

# Navegación por Visión Estereoscópica Asistida por GPS

Emiliano Statello<sup>#1</sup>, Ramiro Verrastro<sup>#2</sup>, Basilio Robino<sup>\*</sup>, Juan Carlos Gomez<sup>#§</sup>, Silvio Tapino<sup>\*</sup>

<sup>#</sup> Grupo de Inteligencia Artificial y Robótica (GIAR), Univ. Tecnológica Nacional Fac. Regional Bs. As.

Av. Medrano 951, Ciudad Autónoma de Buenos Aires, Argentina.

<sup>1</sup> emilianostatello@yahoo.com.ar

<sup>2</sup> ramiroverraastro@frba.utn.edu.ar

<sup>\*</sup> Cátedra Proyecto Final, Ingeniería Electrónica, Univ. Tecnológica Nacional Fac. Regional Bs. As.

Av. Medrano 951, Ciudad Autónoma de Buenos Aires, Argentina.

<sup>§</sup> Centro de Electrónica e Informática, Instituto Nacional de Tecnología Industrial (INTI).

Av. General Paz 5445, San Martín, Buenos Aires, Argentina.

*Abstract*— This paper describes a method for real-time visual odometry, using stereo images combined with information obtained by a GPS system, performing localization in a geo-referenced map.

The coordinates of features (natural landmarks) are obtained from rectified pairs of images. Using these points, the values of roto translation is calculated to trace the trajectory of the cameras.

The selection of the features is essential for the convergence of roto translation values.

In this way the path of the cameras is computed with three degrees of freedom in translation (x, y, z) and three degrees of freedom in rotation (roll, pitch and yaw).

A Kalman filter is used to merge the position obtained by the vision system with that obtained by GPS, so that each measurement is weighed according to their variance.

The photogrammetric calculation gives a differential measurement, so as the vehicle moves, the error increases. Moreover, it is not always possible to obtain the features needed to compute the triangulation, which introduces another cumulative error. Therefore it is of vital importance the use of GPS, as absolute measurement to correct the path. However, the system still working without GPS, using only the visual odometry.

The combination of both measurements using the Kalman filter provides a path with lower error than measurements made independently, and fills sections of the route where the sensor information is not sufficient to trace the trajectory, which demonstrates the robustness of the method.

*Resumen*— Se propone un método para realizar odometría visual en tiempo real, utilizando imágenes estereoscópicas combinadas con información obtenida por GPS, realizando localización en un mapa geo-referenciado.

A partir de pares de imágenes rectificadas, se obtienen los puntos característicos (marcas naturales). Con dichos puntos se calculan los valores de roto-traslación para trazar el recorrido.

La selección de los puntos es esencial para que la roto traslación converja.

Se computa el recorrido de las cámaras con tres grados de libertad en la traslación (x, y, z) y tres grados de libertad en la rotación (roll, yaw y pitch).

Para fusionar la posición obtenida por el sistema de Visión con la obtenida mediante GPS se utiliza un filtro de Kalman,

de forma tal que cada medición es ponderada de acuerdo a su varianza.

El cálculo fotogramétrico es una medición diferencial, por lo cual a medida que el móvil avanza, aumenta el error. Además, no siempre es posible obtener los puntos característicos necesarios para computar la triangulación, lo que introduce otro error acumulativo. Por ello cobra vital importancia el GPS, como medición absoluta, para corregir la trayectoria. Sin embargo el sistema puede seguir funcionando sin GPS, utilizando únicamente la odometría visual.

La combinación de ambas mediciones mediante el filtro de Kalman provee una trayectoria con menor error que las mediciones tomadas independientemente y completa tramos del recorrido en los cuales la información de los sensores es insuficiente, lo cual da muestra de la robustez del método.

## I. INTRODUCTION

La odometría óptica ha cobrado gran interés en estos tiempos ya que el trazado del camino es una capacidad clave para el desempeño de vehículos autónomos [1], así como para implementar rutinas de SLAM (Localización y Mapeo Simultáneo) en todo tipo de móviles dotados de cámaras de video [2] como sensor de posición.

Cuando se trabaja con este tipo de sensores la calidad de la medición está más afectada por la naturaleza de la imagen que por la implementación física del sistema. En superficies uniformes, con poca textura o con patrones repetitivos no se puede establecer la medición, ya que la correspondencia de los puntos medidos no resulta unívoca, lo cual compromete la precisión y la resolución espacial de los modelos generados.

En este trabajo se presenta un método para realizar odometría visual en tiempo real, utilizando imágenes estereoscópicas rectificadas combinadas con información obtenida por GPS, realizando la localización y la estimación de la trayectoria en un mapa geo-referenciado. Se utilizan rutinas con funciones de biblioteca en lenguaje C++.

## II. DESARROLLO

El diagrama en bloques del método de posicionamiento se muestra en la Fig. 1. En la primera etapa se buscan puntos característicos en ambas imágenes logrando que el

seguimiento de los mismos proporcione información robusta y confiable. Seguido a esta etapa se realiza el seguimiento de dichos puntos cuadro a cuadro y con esa información se calcula la roto-traslación del móvil. Posteriormente con la información procesada del GPS se utiliza un filtro de Kalman para fusionar la posición obtenida por el sistema de Visión con la obtenida mediante GPS, de forma tal que cada medición es ponderada de acuerdo a su varianza.

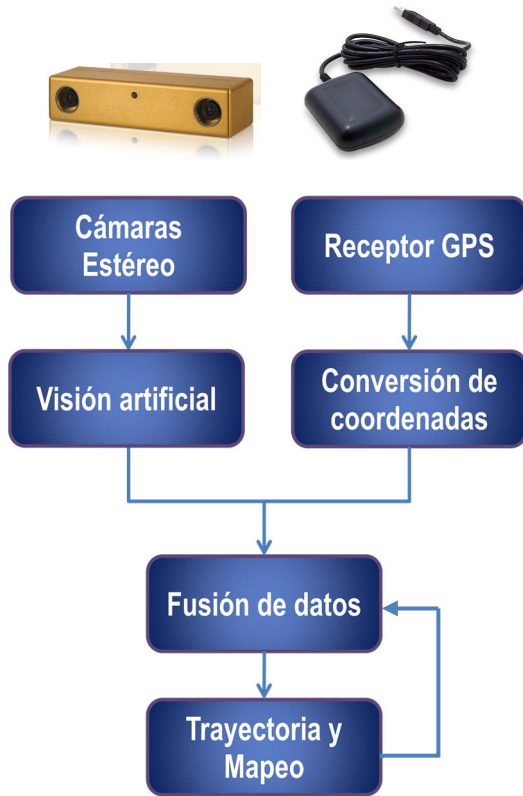


Fig. 1. Diagrama en bloques del método desarrollado

Dentro del bloque de Visión Artificial la información se procesa como se muestra en la Fig. 2.

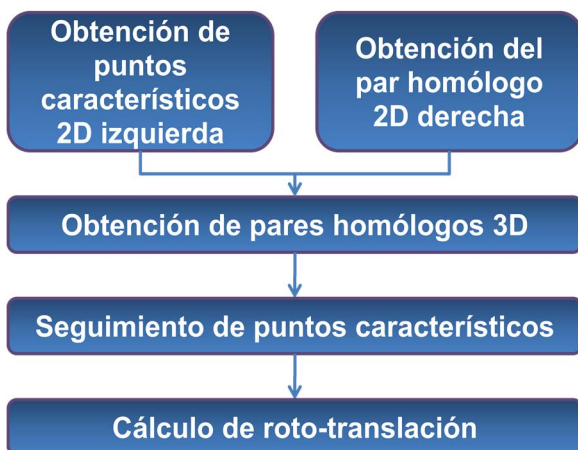


Fig. 2. Proceso de Visión Artificial.

#### A. Extracción de puntos característicos de una imagen

Para garantizar el éxito del sistema es necesario obtener puntos de textura suficiente como para distinguirlos de su entorno [3]. Las regiones de las imágenes donde la

intensidad de los píxeles es uniforme no contienen información suficiente como para determinar la posición del punto dentro de la región. En cambio en las zonas de la imagen donde la intensidad varía de forma abrupta, pueden ser asociados a algún tipo de borde, pero esto no es suficiente, ya que estos puntos pueden ser confundidos con otros a lo largo del mismo borde. Es por ello que debe exigirse una variación abrupta en dos direcciones, en lo posible ortogonales, para asegurar que el punto sea identificado unívocamente en su entorno.

Se utiliza la definición de punto esquina o “corner” [4] como punto característico, para encontrar los puntos característicos de una imagen se evalúa la matriz de derivadas parciales de segundo orden de las intensidades de la imagen, como se describe en la Ec. (1).

$$H(x, y) = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix}_{img} \quad (1)$$

Donde:  $I(x, y)$  representa la imagen rectificada y  $H(x, y)$  la matriz Hessiana (de derivadas parciales).

Se eligen aquellos puntos que arrojen valores altos en dos direcciones significativamente diferentes. Aquí se consideran solo los auto-valores de la correlación de matrices Hessianas y no los auto-vectores, con la ventaja de que esta parametrización es invariante a la rotación.

Se obtienen así puntos que pueden ser identificados y seguidos de forma robusta tanto en las imágenes homólogas del par estereoscópico, como cuadro a cuadro a lo largo del tiempo.

#### B. Obtención de pares

Una vez obtenidos los puntos característicos de una imagen, se buscan los mismos puntos en la imagen homóloga, cuyas filas tendrán una correspondencia epipolar debido a que se trata de imágenes rectificadas y normalizadas. Es decir que los píxeles de cada fila de la imagen se corresponden en la imagen homóloga.

Se utiliza como patrón una región de interés cuadrada en torno a cada punto característico y luego se realiza una correlación, desplazando el patrón horizontalmente (en una ventana rectangular) en la misma posición vertical de la imagen homóloga, el punto de máxima correlación es el par del punto característico de la primera imagen, este proceso se repite para todos los puntos característicos.

#### C. Seguimiento de puntos característicos

Para el cálculo de la trayectoria se utilizan como referencia los puntos característicos, estos poseen una robustez suficiente para ser seguidos entre una imagen y la imagen siguiente. Para dicho seguimiento se utiliza el modelo de flujo óptico el cual analiza el patrón aparente del movimiento. Para implementar este modelo se usa el algoritmo de Lucas-Kanade (LK) [5].

La idea básica del algoritmo de LK se apoya en tres supuestos:

1) *Brillo constante*: Un pixel de la imagen perteneciente a un objeto no cambia de intensidad al moverse de un cuadro al siguiente.

2) *Persistencia temporal*: Los incrementos temporales son lo suficientemente cortos en relación con la escala de movimiento de la imagen, como para suponer que el objeto no se mueve en forma abrupta entre cuadros.

3) *Coherencia espacial*: Puntos vecinos de una escena que pertenecen a la misma superficie tienen un movimiento similar, y por lo tanto también lo tienen los puntos cercanos de la proyección en el plano de la imagen.

A partir del supuesto de brillo constante, se puede igualar a cero la Eq. (2).

$$\Delta I_x \cdot v + \Delta I_y \cdot u + \Delta I_t = 0 \quad (2)$$

Donde:  $\Delta I_x$  y  $\Delta I_y$  son el gradiente de intensidad de la imagen en el pixel  $(x, y)$ ,  $u$  y  $v$  son las velocidades en pixels/frame con las que el objeto se desplaza, y  $\Delta I_t$  es la variación de intensidad en el pixel  $(x, y)$  entre un frame y el siguiente.

Asumiendo persistencia temporal y coherencia espacial, se adopta una ventana de  $M \times M$  pixels alrededor del punto y minimizando por cuadrados mínimos, se resuelve por medio de la Ec. (3).

$$\begin{bmatrix} \sum_{i=0}^M \Delta I_x(p_i) \cdot \Delta I_x(p_i) & \sum_{i=0}^M \Delta I_x(p_i) \cdot \Delta I_y(p_i) \\ \sum_{i=0}^M \Delta I_x(p_i) \cdot \Delta I_y(p_i) & \sum_{i=0}^M \Delta I_y(p_i) \cdot \Delta I_y(p_i) \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum \Delta I_x \cdot \Delta I_t \\ \sum \Delta I_y \cdot \Delta I_t \end{bmatrix} \quad (3)$$

Donde:  $p_i$  es el pixel observado, y  $M$  es el tamaño de la ventana.

#### D. Cálculo de coordenadas espaciales

Se parte de un modelo matemático en el cual se asigna a cada pixel de las imágenes rectificadas una recta en el espacio. De esta forma a cada par de puntos obtenidos ( $Pol$  para la cámara izquierda y  $Por$  para la cámara derecha) le corresponden dos rectas coplanares ( $\vec{R}l$  y  $\vec{R}r$  con versores directores  $\vec{v}l$  y  $\vec{v}r$ ) y su intersección corresponde con la coordenada  $x, y, z$  buscada ( $\vec{P}$ ), con un error que dependerá de la resolución de las cámaras y de la distancia euclídea al punto. A mayor distancia, mayor error en la posición estimada, pero menos error en la determinación de la posición angular. En la Fig. 3 se grafica el esquema del par estereoscópico con las rectas del modelo matemático.

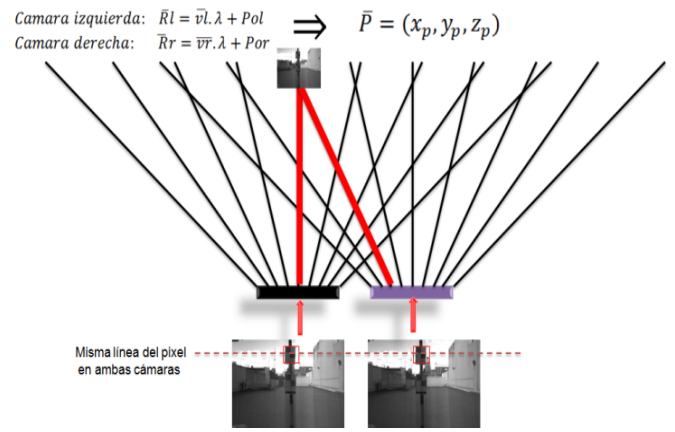


Fig. 3. Esquema de rectas del modelo matemático

#### E. Ponderación de puntos en el espacio

Sabiendo que el cálculo inicial de la rototraslación se realiza con tres puntos del espacio total, se busca que estos tres puntos posean la mejor estabilidad espacial, de acuerdo con Hirschmüller [6] y a su vez conformen un área capaz de darle solidez a los vectores que describen la perspectiva de las cámaras. Para ello del set completo de puntos, obtenidos por medio de los procedimientos ya descritos; se realizan las combinaciones tomadas de a tres y se analizan las distancias euclídeas relativas entre un cuadro y el siguiente. Cada uno de estos tres puntos definen un triángulo, cuya área se toma como factor de ponderación. De esta forma la rototraslación se realiza con la terna de puntos de mayor robustez.

#### F. Rototraslación

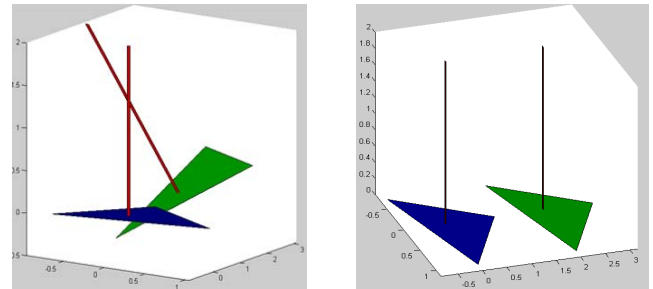


Fig. 4: a) Rotación de los puntos característicos. b) Desplazamiento de los puntos característicos cuadro a cuadro.

Con la terna de puntos mejor ponderados se calcula la rotación en el espacio, para ello se obtienen los vectores normales a los planos conformados por dichos tres puntos y con ellos se mide la rotación de los ángulos azimut, elevación y cenit como se puede observar en la Fig. 4.a. Todo el set de puntos es rotado desde el cuadro anterior al nuevo cuadro para obtener el desplazamiento como se puede observar en la Fig. 4.b.

Dichos valores de rototraslación se usan para la estimación total de la trayectoria.

#### G. Obtención de GPS y conversión de coordenadas

La lectura de los datos de GPS se realiza a través de las definiciones de los tipos de datos soportados por el protocolo NMEA[7], obteniendo como datos de interés la marca temporal (*timestamp*) en que se produce la muestra, así como los valores de latitud, longitud, altitud y *HDOP*(Coeficiente de dilución de la precisión horizontal) [8]. Este último resulta un valor clave para caracterizar el desvío estándar de la medición según Eq. (4) y poder aplicar la fusión por Kalman como se desarrolla en la sección II-I.

$$\sigma_F \cdot HDOP = \sigma_m \quad (4)$$

Donde:  $\sigma_F$  es el desvío estándar especificado por el fabricante, *HDOP* representa el coeficiente de dilución de la precisión horizontal, y  $\sigma_m$  es el desvío estándar de la muestra.

Para confeccionar un mapa unificado con los valores obtenidos mediante visión estéreo se realiza una conversión de coordenadas del sistema geodésico con DATUM WGS-84[9] al sistema cartesiano XYZ local.

#### H. Descarte de valores atípicos (Outliers)

Con el objetivo de robustecer el algoritmo se implementan rutinas de descarte de valores considerados inválidos por encontrarse por fuera del rango aceptable de desvío respecto del valor central esperado según la Eq. (5)

$$|X - \bar{X}| > 3 \cdot \sigma_m \quad (5)$$

Donde:  $X$  es el valor de la medición y  $\bar{X}$  es el valor esperado de la muestra.

#### I. Filtro y fusión por Kalman

Para fusionar la posición obtenida por el sistema de Visión con la obtenida mediante GPS, se utiliza un filtro de Kalman Extendido (*EKF*)[10] de forma tal que cada medición es pesada de acuerdo a su varianza y esta varianza se actualiza dinámicamente con la información obtenida de los sensores.

En forma adicional el Filtro de Kalman genera una predicción del valor esperado de la medición siguiente, utilizando el modelo dinámico del sistema, en esta aplicación se linealiza el sistema para movimiento rectilíneo uniforme, actualizando las variables en cada iteración [11], esta simplificación está en consonancia con la segunda suposición del algoritmo de LK ya que los incrementos temporales son suficientemente cortos como para suponer que se puede linealizar por tramos la trayectoria del vehículo.

Producto de la fusión la medición combinada posee un valor esperado y varianza según Eq. (6) y (7).

$$\mu_{Fusión} = \frac{\mu_{Visión} \cdot \sigma_{GPS}^2 + \mu_{GPS} \cdot \sigma_{Visión}^2}{\sigma_{GPS}^2 + \sigma_{Visión}^2} \quad (6)$$

$$\sigma_{Fusión}^2 = \frac{\sigma_{GPS}^2 \cdot \sigma_{Visión}^2}{\sigma_{GPS}^2 + \sigma_{Visión}^2} \quad (7)$$

### III. IMPLEMENTACIÓN

#### A. Software

El algoritmo propuesto se desarrolló en C++ utilizando rutinas que implementan funciones de biblioteca de OpenCV (rutinas de visión artificial)[12], PCL(confección del mapa y trayectoria del recorrido en tiempo real), GeographicLib(conversión de coordenadas geodésicas)[13], así como los scripts en código Bash de Linux para lanzar las pruebas, administrar los archivos y analizar los resultados mediante GNU Octave(gráficos)[14].

#### B. Hardware

El hardware del ensayo se compone de una Notebook MSI-PR310 con procesador doble núcleo de 64bits y 4Gb de memoria RAM. Cámaras estereoscópicas Point Grey, Modelo Bumblebee 2[15] conectadas a través de un puerto IEEE 1394/Firewire y un módulo receptor de GPS.

#### C. Ensayos

Se realizaron distintas pruebas de campo que demuestran la eficacia del algoritmo como se verá en la sección IV.

Para las mismas se configuraron las cámaras para trabajar con una resolución de 640x480 trabajando con una tasa de refresco de entre 10 y 20fps, según el compromiso de procesamiento dado por la cantidad de puntos característicos a procesar.

El fabricante del GPS utilizado GM-3 N, asegura una exactitud menor a 3m RMS. La cual será afectada por el coeficiente *HDOP*, según lo visto en II-G.

Los recorridos se realizaron con velocidades máximas permitidas para no perder precisión del orden de 0,5 m/seg y 5 grados/seg para los giros. Mayores velocidades implicarán mayores requerimientos en cuanto a la tasa de refresco de las imágenes, menores tiempos entre muestras válidas de GPS y una mayor capacidad computacional que permita seguir procesando esta información en tiempo real ya que el parámetro  $M$  de la Ec. (3) resulta de los fps de la imagen y la velocidad máxima permitida que garantice encontrar el punto característico en la imagen siguiente.

#### D. Análisis teórico de la precisión del sistema de visión

Teniendo en cuenta la resolución y distancia focal de las cámaras utilizadas. Se realiza un estudio geométrico del sistema según el modelo del haz de rectas de la Fig. 5, los errores que se producen al confundir puntos del entorno cercano y aquellos que se producen según si los cruces bajo estudio (0), son confundidos con los cruces cercanos (1, 2, 3, 4 y 5).

En la Fig. 6 se observa la disminución exponencial de la precisión en el eje X a medida que la distancia crece, debido a la distribución de los cruces de las rectas. En cambio la distribución a lo largo de los ejes Y y Z es cuasi-lineal.

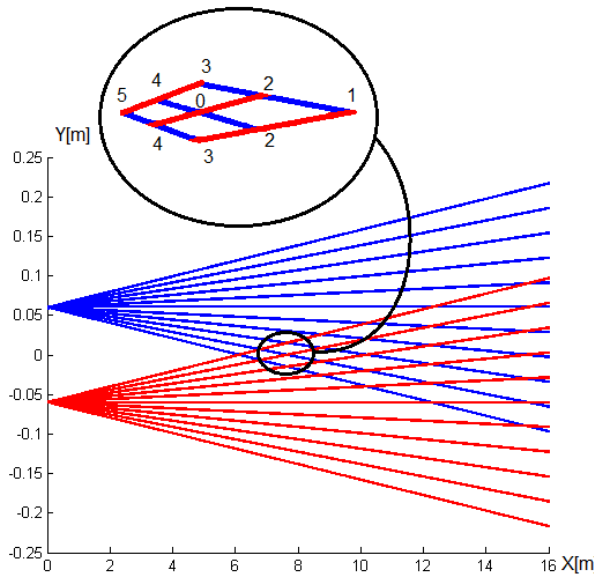


Fig. 5. Vista superior del modelo del haz de rectas,  $z=0$ .

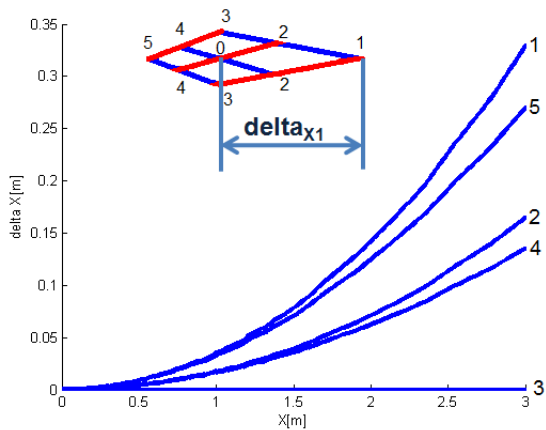


Fig. 6. Errores en la triangulación con cruces cercanos.

A partir del estudio se desprende la conveniencia de trabajar con puntos cercanos para acotar el error en el cálculo del desplazamiento del móvil.

#### IV. RESULTADOS EXPERIMENTALES

##### A. Precisión del sistema de visión artificial

Se realizaron pruebas para los rangos cercanos en los que trabaja el sistema visión. Siendo estos errores acumulativos a medida que las referencias visuales (marcas naturales) son actualizadas conforme avanza el recorrido del móvil. En la Fig. 7 se grafican los errores cometidos al confundir el punto buscado con puntos del entorno cercano, confirmando

la ley de variación exponencial de la precisión expresada en III-D.

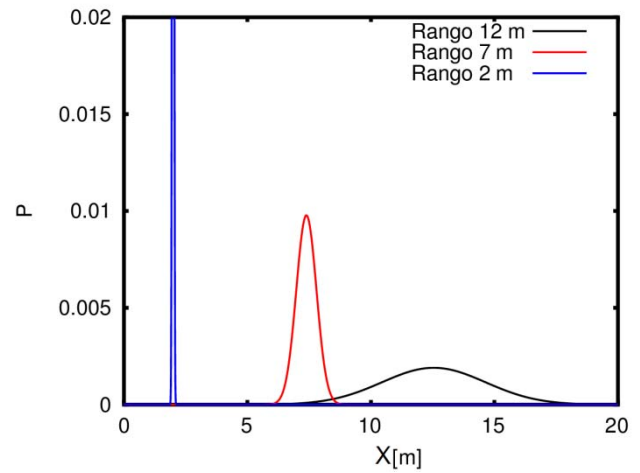


Fig. 7. Distribución de errores de visión en X.

En la TABLA I se expresan los valores de los desvíos estándar de las distribuciones para los distintos rangos.

TABLA I  
DISTRIBUCIÓN DE ERRORES DE VISIÓN

Rango [m]	$\sigma_x$ [m]	$\sigma_y$ [m]	$\sigma_z$ [m]
12m	2,1	0,030	0,041
7m	0,41	0,027	0,036
2m	0,029	0,011	0,015

##### B. Mapeo del recorrido

A continuación se muestran los resultados de las pruebas realizadas en la Ciudad de Buenos Aires, realizando una trayectoria sobre la pista de atletismo de Parque Chacabuco. Fig. 9. Esta trayectoria se realizó sobre el carril interno de la pista teniendo en cuenta que los mojones de medición están referidos a este andarivel.



Fig. 9. Recorrido sobre la pista de atletismo.

En las Fig. 10 y 11 se muestran en un mismo grafico las trayectorias del GPS(Verde), Visión(Rojo) y Kalman(Azul).

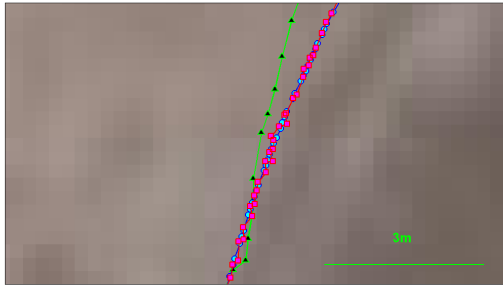


Fig. 10. Detalle del recorrido Escala 3m.

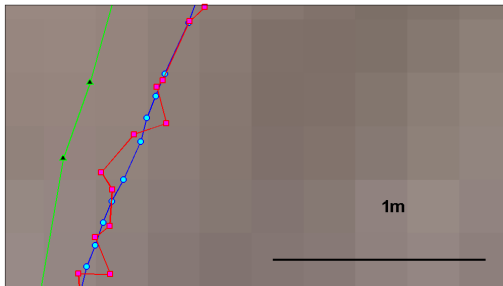


Fig.11. Detalle del recorrido Escala 1m.

### C. Cómputo de la distancia recorrida

En la Tabla II se muestran a modo de referencia una comparación entre las distancias computadas en la trayectoria para GPS, Visión y Kalman, siendo el valor verdadero de la medición 100m.

TABLA II  
DISTANCIAS RECORRIDAS

Sistema	Distancia Recorrida [m]	
	Valor Medio	Desvío Estándar
GPS	119	4,4
Visión	133	18
Kalman	102	1,5

Se desprende de los resultados que el valor calculado por el filtro de Kalman es el que más se ajusta al valor verdadero y es el que menor desvío estándar posee. Se lo atribuye, por un lado a la combinación de ambos sensores ponderando las mediciones de acuerdo a la varianza y por el otro a la inclusión del modelo dinámico del sistema mediante las predicciones sucesivas.

### V. CONCLUSIONES

En este trabajo se muestra que es posible mejorar la precisión diferencial entre puntos de GPS por medio de la fusión con visión estereoscópica. El uso combinado de ambas técnicas permite la mejora de la resolución de las mediciones e incrementar la estabilidad de la navegación ante ocasionales pérdidas de alguna de las mediciones.

La fusión de sensores a través del filtro de Kalman proporciona un método robusto, capaz de funcionar con mediciones ruidosas y pesarlás de acuerdo a su varianza.

Resultado de especial importancia el filtrado de valores atípicos, para descartar aquellos valores que no cumplen con el modelo del sistema.

Con esta técnica se pueden realizar tareas que requerirían una medición GPS diferencial tales como siembra de precisión o exploración geológica y relevamientos topográficos.

### VI. TRABAJOS FUTUROS

Como trabajo futuro se propone la utilización de técnicas de estabilización de imágenes y la incorporación de una unidad de medición inercial IMU.

Se propone como nuevo desafío, el desarrollo de un sistema de navegación reactivo para su implementación en robots móviles, que utilicen el mapeo dinámico provisto por este algoritmo.

### REFERENCIAS

- [1] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in IEEE/RSJ International Conference on Intelligent Robots and Systems, September 2008, pp. 3946 – 3952
- [2] L. Paz, P. Pinies, J. Tardos, J. Neira, Large-scale 6-DOF SLAM with stereo-in hand, IEEE Transactions on Robotics 24 (5) (2008) 946–957.
- [3] J. Shi and C. Tomasi, "Good features to track," 9th IEEE Conference on Computer Vision and Pattern Recognition, June 1994.
- [4] C. Harris and M. Stephens, "A combined corner and edge detector," Proceedings of the 4th Alvey Vision Conference (pp. 147–151), 1988.
- [5] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," Proceedings of the 1981 DARPA Imaging Understanding Workshop (pp. 121–130), 1981.
- [6] H. Hirschmüller, P. R. Innocent and J. M. Garibaldi. Fast, Unconstrained Camera Motion Estimation from Stereo without Tracking and Robust Statistics, In Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision, Singapore, pp. 1099-1104, December 2002.
- [7] GPS - NMEA sentence information: <http://aprs.gids.nl/nmea/> ultima visita 20/12/2015
- [8] R.B. Langley "Dilution of Precision" in GPS World, Vol. 10, No. 5, May 1999, pp. 52–59.
- [9] WGS-84: <http://spatialreference.org/ref/epsg/4326/> ultima visita 20/12/2015
- [10] Julier, S.J.; Uhlmann, J.K., "Unscented filtering and nonlinear estimation," in Proceedings of the IEEE , vol.92, no.3, pp.401-422, Mar 2004 doi: 10.1109/JPROC.2003.823141
- [11] Faragher, R., "Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation [Lecture Notes]," in Signal Processing Magazine, IEEE , vol.29, no.5, pp.128-132, Sept. 2012 doi: 10.1109/MSP.2012.2203621
- [12] Gary Bradski, Adrian Kaehler, "Learning OpenCV: Computer Vision with the OpenCV Library", O'Reilly Media, 1st edition (October 1, 2008).
- [13] GeographicLib library: <http://geographiclib.sourceforge.net/html/> ultima visita 20/12/2015
- [14] GNU Octave: <https://www.gnu.org/software/octave/> ultima visita 20/12/2015
- [15] Bumblebee2 1394a: <https://www.ptgrey.com/bumblebee2-firewire-stereo-vision-camera-systems> ultima visita 20/12/2015