones mencionadas y así obtener un sistema de localización multi-cámara de propósito general.

## II. MÉTODO

El problema principal que se busca resolver en este trabajo es el de obtener una única estimación de pose de un patrón (en este caso, del sistema WhyCon) a partir de la estimación individual obtenida mediante múltiples cámaras, donde cada una localiza al patrón utilizando una instancia del sistema WhyCon. Para resolver la fusión es necesario resolver dos problemas. En primer lugar, es necesario calibrar el sistema (para llevar todas las estimaciones individuales a un marco común) y, en segundo lugar, hacer la fusión de las estimaciones individuales (dado que no necesariamente coincidirán).

### II-A. Calibración

En la fig. 1 se pueden ver tres cámaras  $C_i$  las cuales observan un punto  $\mathbf{x}_i$  en el plano de la imagen de cada una de ellas respectivamente. Cada cámara posee su propio sistema de coordenadas. El objetivo de la calibración es obtener  $\mathbf{H}_{C_i}^W$  que permite realizar la transformación del punto  $\mathbf{X}$  que se encuentra en el sistema de coordenadas de la cámara al sistema de coordenadas del mundo  $W$  definido por el usuario.

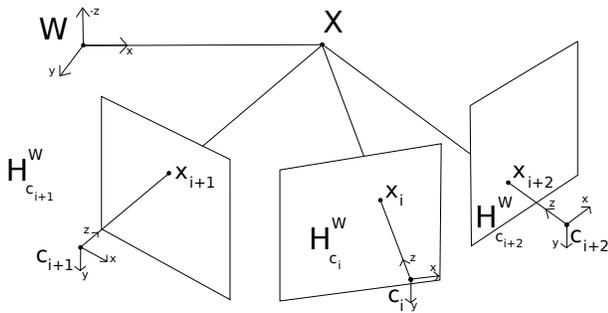


Figura 1. Modelo Multi-cámara.

Para realizar la calibración se utiliza el propio sistema WhyCon, estimando la pose de la cámara respecto de una

serie de patrones fijos que definen el sistema de coordenadas dado (dicho sistema es compartido y observado por todas las instancias). La calibración intrínseca de cada cámara se asume resuelta de antemano.

Para estimar la transformación  $\mathbf{H}_{C_i}^W$  de cada cámara  $C_i$ , se requiere definir el marco de referencia  $W$  colocando tres patrones en la escena que determinen los ejes  $X$  e  $Y$  sobre el plano de trabajo (generalmente el piso) y el origen del sistema (el eje  $Z$  quedará definido implícitamente a partir de los ejes previamente definidos).

La transformación que permite llevar un vector  $\mathbf{x}_c$  representado en el sistema de coordenadas  $C$  de la cámara, a uno  $\mathbf{x}_w$  en  $W$ , esta dada por (1).

$$\mathbf{x}_w = \mathbf{R} (\mathbf{x}_c - \mathbf{x}_0) \quad (1)$$

con:

$$\mathbf{R} = (\mathbf{t}_1^T \quad \mathbf{t}_2^T \quad \mathbf{t}_3^T)^{-1} \quad (2)$$

tal que:

$$\mathbf{t}_1 = \mathbf{x}_1 - \mathbf{x}_0 \quad \mathbf{t}_2 = \mathbf{x}_2 - \mathbf{x}_0 \quad \mathbf{t}_3 = \mathbf{t}_1 \times \mathbf{t}_2 \quad (3)$$

donde  $\mathbf{x}_i$  son las coordenadas de los tres patrones utilizados para definir al sistema de referencia  $W$  ( $\mathbf{t}_1$  define el eje  $X$ ,  $\mathbf{t}_2$  el eje  $Y$  y  $\mathbf{t}_3$  el eje  $Z$ ) y  $\mathbf{x}_0$  la posición del sistema  $C$  respecto de  $W$ .

### II-B. Fusión de las poses

Cada cámara tiene asociada una instancia del sistema WhyCon que procesa las imágenes capturadas y produce una estimación de la posición del patrón mientras se encuentre en su campo de visión. Con el fin de obtener una única estimación de la posición del patrón, se utiliza el Filtro de Kalman Extendido (EKF), para fusionar las distintas estimaciones asociadas a cada cámara.

En la fig. 2 puede verse un esquema del sistema de localización. A partir de la observación  $\mathbf{z}_i$ , que describe la posición del patrón respecto del sistema de coordenadas asociado a la cámara  $C_i$ , se busca obtener una estimación  $\mathbf{x}_i$  que describa la posición del patrón respecto de  $W$ . Por otro lado, como se puede observar en la fig. 2, las observaciones individuales tienen ruido debido al sistema de sensado. El EKF permite encontrar una estimación teniendo en cuenta estas incertidumbres.

Para utilizar el EKF es necesario definir previamente cual es el estado que se quiere estimar, el modelo de movimiento y el modelo de las observaciones, como así también los jacobianos de los modelos.

**Estado:** está representado por una variable aleatoria  $\mathbf{X}_t \sim \mathcal{N}(\mathbf{x}_t, \mathbf{P}_t)$ , de media  $\mathbf{x}_t \in \mathbb{R}^3$  y covarianza  $\mathbf{P}_t \in \mathbb{R}^{3 \times 3}$ , donde  $\mathbf{x}_t = (x, y, z)^T$  es la posición del patrón o robot a localizar.

**Modelo de Movimiento:** permite predecir la nueva posición del patrón en el instante siguiente. La entrada de modelo  $\mathbf{u}_t$  se define como  $\mathbf{u}_t = \mathbf{v}_t$ , donde  $\mathbf{v}_t$  corresponde a una estimación de la velocidad lineal del patrón, entonces el modelo de movimiento queda definido según (4).

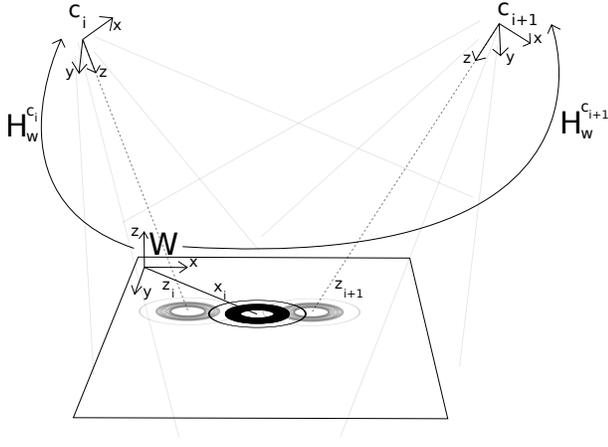


Figura 2. Sistema WhyCon con múltiples sensores.

$$f(\mathbf{x}_{t-1}, \mathbf{u}_{t-1} + \xi_{\mathbf{u}_{t-1}}, \mathbf{w}_{t-1}) = \begin{pmatrix} x_{t-1} + \Delta t(u_x + \xi_{u_x}) + w_x \\ y_{t-1} + \Delta t(u_y + \xi_{u_y}) + w_y \\ z_{t-1} + \Delta t(u_z + \xi_{u_z}) + w_z \end{pmatrix} \quad (4)$$

donde  $\mathbf{w}$  corresponde al ruido asociado al modelo de movimiento, con  $\mathbf{w} \sim \mathcal{N}(0, \mathbf{Q}_w)$ ,  $\xi_{\mathbf{u}}$  es ruido gaussiano aditivo que la velocidad tiene asociado, con media 0 y covarianza  $\mathbf{P}_u$ .

La velocidad  $\mathbf{v}_t$  se estima a partir de tomar la diferencia de posiciones de las observaciones, y utilizando las observaciones de todas las cámaras que estén detectando el robot. Luego, se aplica un filtro pasabajo para disminuir el ruido. En determinados casos la velocidad se puede obtener directamente a partir de un sensor propio del robot (por ejemplo, de la odometría obtenida mediante encoders). En este trabajo se optó por la primera opción, haciendo del robot un objeto pasivo que se sigue mediante observaciones externas.

Debido a que  $f$  es una función lineal respecto de  $\mathbf{x}$  y  $\mathbf{w}$ , los jacobianos de dicha función con respecto a estas variables resulta ser la identidad, es decir,  $\frac{\delta f}{\delta \mathbf{x}} = I$  y  $\frac{\delta f}{\delta \mathbf{w}} = I$ .

Por otra parte, el jacobiano  $\mathbf{J}_u$  de  $f$  con respecto a  $\mathbf{u}$  se muestra en (5).

$$\mathbf{J}_u = \frac{\delta f}{\delta \mathbf{u}} = \Delta t I \quad (5)$$

Utilizando  $\mathbf{J}_u$  se propaga la covarianza  $\mathbf{P}_u$  para obtener la covarianza tras el paso de predicción, como se observa en (6).

$$\mathbf{Q}_u = \mathbf{J}_u \mathbf{P}_u \mathbf{J}_u^T \quad (6)$$

Por último, la matriz de covarianza total se muestra en (7).

$$\mathbf{Q} = \mathbf{Q}_w + \mathbf{Q}_u \quad (7)$$

donde  $\mathbf{Q}$  es la matriz usada en el paso de predicción de *EKF*.

**Modelo de Observaciones:** El modelo de observaciones utiliza la información de los sensores para mejorar la precisión de la estimación del estado durante el paso de corrección del *EKF*. Teniendo en cuenta que cada una de las instancias del sistema WhyCon arroja observaciones sobre la posición del patrón respecto de  $W$ , el modelo de observación se muestra en (8).

$$h(\mathbf{x}_{t-1}, \mathbf{v}_{t-1}) = \mathbf{x}_{t-1} + \mathbf{v}_{t-1} \quad (8)$$

donde  $\mathbf{v} \sim \mathcal{N}(0, \mathbf{R})$  corresponde al ruido de observación.

Se puede observar en (8) que  $h$  depende directamente de  $\mathbf{x}_{t-1}$ . Esto se debe a que se utiliza directamente el valor obtenido por la observación para actualizar la estimación. Dada la simplicidad del modelo, los jacobianos resultan la identidad, es decir,  $\mathbf{H} = \frac{\delta h}{\delta \mathbf{x}} = I$  y  $\mathbf{V} = \frac{\delta h}{\delta \mathbf{v}} = I$ .

La simplicidad de este modelo de observación radica en el uso de un único sistema de coordenadas  $W$ , por lo que la observación esperada equivale directamente al estimado que se tiene actualmente.

Al ejecutar el paso de corrección ante un sensado  $\mathbf{z}_i$ , se deben resolver dos cuestiones: transformar la observación  $\mathbf{z}_i$  de  $C_i$  a  $W$  y obtener la matriz de covarianza  $\mathbf{R}$  asociada a  $\mathbf{z}_i$ , también representada respecto de  $W$ .

Por lo tanto, se definió la covarianza  $\mathbf{R}$  que representa el modelo de error del sistema WhyCon. Dicho modelo depende directamente de la distancia entre el patrón y la cámara y los ángulos de la misma. En este trabajo los errores angulares,  $e_\alpha$  y  $e_\beta$  (ecuación 9), son considerados constantes y calculados en base a la distancia focal  $f_{x,y}$ , al campo de visión  $a$  de la cámara y al error  $p$  en píxeles del sistema WhyCon.

$$e_{\alpha, \beta} = \frac{p \times a}{f_{x,y}} \frac{\pi}{180} \quad (9)$$

Por otro lado, el error en distancia utilizado se planteó en [4] según la ecuación (10).

$$e_\rho = \frac{\Delta f}{f_x} + \Delta e_0 \frac{\mathbf{x} f_x}{d_0} + \eta_{rad} \quad (10)$$

con  $\Delta f$  el error de la estimación de la distancia focal,  $\Delta e_0$  el error de la estimación de los ejes de la elipse,  $d_0$  el diámetro del círculo negro del patrón,  $\eta_{rad}$  el error relativo de la distorsión radial del modelo. Las variables  $\Delta f$  y  $\eta_{rad}$  se estiman en la calibración de la cámara, pero  $\Delta e_0$  es influido por muchos factores, como por ejemplo la luz o el ruido termal de la cámara. En particular, en los experimentos de [4],  $\Delta e_0$  tiene un promedio de 0,15 píxeles. En el caso general, estos parámetros deben ser encontrados empíricamente.

En resumen, la covarianza asociada a una observación  $\mathbf{z}_i$ , que se asume en coordenadas polares, se define en (11).

$$\mathbf{R}_{pol} = \begin{bmatrix} e_\rho & 0 & 0 \\ 0 & e_\alpha & 0 \\ 0 & 0 & e_\beta \end{bmatrix} \quad (11)$$

Finalmente para obtener el modelo deseado es necesario propagar  $\mathbf{R}_{pol}$  a una covarianza en coordenadas cartesianas y respecto del sistema  $W$ . Para ello, dada una observación  $\hat{\mathbf{z}}_i = (\rho, \alpha, \beta)$  se define la función  $g(\hat{\mathbf{z}}_i)$  en (12).

$$g(\hat{\mathbf{z}}_i) = \mathbf{H}_{C_i}^W PaC(\hat{\mathbf{z}}_i) \quad (12)$$

donde  $\mathbf{H}_{C_i}^W$  es la transformación del sistema de coordenadas de la cámara  $C_i$  al sistema  $W$  y  $PaC$  la función para convertir de coordenadas polares a cartesianas definida según (13).

$$PaC(\rho, \alpha, \beta) = (\rho \cos(\alpha) \cos(\beta), \rho \cos(\alpha) \sin(\beta), \rho \sin(\alpha)) \quad (13)$$

Luego, para propagar  $\mathbf{R}_{\text{pol}}$  se debe hallar el jacobiano  $\mathbf{J}_g$  por regla de la cadena como se muestra en (14).

$$\mathbf{J}_g = \frac{\partial g}{\partial \hat{\mathbf{z}}_i} = \mathbf{R}_i \mathbf{J}_{PaC}(\hat{\mathbf{z}}_i) \quad (14)$$

con  $\mathbf{R}_i$  la matriz de rotación perteneciente a la transformación  $\mathbf{H}_{C_i}^W$  y  $\mathbf{J}_{PaC}$  evaluado en la observación  $\hat{\mathbf{z}}_i$  se define en (15).

$$\mathbf{J}_{PaC} = \begin{bmatrix} \cos(\alpha) \cos(\beta) & -\rho \sin(\alpha) \cos(\beta) & -\rho \cos(\alpha) \sin(\beta) \\ \cos(\alpha) \sin(\beta) & -\rho \sin(\alpha) \sin(\beta) & \rho \cos(\alpha) \cos(\beta) \\ \sin(\alpha) & \rho \cos(\alpha) & 0 \end{bmatrix} \quad (15)$$

Por lo tanto,  $\mathbf{R}$  queda definida según (16).

$$\mathbf{R} = \mathbf{J}_g \mathbf{R}_{\text{pol}} \mathbf{J}_g^T \quad (16)$$

### III. EXPERIMENTOS

Con el fin de evaluar la precisión del método y de poder evaluar fácilmente distintas configuraciones y escenarios de prueba sin requerir una infraestructura compleja para los experimentos (sistema de cámaras interconectadas), en todos los experimentos se trabajó en simulación. La herramienta de simulación que se utilizó es V-Rep (*Virtual Robot Experimentation Platform*) [14], con la cual se diseñaron diversos escenarios acorde a cada experimento. En todas las simulaciones se cuenta con un robot *Pioneer P3-DX*, el cual posee un patrón circular adosado a su chasis (fig. 3).

El robot fue controlado manualmente mediante comandos de velocidad. Los escenarios cuentan con cámaras ubicadas en distintas posiciones, las mismas fueron calibradas según la sección II-A. Las imágenes obtenidas con estas son procesadas por instancias individuales del sistema WhyCon. Los resultados de localización de dichas instancias son tomados como entrada al sistema de localización propuesto.

El simulador permite conocer además las posiciones reales de los elementos de la escena (*ground truth*), en particular la del patrón, esta información fue utilizada para contrastar con los resultados obtenidos por el sistema de localización propuesto. Las medidas de dicho patrón son: 0,2m para el diámetro externo (círculo negro) y 0,1m para el diámetro interno (círculo blanco). Las cámaras que se simularon se configuraron con resolución de  $800 \text{ px} \times 600 \text{ px}$ , un ángulo de visión de 60 grados y longitud focal de  $682 \text{ px}$ . Esta configuración corresponde a una situación típica usada comúnmente con el sistema de localización WhyCon.

#### III-A. Máximo Solapamiento 2 y 4 cámaras

*Solapamiento con 2 Cámaras:* El objetivo de este experimento es evaluar la precisión del método de localización en comparación con la precisión del método WhyCon al usar una sola cámara. Además, ambas estimaciones se contrastan con el *ground truth* del simulador. La hipótesis planteada es que al utilizar dos cámaras con máximo solapamiento entre ellas, la precisión aumenta con respecto a la utilización de una única cámara en la escena. Además, se quiere evaluar en este caso el sistema multicámara con la menor cantidad de cámaras posibles (2 cámaras). La fig. 3 muestra la escena utilizada para este experimento.

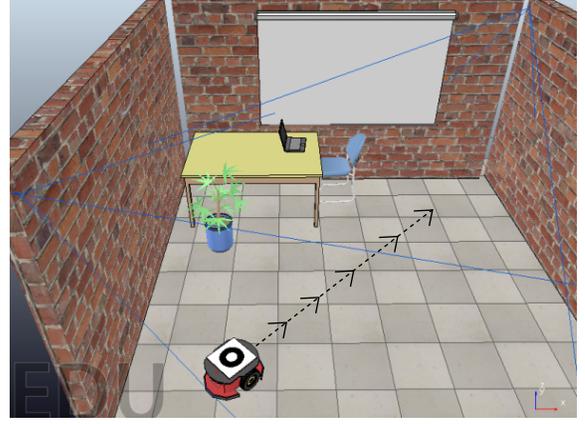


Figura 3. Escena con dos cámaras.

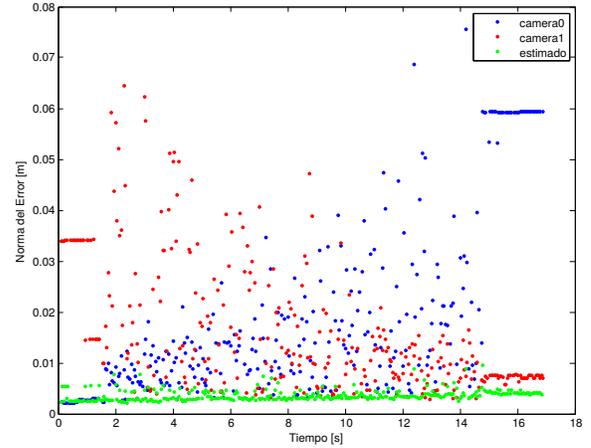


Figura 4. Error de estimación para el caso de dos cámaras con máximo solapamiento.

En la fig. 4 se presenta la norma del error de precisión respecto del *ground truth* para cada estimación. Se observa que a medida que aumenta el tiempo, el error crece para *camera0* y decrece para *camera1*. Este comportamiento se atribuye a la distancia en cada momento del patrón a cada cámara, ya que a medida que avanza el tiempo, el patrón se aleja de *camera0* y se acerca a *camera1*. Además, se puede ver que en todo momento la norma del error estimado es menor que cada instancia de WhyCon. Los errores de estimación cae para toda la escena por debajo de 0.01m.

*Solapamiento con 4 cámaras:* El objetivo de este experimento es evaluar como varía la precisión del método al agregar más cámaras en la escena. Por lo tanto, se agregan dos cámaras adicionales a lo planteado en el experimento anterior. Las cámaras en las esquinas son iguales a las del caso previo, las cuales ven todo el recorrido en todo momento. Las cámaras que se encuentran en el techo, una de ellas ve la primera mitad del recorrido y la otra la segunda mitad. Entre estas dos últimas cámaras existe un mínimo solapamiento en la trayectoria del robot. De esta forma, en cada instante hay tres cámaras observando el patrón. En la fig. 5 se puede ver un

esquema de este escenario.

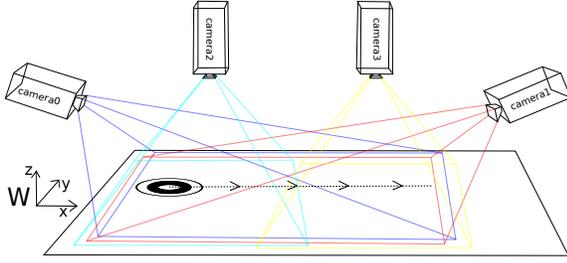


Figura 5. Esquema del recorrido en la escena en la cual el patrón se aleja de camera0 y se acerca a camera1, atravesando a su paso el trayecto cubierto por camera2 y camera3.

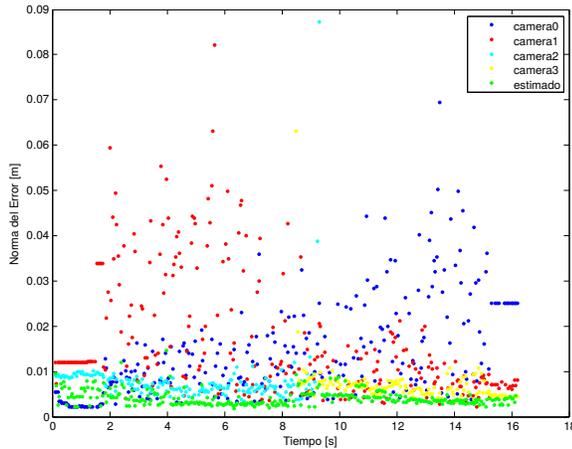


Figura 6. Error de estimación para el caso de cuatro cámaras.

En la fig. 6 se presenta la norma del error de precisión respecto del *ground truth* para cada estimación. Se puede observar que el comportamiento de la norma del error de las instancias de WhyCon asociadas a las cámaras en los extremos de la habitación es igual al del experimento anterior. Al igual que en el experimento anterior, los errores de estimación cae para toda la escena por debajo de 0.01m. Para este caso, agregar más cámaras en la escena no mejora la calidad del método.

### III-B. Máxima Cobertura

A diferencia de los experimentos anteriores, en esta sección se desea analizar la precisión del método en una escena en la cual las cámaras están colocadas con el menor solapamiento posible con el objetivo de maximizar la cobertura de la trayectoria del robot. En esta sección se estudió el caso de mayor cobertura, es decir, las cámaras tienen poco solapamiento entre ellas. Por lo tanto, se colocaron 4 cámaras en el techo en un pasillo largo como muestra la fig. 7. El recorrido consiste en mover el robot de una punta a la otra del pasillo.

En la fig. 8 se presenta la norma del error respecto del *ground truth*. Se obtuvo que el error es menor a 0.05 m para todo el recorrido. Además, se observa en la fig. 8 que el error de cada instancia de WhyCon crece al aumentar la distancia

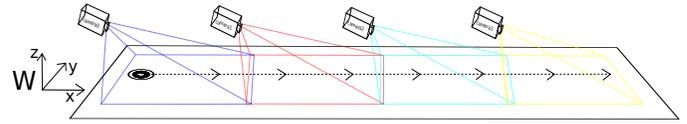


Figura 7. Esquema del recorrido en la escena en la cual el patrón llega al final del pasillo.

del patrón a la cámara, y debido a que una única cámara está observando el patrón, el error del estimado también crece. Cuando el patrón es visto por dos cámaras el error en el estimado decrece nuevamente.

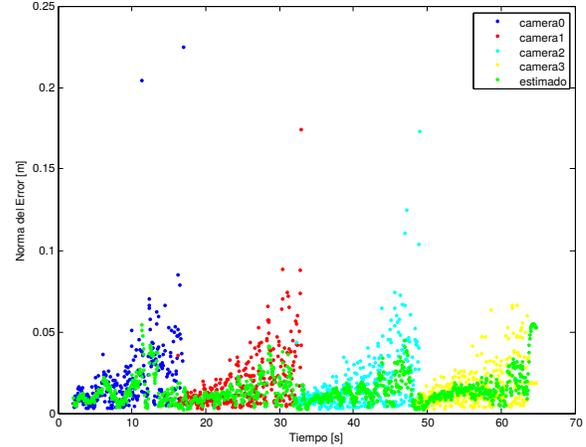


Figura 8. Error de estimación para el caso de máxima cobertura.

### III-C. Escena General

Como último experimento, se propuso un escenario semejante a una situación real, de un área  $16m \times 16m$ , conteniendo distintas habitaciones separadas a las cuales el robot ingresa una por una. La distribución de las cámaras se diseñó intentando cubrir todo el espacio posible sin sobrecargar de cámaras el lugar, en lo que se consideró la mejor ubicación para cada una de ellas.

En la fig. 9 se muestra la trayectoria estimada por el método propuesto junto a la trayectoria real. A su vez, se incluyen las trayectorias individuales estimadas por cada instancia de WhyCon. Además, cada 25 muestras y aumentadas por un factor de escala de 5000, se graficó la covarianza del estimado.

El sistema es capaz de estimar la posición del robot en todo momento, incluso cuando el mismo no es captado por ninguna cámara. En estos casos, la covarianza de la estimación crece. En el recuadro a de la fig. 9, se observa un crecimiento grande de la covarianza ya que la última vez que se vió el patrón fue por una cámara lejana. En el caso del recuadro b, la última vez corresponde a una cámara cercana y por ende el crecimiento de la covarianza en este intervalo es menor que en a.

En la fig. 10 se muestra el valor de la norma del error en función del tiempo para este experimento. El promedio de la norma del error estimado es de  $0,012m$  y el máximo valor que tiene la norma del error es de  $0,496m$ . Los picos del gráfico corresponden a etapas de la trayectoria en las cuales

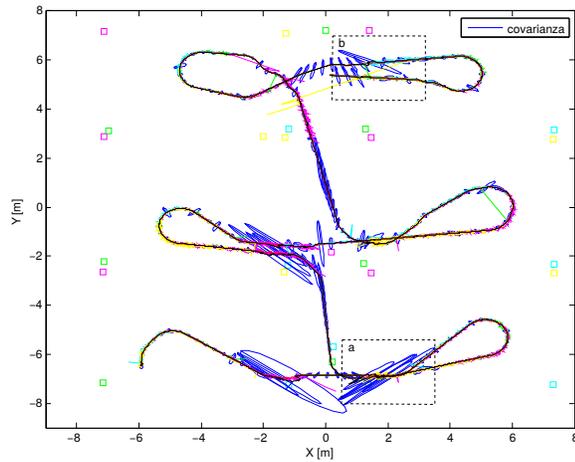


Figura 9. Recorrido del robot sobre el escenario de oficinas. Las elipses azules muestran la covarianza del estimado en los ejes  $XY$ . Los cuadrados de colores representan las cámaras. El recorrido real se encuentra en rojo. Por encima de éste se ve el de cada cámara en diferentes colores. Por sobre estos se muestra en negro el recorrido estimado por el método.

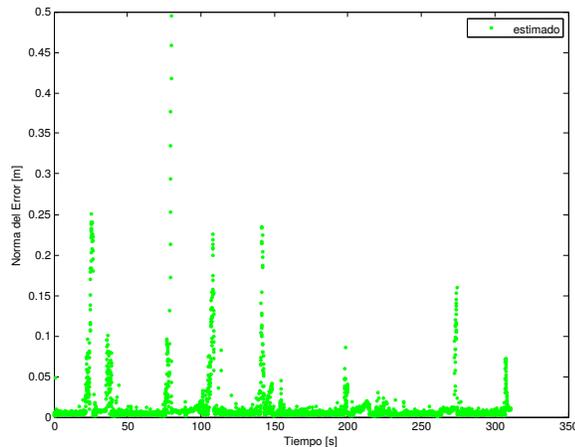


Figura 10. Error de estimación para el caso de la escena general.

ninguna cámara observa el patrón, o a casos en los cuales la cámara se encuentra a una distancia lejana. Por otra parte, la máxima distancia del patrón a una cámara en una observación es de  $10,016\text{ m}$  y en dicha observación la norma del error de WhyCon sobre esa cámara es de  $1,174\text{ m}$ , cuando la norma del error estimado en ese mismo instante es de  $0,011\text{ m}$ . Es decir, en esta situación, la ganancia en la precisión de la estimación obtenida mediante la fusión de las cámaras es 100 veces mayor que con una única cámara.

#### IV. CONCLUSION

En este trabajo se presenta un método de localización externa basado en visión que utiliza múltiples cámaras, lo que permite obtener la pose de un robot autónomo móvil mediante la estimación individual obtenida por cada cámara y la fusión de dicha información utilizando el Filtro de Kalman Extendido. Para las estimaciones individuales se emplea el sistema de localización WhyCon.

Se realizaron experimentos de máximo solapamiento con 2 y 4 cámaras y se obtuvo una precisión estimada de  $0,01\text{ m}$  para ambos casos. Se observó también que agregar más de 2 cámaras observando al mismo patrón no mejora la precisión del sistema. Por otro lado, se realizaron experimentos de máxima cobertura obteniendo una precisión estimada de  $0,05\text{ m}$ . Por último, se simuló una escena general y se colocaron 29 cámaras para estimar la posición de un robot. En este caso, el promedio de la norma del error estimado es de  $0,012\text{ m}$ . La máxima distancia del patrón a una cámara en una observación es de  $10,016\text{ m}$  y en dicha observación la norma del error de WhyCon sobre esa cámara es de  $1,174\text{ m}$ , cuando la norma del error estimado en ese mismo instante es de  $0,011\text{ m}$ .

En general, se pudo ver que el uso de múltiples cámaras permite mejorar el error de localización considerablemente respecto de el uso de una única cámara. Además, se logró ampliar el área de cobertura del sistema a un entorno mayor.

#### REFERENCIAS

- [1] D. H. Kato, "ARToolKit," [cited 8 Jan 2014]. [Online]. Available: <http://www.hitl.washington.edu/artoolkit/>
- [2] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *Robotics and Automation (ICRA)*. IEEE, 2011, pp. 3400–3407.
- [3] D. L. De Ipiña, P. R. S. Mendonça, and A. Hopper, "TRIP: A low-cost vision-based location system for ubiquitous computing," *Personal and Ubiquitous Computing*, vol. 6, no. 3, pp. 206–219, 2002.
- [4] T. Krajník, M. Nitsche, J. Faigl, P. Vanek, M. Saska, L. Preucil, T. Duckett, and M. Mejail, "A Practical Multirobot Localization System," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 76, no. 3–4, pp. 539–562, 2014.
- [5] M. Faessler, E. Mueggler, K. Schwabe, and D. Scaramuzza, "A monocular pose estimation system based on infrared LEDs," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 907–913, 2014. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6906962>
- [6] A. Richardson, J. Strom, and E. Olson, "AprilCal: Assisted and repeatable camera calibration," *IEEE International Conference on Intelligent Robots and Systems*, pp. 1814–1821, 2013.
- [7] C. Feng and V. R. Kamat, "Plane registration leveraged by global constraints for context-aware AEC applications," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 5, pp. 325–343, 2013.
- [8] F. Duvallet, M. R. Walter, T. Howard, S. Hemachandra, J. Oh, S. Teller, N. Roy, and A. Stentz, "Inferring maps and behaviors from natural language instructions," *Springer Tracts in Advanced Robotics*, vol. 109, pp. 373–388, 2016.
- [9] Vicon, "Vicon MX Systems," [cited 8 Jan 2014]. [Online]. Available: <http://www.vicon.com>
- [10] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti, and P. Taddei, "Rawseeds ground truth collection systems for indoor self-localization and mapping," *Autonomous Robots*, vol. 27, no. 4, pp. 353–371, 2009.
- [11] B. M. Yu, K. V. Shenoy, and M. Sahani, "Derivation of extended kalman filtering and smoothing equations forward recursions," *Matrix*, no. 1, pp. 1–5, 2004.
- [12] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*, 2013, vol. 53, no. 9.
- [13] C. Losada, M. Mazo, S. Palazuelos, D. Pizarro, and M. Marrón, "Multi-camera sensor system for 3D segmentation and localization of multiple mobile robots," *Sensors*, vol. 10, no. 4, pp. 3261–3279, 2010.
- [14] E. Rohmer, S. P. N. Singh, and M. Freese, "V-REP : a Versatile and Scalable Robot Simulation Framework," pp. 0–5.