4^to Congreso Argentino de Ingeniería Aeronáutica

POSICIONAMIENTO BASADO EN UNA SECUENCIA DE IMAGENES

Dario F. Mendieta[a], **Marcelo L. Moreyra**[a]

[a] *Departamento de Electrotecnia, Facultad de Ingeniería, Universidad Nacional del Comahue. Buenos Aires 1400, Neuquén CP 8300, Argentina.*

Palabras claves: SLAM visual, Odometría, Extracción de características, Alineamiento de imágenes.

Resumen

Cuando un vehículo no conoce el medio que lo rodea debe construir un mapa de su entorno para posicionarse. Este problema es ampliamente conocido como SLAM (por su siglas en inglés de Simultaneous Localization and Mapping). El hecho de que la mayoría de los seres vivos utilicen la visión para posicionarse y moverse dentro del ambiente, ha motivado en los últimos años el empleo sensores de visión monocular o estéreo en SLAM. Este problema se conoce como SLAM Visual. En esta línea, cuando se prioriza estimar la posición del vehículo antes que la estructura del entorno en pos de una estrategia menos costosa computacionalmente y de tiempo real, el problema se denomina Odometría Visual (VO por su siglas en inglés).

Este trabajo aborda el problema de odometría visual: obtener una estimación confiable de la posición y actitud de un vehículo solamente desde una secuencia de imágenes. Una primera estimación del movimiento se obtiene desde la geometría proyectiva. Para cada par de imágenes consecutivas se extrae un conjunto de características salientes (puntos o líneas) y se agrupan aquellas correspondientes. El seguimiento de estas correspondencias permite obtener una estimación del desplazamiento del móvil. Una segunda aproximación de la posición del vehículo se obtiene operando directamente sobre la intensidad de los pixeles.

Finalmente estos resultados se combinan y se muestra experimentalmente que esta combinación mejora los resultados obtenidos individualmente en ciertas condiciones. Este trabajo no se enfoca en la implementación en tiempo real de los algoritmos. Para el diseño y análisis de los mismos, se utiliza como software principal Matlab, junto con herramientas y librerías de código abierto.

Los resultados experimentales se obtienen utilizando secuencias de imágenes en condiciones reales de ambiente e iluminación disponibles de un dataset público. Este provee datos de diversas trayectorias realizadas por un vehículo dotado con 4 cámaras, un LIDAR 3D y una unidad GPS/IMU. Los datos de telemetría se utilizan como referencia para validar las estimaciones obtenidas de la trayectoria del vehículo.

e-mail: dario.mendieta@fain.uncoma.edu.ar

1. INTRODUCCIÓN

Típicamente, para determinar la ubicación de un vehículo se emplean sensores activos de medición inercial (IMU) o GPS. En el área de robótica móvil cuando además no se conoce el entorno, es posible la navegación y la construcción de un mapa del medio mediante radares y sensores de largo LIDAR. Este problema es ampliamente conocido en la comunidad como SLAM (por su siglas en inglés de Simultaneous Localization and Mapping) [1], así también el filtro EKF (por su siglas en inglés de Extend Kalman Filter) es la solución estándar esencial.

En estos últimos años ha cobrado interés el empleo de cámaras para la localización de un vehículo dado que son sensores no invasivos, de bajo costo, tamaño compacto y un excelente complemento en los intervalos de ausencia de la señal GPS. Debe decirse que una de las desventajas de una cámara, al capturar una proyección de su entorno, es que no proporciona información de rango y es sensible a las condiciones de iluminación.

La estimación de la postura de una cámara, aborda diversos problemas como tipo de cámara a emplear, características extraídas en la imagen, estrategia para la estimación de la postura, representación de los puntos 3D que se reconstruyen e incorporación de telemetría entre otros. Investigaciones realizadas en Estructura desde el Movimiento (SFM por sus siglas en ingles de Struct From Motion) a principios de la década anterior, ya permitían estimar la trayectoria de la cámara y el entorno observado como una nube de puntos 3D, procesando toda la secuencia de imágenes [2]. Sin embargo, el elevado tiempo de cómputo de este procedimiento sumado al del refinamiento recursivo de las estimaciones obtenidas, que se conoce como BA (por su siglas en inglés de Bundle Adjustment), no permiten que el SFM sea una técnica viable para la localización de un vehículo en tiempo real. En el año 2004, Nistér en su trabajo [3] impone el termino Odometría Visual (VO por su siglas en inglés de Visual Odometry) haciendo referencia a las estrategias enfocadas en la localización antes que en la estructura. Su algoritmo estima el movimiento basado en la geometría proyectiva empleada en SFM y puede implementarse en tiempo real para cámaras monoculares y estereros. Por otra parte, Davison con su aporte conocido como enfoque probabilístico [4], extiende la solución EKF del SLAM para cámaras monoculares. Trabajos como [5, 6] enfocan el problema de VO desde las apariencias o intensidad en las imágenes.

El aporte de este trabajo radica en analizar por un lado una estrategia clásica para la localización de una cámara que utilice puntos característicos en la imagen, y que recupere el movimiento desde la geometría proyectiva y por otro lado una estrategia basada directamente en los valores de intensidad en las imágenes que tiene el potencial de un menor costo computacional pero que puede resultar menos precisa por la sencillez de su enfoque. Combinar estas estrategias para conseguir en ciertas condiciones mejorar el resultado individual de cada una de ellas.

Este trabajo se organiza como sigue. En la sección II se describen en detalle las estrategias empleadas para la estimación del movimiento de una cámara. La sección III presenta el enfoque propuesto para combinar las estrategias. La sección IV describe los dos datasets empleados y las condiciones en las que se realizan las simulaciones. La sección V muestra los resultados obtenidos. Finalmente en la sección VI plantea las conclusiones y los pasos futuros en esta línea de trabajo.

2. ESTIMACION DEL MOVIMIENTO

Los parámetros que deben estimarse y que definen el movimiento entre frames consecutivos son una matriz de rotación R y un vector de traslación t. La integración de estas variables a lo largo del tiempo permite reconstruir el recorrido del vehículo. En esta sección se describen los tipos de cámara analizados y el modelo a utilizar. Se presentan además las dos estrategias empleadas para la estimación del movimiento de la cámara.

2.1. Tipo de cámara y modelo a utilizar

Las técnicas de VO pueden dividirse según si se emplea una cámara monocular o una estéreo.

Una cámara estéreo es un arreglo perfecto de dos sensores de visión alineados en donde se conoce la distancia entre sus centros ópticos (baseline). Este arreglo proporciona una medida de rango, es decir que puede calcularse directamente la profundidad de un punto similar a un LIDAR. Esto permite un tratamiento de los puntos en coordenadas 3D como restricción de rigidez de los objetos para la detección de falsas correspondencias, convirtiéndolo en sensor preciso y confiable. La principal desventaja de esta cámara es que al observar objetos cuya profundidad es mucho mayor que su baseline, su comportamiento degenera al caso monocular.

Una cámara monocular es simplemente el arreglo de un único sensor, por lo que es más económico y accesible. Ofrece un reto interesante al no disponer de la información de rango. Con este tipo de sensores es posible estimar el movimiento de la cámara en una escala relativa. Este factor puede corregirse con otro sensor como un sonar o LIDAR que proporcione una medida directa de rango, o si se conocen las dimensiones de los objetos en la imagen.

En la mayoría de los casos, las cámaras emplean un sensor proyectivo de visión y el modelo ampliamente usado es el de proyección pinhole como se muestra en la Figura 1. La imagen se forma por la intersección del rayo que va desde el punto tridimensional $A = [A_x, A_y, A_z]^T$ hacia el centro de la cámara y el plano de la imagen. En la Figura 1 se observan dos sistemas de coordenadas, uno ubicado en el centro de la cámara $(X_c - Y_c)$ y otro en el plano de la imagen con su origen en el vértice superior izquierdo $(u - v)$. Ambos sistemas tienen la misma orientación y difieren en el vector de traslación $[C_x, C_y, -f]^T$. La relación entre el punto tridimensional A y el punto bidimensional $a = [a_u, a_v]^T$ es

$$a_u = f\frac{A_x}{A_z} + C_x, \qquad a_v = f\frac{A_y}{A_z} + C_y \tag{1}$$

El punto principal $[C_x, C_y]^T$ y el foco f definen los parámetros intrínsecos de la cámara y se resumen en la matriz de calibración K [7].

En este trabajo en particular se emplea como sensor de visión una cámara monocular proyectiva por el desafío que presenta y por su bajo costo. Además se considera que la matriz de calibración K es conocida.

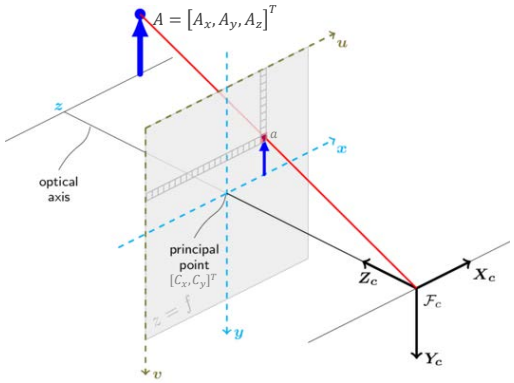

Figura 1: Modelo pinhole para una cámara proyectiva

2.2. Localización basada en puntos característicos

Para la estimación del movimiento de una cámara, una estrategia clásica de VO puede descomponerse en tres partes: una fase de extracción de características en las imágenes (servirán de referencia en la definición del movimiento de la cámara), asociación o correspondencias de estas características y finalmente el método para estimar el movimiento de la cámara.

En la fase de detección, las características que se buscan son puntos o líneas que sean representativos, fuertes, repetibles y robustas, que puedan encontrarse en al menos dos frames consecutivos. Para esto es necesario que la secuencia de imágenes procesadas tenga buena iluminación, textura y detalle. En particular, en la detección de puntos característicos se busca aquellos pixeles que por su brillo, contraste, orientación o gradiente se destacan. Hay una gran diversidad de detectores [8] y la elección final depende de la estructura del ambiente, costo computacional y precisión.

La etapa de correspondencias se realiza entre frames consecutivos y se resuelve calculando la distancia euclidiana entre descriptores. Un descriptor es un vector n-dimensional, que describe el punto de interés detectado en función de la intensidad, orientación o gradiente de su entorno. De igual manera que los detectores, hay una extensa diversidad de descriptores [8] y la elección depende también de la aplicación. En este trabajo se emplea como detector y descriptor de características el SURF (del inglés Speeded-Up Robust Features) por su robustez, eficiencia e invariancia a las rotaciones y escala en la imagen.

Finalmente, con los puntos correspondientes entre frames consecutivos puede estimarse el movimiento de la cámara desde la geometría proyectiva [7]. La Figura 2 muestra el problema que debe resolverse. Sin pérdida de generalidad, se supone que la matriz de proyección para la primer y segunda cámara son $P_1 = K[I|0]$ y $P_2 = K[R|t]$ respectivamente. El problema de la estimación de movimiento se resuelve encontrando la matriz R y el vector de traslación t para cada frame.

La Figura 3 muestra la geometría epipolar para dos vistas. Dado el punto tridimensional A y sus proyecciones a_1 y a_2 se define el plano epipolar. La intersección de este plano con cada plano-imagen define las rectas epipolares l_1 y l_2. Puede verse que cualquier punto en la dirección del rayo C_1 y A va a proyectarse sobre la línea epipolar l_2 en la segunda cámara. Toda esta geometría queda resumida en la matriz fundamental F [7], tal que

$$l_2 = Fa_1 \tag{2}$$

Puede demostrarse que,

$$a_2^T F a_1 = 0. \tag{3}$$

Cuando se conoce la matriz de calibración K, la ecuación (3) se reescribe como,

$$q_2^T E q_1 = 0 \tag{4}$$

Los puntos q son las proyecciones a normalizadas, es decir $q = aK^{-1}$, mientras que a E se la denomina matriz esencial. De [7, 9] se tiene que al realizar la descomposición en valores singulares de la matriz esencial E, surgen cuatro posibles soluciones para el movimiento de la segunda cámara R y t. Para poder elegir el caso correcto, se realiza para cada caso de R y t un proceso de triangulación entre las proyecciones p_{k-1} y p_k. El caso correcto es aquel en donde la mayoría de los puntos triangulados se encuentren delante de ambas cámaras.

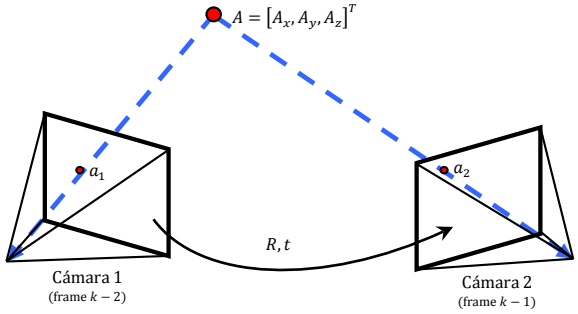

Figura 2: Rotación y traslación en dos vistas

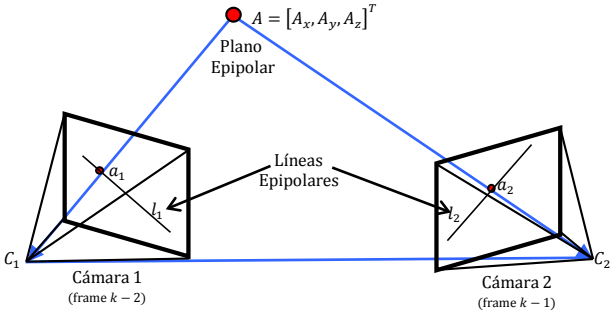

Figura 3: Geometría Epipolar para dos vistas

La estimación obtenida del movimiento entre dos frames tiene el factor limitante de que solamente se conoce la dirección del vector t, no así el módulo. Para solucionar esto y llevarlo a una escala relativa, es necesario encontrar correspondencias entre puntos característicos en un conjunto de al menos tres frames como se muestra en la Figura 4. Con el punto triangulado A_T entre a_1 y a_2 se calcula la transformación que explique el movimiento entre P_T y su punto correspondiente a_3 en el tercer frame. Este problema se conoce como P3P(del inglés Perspective Three Point) [10]. De esta transformación se obtiene la matriz de rotación R' y el vector de traslación t' con escala relativa al primer frame.

De esta manera, el procesamiento de tres frames consecutivos permite estimar el movimiento de la cámara en una escala relativa. Si el escenario recorrido es estático, entonces el movimiento encontrado se debe únicamente a la cámara.

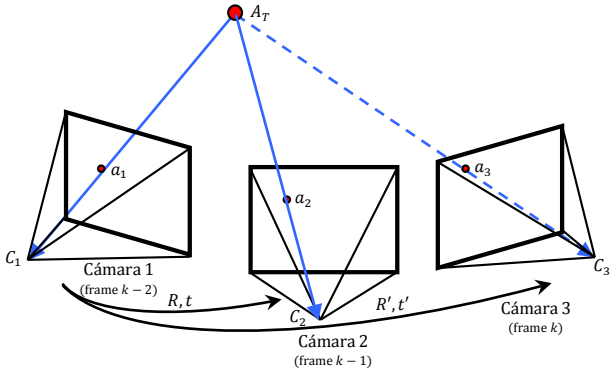

Figura 4: Análisis de 3 vistas. Perspectiva 3D-2D.

2.3. Localización basada en las apariencias

En la sección anterior, se utilizaron puntos característicos en la fase de detección como referencia en la imagen eran puntos característicos. Una forma directa de interpretar la información cruda en una imagen es utilizar la dimensión de intensidades de sus pixeles. Milford en [11] introduce la estrategia en la que utiliza los perfiles de intensidad de una imagen, mostrando un buen desempeño para vehículos terrestres en la estimación de su orientación (yaw). Un perfil de imagen consiste simplemente en la suma de los valores de intensidad en cada columna de la imagen presente (orientación vertical), es decir que se obtiene una señal de tiempo discreto I_n^k asociada al frame k-ésimo de la misma longitud que el ancho de la imagen en pixeles. Su sencillez lo convierte en una herramienta con un gran potencial y escaso costo computacional.

Inicialmente, La estimación del yaw $\hat{\Delta}_{k-1,k}$ en pixeles desde el frame anterior al actual se obtiene de la alineación de sus perfiles correspondientes de la siguiente forma

$$\hat{\Delta}_{k-1,k} = \arg\min_{\Delta} p(\Delta) \tag{5}$$

en donde $p(\Delta)$ es la diferencia entre perfiles para un desplazamiento de Δ pixeles y se define como

$$p(\Delta) = \frac{1}{N - |\Delta|} \sum_{n=1}^{N-\Delta} \left| I_{n+max(\Delta,0)}^{actual} - I_{n-min(\Delta,0)}^{anterior} \right| \tag{6}$$

Finalmente, el yaw en radianes es

$$\psi = \lambda \Delta_{k-1,k} \tag{7}$$

Donde λ es una constante de proporcionalidad que convierte el desplazamiento de pixeles en grados o radianes y se obtiene empíricamente en este trabajo.

La Figura 5 muestra dos frames distintos con su correspondiente perfil y la Figura 6 muestra solamente los perfiles obtenidos. En este caso, el de Δ que mejor ajusta los perfiles es de 9 pixeles.

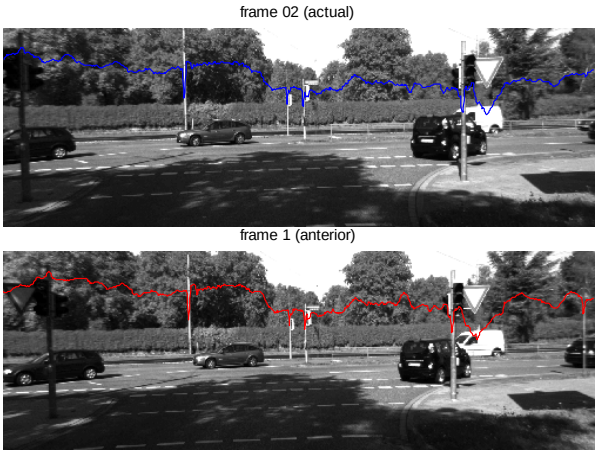

Figura 5: Perfiles de intensidad en su correspondiente frame

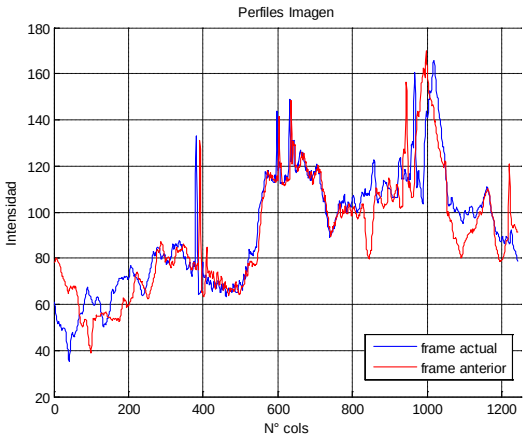

Figura 6: Perfiles para dos frames consecutivos

3. ESTRATEGIA PROPUESTA

La sección anterior presentó dos formas de estimar el movimiento de un vehículo a partir de una cámara solidaria al mismo. Por un lado, la estimación del yaw puede realizarse con una técnica directa y sencilla, procesando dos frames consecutivos $k-1$ y k y buscando el mejor Δ que alinea los perfil-imagen correspondientes. Por otro lado, un enfoque clásico de VO procesa los frames $k-2$, $k-1$ y k iterativamente. Al triangular los puntos correspondientes entre el frame $k-2$, $k-1$ se obtiene la postura $[R|t]$ y una nube de puntos tridimensional. El movimiento local $[R'|t']$ surge de resolver el problema P3P.

En condiciones reales de aplicación, es común que la secuencia de imágenes a procesar presente frames con condiciones de iluminación desfavorables, poca textura y objetos en movimiento entre otros. Esto tiene consecuencias directas sobre los algoritmos para la detección de puntos característicos y descriptores, ocasionando

falsas correspondencias en la etapa de asociación de puntos característicos. Estas falsas correspondencias afectan a su vez la estimación $[R|t]$, la triangulación y el movimiento local $[R'|t']$. Dado que la localización final respecto del primer frame es una acumulación de los movimientos parciales $[R|t]$ y $[R'|t']$, estos errores se acumulan inevitablemente con el tiempo.

Para mitigar este problema la estrategia que aquí se propone es utilizar los puntos triangulados que se obtienen vía un enfoque clásico, y combinarlos con los perfiles-imagen para estimar la matriz de rotación R' con el supuesto de que los ángulos de navegación pitch y roll son nulos. Sean entonces dos puntos tridimensionales $A' = [A'_x, A'_y, A'_z]^T$ y $B' = [B'_x, B'_y, B'_z]^T$ y puntos sus correspondientes $A = [A_x, A_y, A_z]^T$ y $B = [B_x, B_y, B_z]^T$ en el marco de referencia $k-2$ y k respectivamente. Las proyecciones de A y B en el frame k son $a = [a_u, a_v]^T$ y $b = [b_u, b_v]^T$ y $[R'|t']$ es el movimiento del frame $k-2$ al k. Los datos conocidos son los puntos A' y B', sus proyecciones a y b y la matriz de rotación R'. Por ser puntos tridimensionales,

$$\begin{aligned} A &= R'A' + t' \\ B &= R'B' + t' \end{aligned} \tag{8}$$

Reemplazando las ecuación de (8) en (1) queda,

$$\begin{aligned} a_u &= f\frac{A_x}{A_z} + C_x = f\frac{r_1A' + t'_x}{r_3A' + t'_z} + C_x \\ a_v &= f\frac{A_y}{A_z} + C_y = f\frac{r_2A' + t'_x}{r_3A' + t'_z} + C_y \end{aligned} \tag{9}$$

Donde r_1, r_2 y r_3 son las filas de la matriz de rotación R'. Despejando t'_x, t'_y y t'_z de las ecuaciones en (9) puede demostrarse [12] que:

$$\begin{aligned} t'_x &= \frac{(a_u - C_x)A_z}{f} - r_1A', \\ t'_y &= \frac{(a_v - C_y)A_z}{f} - r_2A', \\ t'_y &= A_z - r_3A' \end{aligned} \tag{10}$$

donde

$$\begin{aligned} A_z &= \frac{[fr_1 - (b_u - C_x)r_3](B' - A')}{b_u - a_u} \text{ si } b_u \neq a_u \\ A_z &= \frac{[fr_2 - (b_v - C_y)r_3](B' - A')}{b_v - a_v} \text{ si } b_v \neq a_v \end{aligned} \tag{11}$$

La Figura 7 resume la estrategia propuesta.

Lazo principal
Para los frames $k-2$, k-1 y k:
.- Triangulación entre el frame $k-2$ y $k-1$, para obtener una nube de puntos tridimensionales.
.- Cálculo de matriz de rotación R', entre el frame $k-2$ y k con el método de perfiles.
.- Con dos puntos 3D en el marco de referencia del frame $k-2$ y la matriz de rotación R', calcular el vector de traslación t' por media de la ecuación (8).

Figura 7: Pseudocódigo de la estrategia propuesta

4. SIMULACIONES Y RESULTADOS

Para evaluar experimentalmente los métodos de VO analizados y la validez de la propuesta realizada se emplearon los datasets públicos "The KITTI Vision Benchmark Suite" conocidos en la comunidad de sistemas de visión [13]. Estos datasets fueron generados en diferentes recorridos de la ciudad de Karlsruhe (Alemania) y pueden descargarse en [14]. Se utilizó un vehículo VW Passat equipado con 4 cámaras PointGray Flea2 (2 a color y 2 en

escala de grises), un escáner láser 3D Velodyne HDL-64E, y un sistema de navegación GPS/IMU OXTS RT3003 de alta precisión. La disposición de los sensores en el vehículo está detallada en la Figura 8, que es una reproducción de una imagen en [13]. Los datos de navegación del GPS/IMU se utilizan como referencia para evaluar y validar las estimaciones obtenidas de la trayectoria del vehículo.

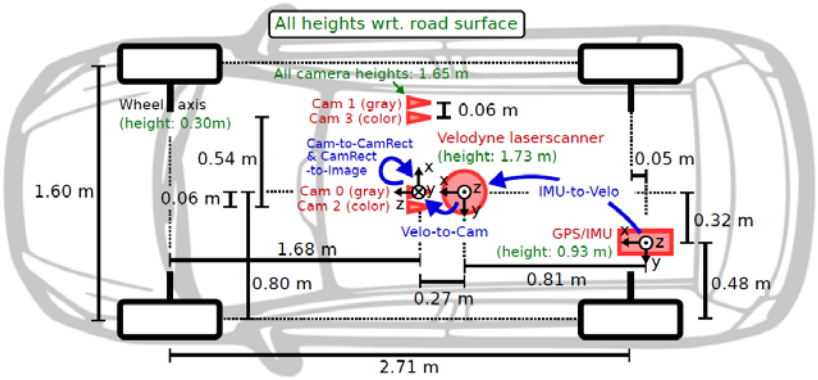

Figura 8: Distribución de los sensores en el vehículo. Reproducción original de la Fig. 3 de [13].

Las estrategias descriptas en la sección 2 y 3 se testearon con varios de los datasets disponibles en [14]. En esta sección se muestran los resultados solamente del dataset que se denomina “2011_09_26_drive_0014 (synced + rectified data)”y está compuesto de 320 frames. Se ubica en la categoría City y describe el recorrido por una ruta principal con gran cantidad de árboles y vehículos en movimiento. Del dataset se emplean solamente las imágenes de la cámara etiquetada como Cam 0 (gray) que está ubicada en el eje central del vehículo.

Para obtener el yaw a partir de los perfiles-imagen se implementan las ecuaciones (5) y (6). El valor de Δ que mejor ajusta se busca en un intervalo de $\pm$ 150 pixeles.

Para la implementación de una técnica de VO clásica como la descripta en la sección 2, se trabaja con la plataforma Matlab. Los parámetros de calibración de la cámara son $f = 721.53$, $C\_x = 609.55$ y $C\_y = 172.85$. Se adopta como sistema de coordenadas de referencia es el de la IMU/GPS. En las etapas de detección, descripción y correspondencia de puntos característicos, cálculo de la matriz fundamental F y el problema P3P se usaron funciones del toolbox de *Computer Vision* de Matlab. El algoritmo para la definición de los 4 casos de $[R|t]$, triangulación e integración final fue escrito por los autores de este trabajo basado principalmente en [7].

La Figura 9 muestra la estimación obtenida del yaw entre frame y frame en pixeles y su valor real obtenido del dataset, y en la Figura 10 se tiene el yaw acumulado en grados desde el inicio del experimento. Claramente puede notarse entre el frame 0 y 50 que el estimador sigue fielmente el movimiento de giro cuando este está cambiando. En la Figura 10, puede observarse alrededor del frame 50 un error de aproximadamente 7° producto de la acumulación de pequeños errores en los anteriores frames.

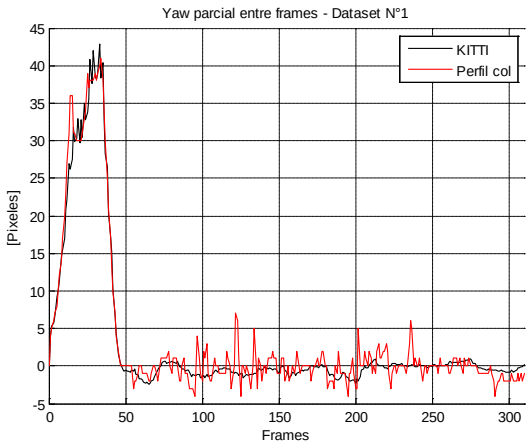

Figura 9: yaw estimado entre perfiles consecutivos

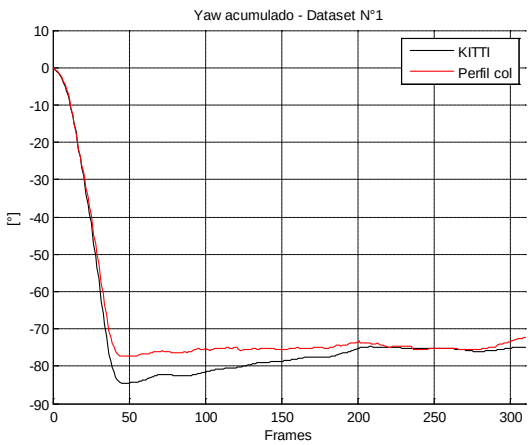

Figura 10: yaw acumulado desde el primer frame

Como se observa en la Figura 8, el movimiento lateral se realiza en el plano X-Y, y el eje Z queda orientado hacia arriba.

Las Figuras 11 y 12 muestran los resultados obtenidos para la estimación de la localización del vehículo a lo largo del tiempo en color verde respecto del dataset de KITTI en color negro. La Figura 11 deja ver que en líneas generales el método VO clásica analizado sigue el movimiento real del vehículo. Para un análisis minucioso, la Figura 12 muestra el error de la estimación respecto a la localización real en cada uno de los ejes de traslación. Se observa un error de hasta 20m en el eje X, 80m en el eje Y y 5m en el eje Z. Hasta el frame 150, el error se mantiene en un valor aceptable. Después, alrededor del frame 160 y 250 se observa errores puntuales debido a errores graves en la etapa de extracción de puntos y correspondencia.

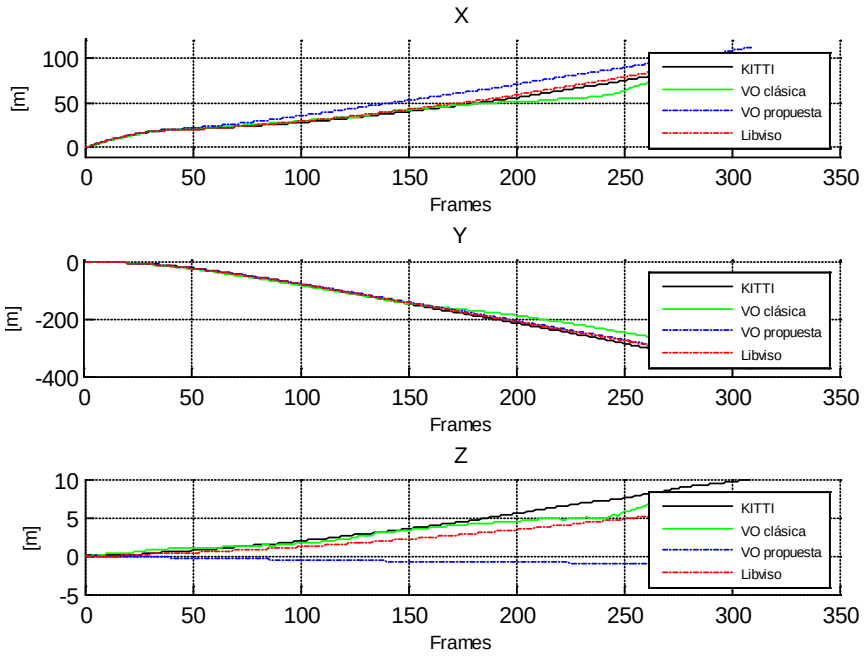

Figura 11: Coordenadas X, Y, Z de localización del vehículo

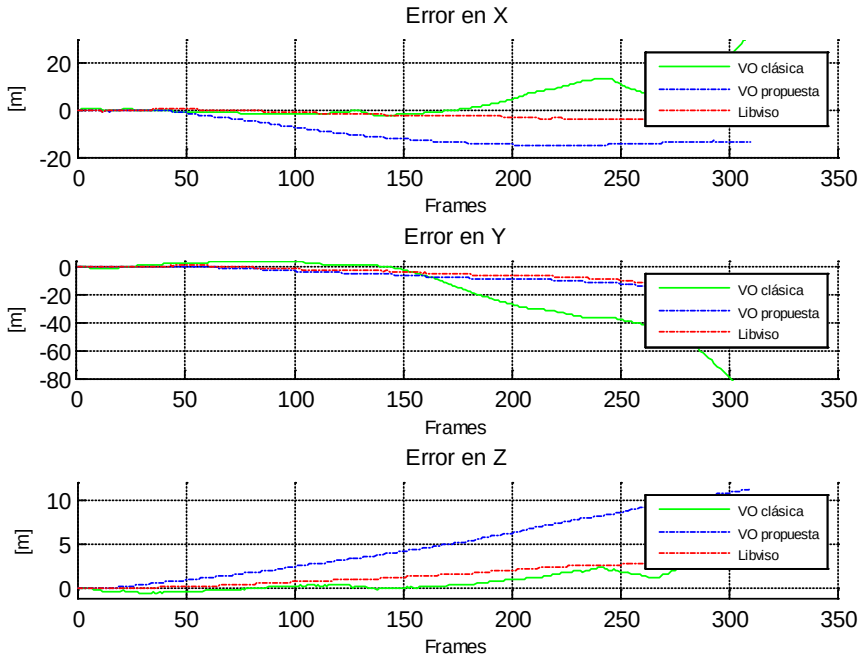

Figura 12: Error en la localización respecto del dataset

Finalmente las Figuras 13 y 14 muestran los resultados obtenidos para la estimación de la orientación del vehículo a lo largo del tiempo en color verde respecto del dataset de KITTI en color negro. Tanto el roll y el pitch

presentan oscilaciones propias de la no uniformidad de la vía transitada. El ángulo de importancia en este trabajo, el yaw, parece estimarse de manera aceptable hasta el frame 150. La implementación de la estrategia propuesta sigue el pseudocódigo descripto en la Figura 7. Los resultados obtenidos se muestran en las Figuras 11 a 14 en color azul. De la Figura 12, el error alcanza los 20m en el eje X, 20m en el eje Y y 10m en el eje Z. A pesar de que sobre el eje X no se disminuye el error, el comportamiento de la estimación no sufre errores grandes y mantiene su tendencia. En el caso del eje Y, el resultado es satisfactorio dado que se consigue bajar la amplitud del error. En lo que respecta al error sobre el eje Z, el aumento notable se atribuye a considerar nulas la orientación de pitch y roll.

Se incorpora como una referencia del potencial detrás de las estrategias de odometría visual, los resultados del toolbox libviso2 [15] en color rojo en las Figuras 11 a 14. Con esta herramienta, el problema de la estimación del módulo del vector traslación se resuelve con puntos que se estima están en el camino.

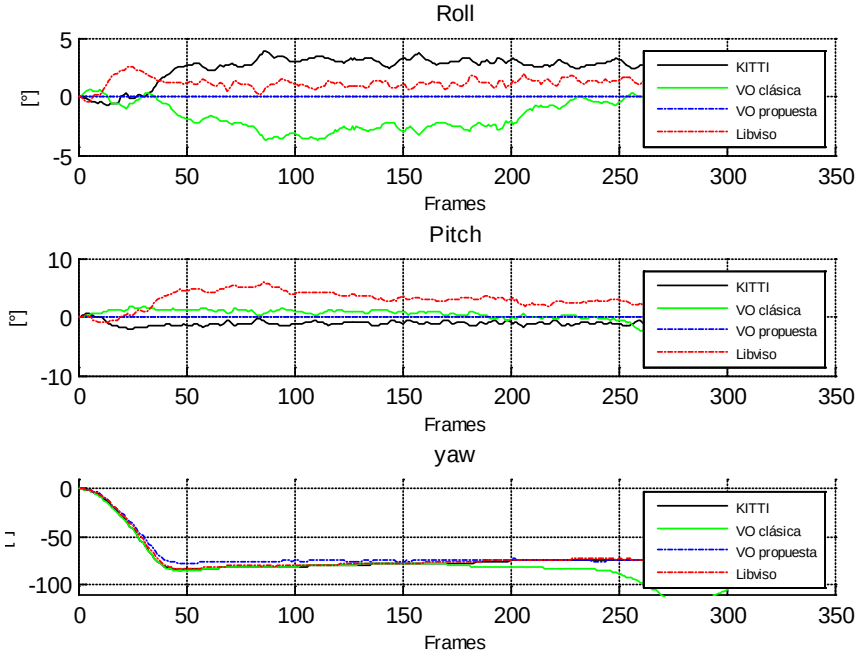

Figura 13: Orientación del vehículo

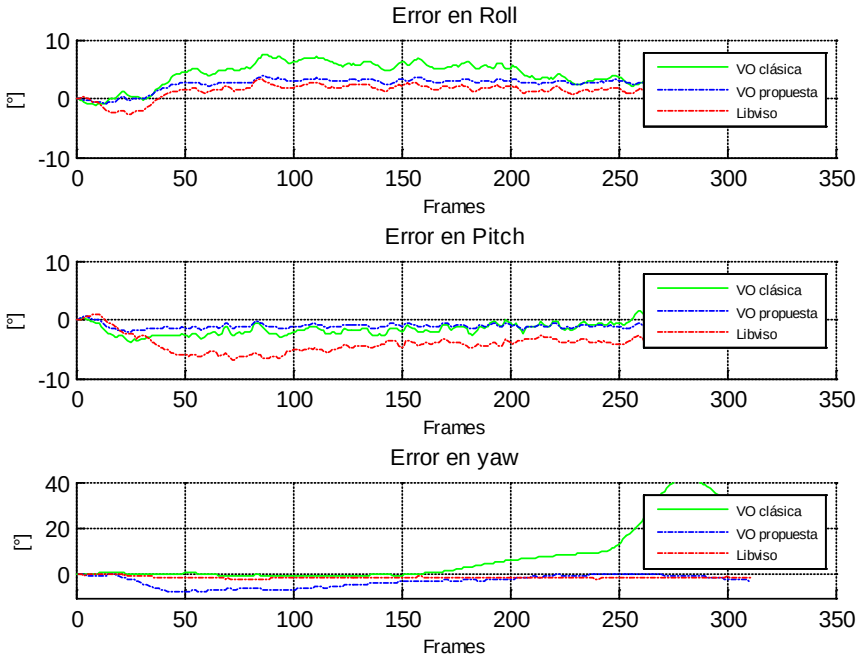

Figura 14: Error en la orientación respecto del dataset

5. CONCLUSIONES

En este trabajo se analiza una estrategia de VO clásica y otra directa basada en perfiles-imagen para la estimación del movimiento de un vehículo a partir de una cámara monocular. La propuesta del trabajo es combinar los puntos tridimensionales que puede generar una técnica VO clásica con la estimación de la matriz de rotación R obtenida por perfiles-imagen y calcular el vector de traslación. Los resultados experimentales presentados se obtuvieron a partir de datos públicos reales. Se demostró la validez de la propuesta tomando como referencia la información de la orientación y localización del vehículo obtenida del sistema GPS/IMU. De esta forma se consigue minimizar el efecto de la propagación de errores en la triangulación hacia la etapa de estimación de $[R|t]$, y a su vez la acumulación de errores a lo largo del tiempo.

Los trabajos a futuro incluyen minimizar los efectos de las falsas correspondencias en la etapa de triangulación y explorar el uso de los perfiles-imagen para la estimación del pitch. También es necesaria la incorporación de una etapa que permita evaluar la utilidad de un frame para la detección del movimiento, caso contrario lo descarte y espere un frame más significativo.

REFERENCIAS

[1] Durrant-Whyte, Hugh; Bailey, Tim. Simultaneous localization and mapping: part I. IEEE robotics & automation magazine, 2006, vol. 13, no 2, p. 99-110.

[2] Fitzgibbon, Andrew W.; Zisserman, Andrew. Automatic camera recovery for closed or open image sequences. En European conference on computer vision. Springer Berlin Heidelberg, 1998. p. 311-326.

[3] Nistér, David; Naroditsky, Oleg; Bergen, James. Visual odometry. En Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004. p. I-652-I-659 Vol. 1.

[4] Davison, Andrew J. Real-time simultaneous localisation and mapping with a single camera. En Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003. p. 1403-1410.

[5] Smith, Devin; Dodds, Zachary. Visual navigation: Image profiles for odometry and control. Journal of Computing Sciences in Colleges, 2009, vol. 24, no 4, p. 168-179.

[6] Scaramuzza, Davide; Siegwart, Roland. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. IEEE transactions on robotics, 2008, vol. 24, no 5, p. 1015-1026.

[7] Hartley, Richard; Zisserman, Andrew. Multiple view geometry in computer vision. Cambridge university press, 2003.

[8] Fraundorfer, Friedrich; Scaramuzza, Davide. Visual odometry: Part II: Matching, robustness, optimization, and applications. IEEE Robotics & Automation Magazine, 2012, vol. 19, no 2, p. 78-90.

[9] Nistér, David. An efficient solution to the five-point relative pose problem. IEEE transactions on pattern analysis and machine intelligence, 2004, vol. 26, no 6, p. 756-770.

[10] Haralick, Bert M., et al. Review and analysis of solutions of the three point perspective pose estimation problem. International journal of computer vision, 1994, vol. 13, no 3, p. 331-356.

[11] Milford, M. J. and Wyeth, G. F. Single Camera Vision-only SLAM on a Suburban Road Network, in Proceedings, 2008 Int. Conf. on Robotics and Automation, Pasadena, CA, USA, May 19-23, pp. 3684-3689, 2008.

[12] Merckel, Loic; Nishida, Toyoaki. Evaluation of a method to solve the perspective-two-point problem using a three-axis orientation sensor. En Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on. IEEE, 2008. p. 862-867.

[13] Geiger, Andreas, et al. Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research, 2013.

[14] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "The KITTI Vision Benchmark Suite," Disponible en http://www.cvlibs.net/datasets/kitti, 2012.

[15] Geiger, Andreas; Ziegler, Julius; Stiller, Christoph. Stereoscan: Dense 3d reconstruction in real-time. En Intelligent Vehicles Symposium (IV), 2011 IEEE. IEEE, 2011. p. 963-968.