

EVALUACIÓN DE DESCRIPTORES LOCALES EN LOCALIZACIÓN VISUAL CON IMÁGENES OJO DE PEZ

María Flores, David Valiente, Arturo Gil, Adrián Peidro, Oscar Reinoso, Luis Payá
 Departamento de Ingeniería de Sistemas y Automática
 Universidad Miguel Hernández, Avenida de la Universidad s/n 03202 Elche (Alicante), España
 {m.flores, dvaliente, arturo.gil, apeidro, o.reinoso, lpaya}@umh.es

Resumen

Un robot móvil debe ser capaz de localizarse en el entorno en el que se encuentra para navegar de forma autónoma. Para tal fin, en este trabajo se utiliza un método denominado Adaptive Probability-Oriented Feature Matching (APOFM). Este algoritmo consiste en obtener la pose relativa a partir de información visual de un par de imágenes, como un algoritmo de odometría visual. La principal característica de este método es que realiza un filtrado de los puntos característicos detectados utilizando información proporcionada por un modelo 3D del entorno. En dicho modelo, cada punto tiene asignada una probabilidad de existencia de correspondencia. Con esto se consigue un menor número de falsos positivos. Los objetivos principales de este trabajo son evaluar este método con distintos tipos de características locales y comparar su efectividad con un método de odometría visual estándar, además de analizar los resultados empleando imágenes tomadas con una cámara con lente de ojo de pez. Tras los distintos experimentos, la solución más precisa al estimar la pose relativa se ha obtenido con el método APOFM extrayendo la información visual con ORB.

Palabras clave: Adaptive Probability-Oriented Feature Matching, odometría visual, localización, características locales.

1. INTRODUCCIÓN

Con el fin de conseguir una navegación autónoma eficaz, un robot móvil debe ser capaz de resolver el problema de localización. En la literatura, se han empleado varios tipos de sensores y técnicas [1] para obtener la posición y orientación de un robot móvil de forma precisa. De entre todos los tipos de sensores que pueden ir a bordo del robot móvil, los sistemas de visión han recibido un gran interés. Este hecho se debe a que la información que proporciona al robot móvil puede aplicarse tanto para resolver el problema de localización como para realizar otras tareas de navegación autónoma.

En cuanto a las técnicas de localización que utilizan la información que ofrece este tipo de sen-

sor, se encuentra la odometría visual [12]. En este método, la pose se estima de forma incremental a partir de los cambios que el movimiento causa en las imágenes [4]. Las ventajas que presenta este enfoque son varias, por ejemplo, que no se ve afectado por el posible deslizamiento de las ruedas y que puede utilizarse para resolver el problema de localización en diferentes tipos de robots, no solo en los robots móviles con ruedas. Un ejemplo de ello se puede ver en [17] donde utilizan un proceso de odometría visual estéreo a partir de imágenes que han sido tomadas a bordo de un vehículo submarino autónomo.

Cabe señalar que el empleo de las cámaras con gran campo de visión tiene muchas ventajas en una gran cantidad de aplicaciones robóticas. Las cámaras omnidireccionales pueden capturar imágenes con un campo de visión de 360° alrededor del robot. Hay diferentes configuraciones posibles para obtener imágenes omnidireccionales [11] aunque los sistemas más utilizados son los catadióptricos y de ojo de pez. Un sistema catadióptrico está compuesto por una cámara de perspectiva convencional con un espejo convexo. De esta manera, se genera una vista completa de 360 grados (una esfera completa). Por ejemplo, Román et al. [8] muestran el desarrollo y evaluación de un enfoque de agrupamiento incremental para obtener modelos jerárquicos compactos del entorno utilizando un sistema de visión catadióptrico como fuente de información. Otra forma de aumentar el campo de visión es mediante la combinación de una lente ojo de pez y una cámara de perspectiva convencional. Por ejemplo, Matsuki et al. [7] proponen un método que extiende la odometría dispersa directa para utilizar la imagen completa incluso con una fuerte distorsión. Por ello, evalúan el enfoque utilizado con imágenes de ojo de pez.

La principal diferencia entre los sistemas de visión comentados en el párrafo anterior es que un sistema formado por una única lente de ojo de pez presenta un menor campo de visión en comparación con un sistema catadióptrico. A pesar de tal hecho, resulta interesante evaluar un algoritmo de odometría visual utilizando imágenes de este tipo de cámara ya que dispone de algunas característi-

cas relevantes comparado con el catadióptrico, como su reducido tamaño y ligereza. Además, el sistema de visión catadióptrico es estructuralmente más complejo.

Para resolver el problema de odometría visual, es necesario extraer y relacionar la información más relevante de las imágenes. En este trabajo, se emplea un algoritmo de odometría visual propuesto en trabajos de investigación previos [15], llamado Adaptive Probability-Oriented Feature Matching (APOFM). La finalidad de esta propuesta es resolver el problema de localización basado en el método estándar de odometría visual (en adelante, se mencionará como SVOM, Standard Visual Odometry Method) pero utilizando información de probabilidad asociada a la existencia de puntos característicos dentro del entorno. Dicha información es proporcionada por un modelo de escena que establece relaciones entre puntos 3D con alta probabilidad de existencia y sus proyecciones sobre un par de imágenes. De esta manera, estas proyecciones definen áreas de las imágenes donde es más probable que aparezcan puntos con correspondencias. El APOFM mejora el procesamiento de imágenes (detección y descripción de características y búsqueda de correspondencias) en SVOM, obteniendo una búsqueda de coincidencias más robusta y un rechazo de valores atípicos. De esta manera, la solución de localización obtenida es más precisa.

En el trabajo anterior [15] se evaluó este método utilizando las imágenes capturadas por un sistema catadióptrico y utilizando solo puntos característicos SURF para extraer la información visual. Teniendo en cuenta esto, el presente estudio emplea diferentes tipos de detectores y descriptores de puntos característicos (SURF, ORB, FAST y KAZE) y evalúa la influencia del tipo de característica sobre el rendimiento de la odometría visual utilizando APOFM con imágenes tomadas con una cámara de lente de ojo de pez. En la sección dedicada a los experimentos (Sección 4), se analizan tanto los resultados del proceso de búsqueda de correspondencias como la precisión de la odometría visual en la estimación de la posición y orientación del robot, y estos resultados se comparan con los obtenidos con un método de odometría visual estándar. Como ya se ha comentado, otro objetivo de los experimentos es analizar el comportamiento de este método con imágenes de lente de ojo de pez. Para lograr tal fin, hemos utilizado un conjunto de imágenes de este tipo disponible públicamente [19].

Este trabajo se encuentra estructurado de la siguiente forma: los diferentes tipos de características locales utilizados en este trabajo se presentan

en la Sección 2; en la Sección 3 se describe el método empleado; los resultados obtenidos durante los experimentos se muestran en la Sección 4; y, finalmente, las conclusiones se exponen en la Sección 5.

2. CARACTERÍSTICAS LOCALES

Para extraer y describir la información relevante de una imagen, se pueden emplear principalmente dos tipos de métodos: características globales o locales. En cuanto a los descriptores de apariencia global, cada imagen se describe con un único vector. Se espera que este descriptor sea invariable frente a cambios globales. Por ejemplo, Amorós et al. [3] presentan una comparación de técnicas de descripción de apariencia global (incluyendo el uso de la información de color) para creación de mapas y localización utilizando solo la información proporcionada por imágenes omnidireccionales.

Por el contrario, las características locales son patrones o estructuras (por ejemplo, puntos, bordes o pequeñas áreas) presentes en una imagen. Difieren de su vecindario inmediato en términos de intensidad, color y/o textura [13]. En este sentido, Valiente et al. [14] comparan los resultados de un método de odometría visual con imágenes omnidireccionales extrayendo la información visual con estas dos técnicas.

Las citadas características locales se pueden considerar como la combinación de un detector de características y un descriptor. Los detectores se utilizan para encontrar las características esenciales de la imagen, mientras que los descriptores describen las características extraídas y generan un vector descriptivo. Hay varios tipos de características locales propuestas en la literatura. Joshi y Patel [6] exponen una evaluación de distintos métodos de detección y descripción.

En este trabajo, hemos usado los siguientes tipos de características locales: SURF [5] (basado en blobs y con el vector descriptor real), FAST [9] (esquinas y con vector descriptor binario), ORB [10] (esquinas y con vector descriptor binario) y KAZE [2] (blobs y con vector descriptor real).

3. MÉTODO APOFM

APOFM consiste en resolver el problema de localización basado en un algoritmo de odometría visual estándar, pero incorporando información de probabilidad proporcionada por un modelo de la escena. Este modelo consiste en una distribución de probabilidad que caracteriza dinámicamente la aparición de correspondencias encontradas en ite-

raciones anteriores. La técnica empleada para este propósito es el Proceso Gaussiano (GP) [16]. En la Figura 1 se muestra el digrama de APOFM con los distintos pasos que lo componen. En la primera iteración ($t = 1$) del algoritmo, dado que todas las proyecciones de los puntos 3D de la escena tienen la misma probabilidad de ser una correspondencia, la pose relativa se estima resolviendo el SVOM. De este modo, primero se detectan los puntos característicos en cada imagen (I_0 e I_1), después se extraen los vectores descriptores de todos los puntos detectados y, posteriormente, se buscan correspondencias en función de la similitud entre descriptores. La medida de similitud utilizada para los descriptores de características binarias es la distancia Hamming y, por el contrario, para otros formatos de descriptores es la distancia euclídea al cuadrado. Por lo tanto, de estos tres pasos se ha obtenido un conjunto de correspondencias 2D a 2D. El último paso es estimar el movimiento relativo usando la geometría epipolar y el conjunto de puntos obtenido en el paso anterior. Dado que para emplear la matriz esencial es necesario conocer los parámetros intrínsecos de la cámara, se ha realizado el proceso de calibración previamente. De esta forma, al ejecutar el algoritmo, los valores de estos parámetros son conocidos.

En este trabajo, la pose relativa se expresa mediante cinco parámetros angulares ($\theta, \gamma, \alpha, \phi, \beta$) y un factor de escala (ρ). Los tres primeros parámetros angulares están asociados con la orientación (θ, γ, α). Los otros dos (ϕ, β), junto con el factor de escala (ρ), están asociados a la traslación expresada en coordenadas esféricas, donde ρ es la distancia relativa entre los centros de ambas cámaras (excepto por un factor de escala), ϕ es el ángulo en el plano x-y y β el ángulo de elevación desde dicho plano. Los valores de los cinco parámetros angulares son los que se estiman en este trabajo y se muestran en la sección correspondiente a los experimentos. Después de esta primera iteración, ya se ha obtenido información sobre correspondencia de características. De esta forma, es posible obtener el modelo y emplear esta información en el algoritmo.

3.1. MODELO 3D DE PROBABILIDAD

Para poder crear el modelo, cuyos pasos se pueden ver en la Figura 2, en primer lugar, se deben recuperar las coordenadas 3D de cada par del conjunto de correspondencias 2D a 2D con los que se ha estimado la pose anterior. A este ejercicio se le conoce como problema de triangulación. En este sentido, dado un par de imágenes, si los puntos característicos coincidentes son en realidad la proyección del mismo punto 3D, sus rayos deben intersectar en este punto 3D. Sin embargo, este hecho no siempre ocurre debido a la presencia de varios tipos de ruido (por ejemplo, error debido a que los parámetros de calibración no son precisos o ruido durante la detección de características). Por lo tanto, el problema de triangulación se reduce a encontrar la mejor solución, por ejemplo, utilizando el método del punto medio donde se supone que el punto 3D es el punto medio de la perpendicular común a ambas líneas 3D. No obstante, las dos líneas 3D no se cruzan en algunos casos porque la coincidencia de este par de puntos característicos es un falso positivo, lo que significa que no son la proyección del mismo punto 3D, aunque sus descriptores sean similares y por ello han sido asociados incorrectamente durante el paso de búsqueda coincidente.

Para poder evaluar la efectividad del APOFM y SVOM con respecto a los falsos positivos, hemos agregado un bloque, denominado contador de falsos positivos. En este bloque, dado un punto 3D, se vuelve a proyectar sobre la segunda imagen, empleando el modelo de la cámara. Si está cerca del punto característico con el que se han obtenido (junto con su correspondencia en la primera imagen) sus coordenadas 3D, significa que el punto característico es la proyección de este punto 3D y, por lo tanto, el par de puntos coincidentes es un verdadero positivo. En caso contrario, estamos ante un falso positivo. De este modo, se calcula la distancia entre estos dos puntos, como se puede ver en la Figura 2, y si es mayor a un determinado valor (a) se trata de un falso positivo.

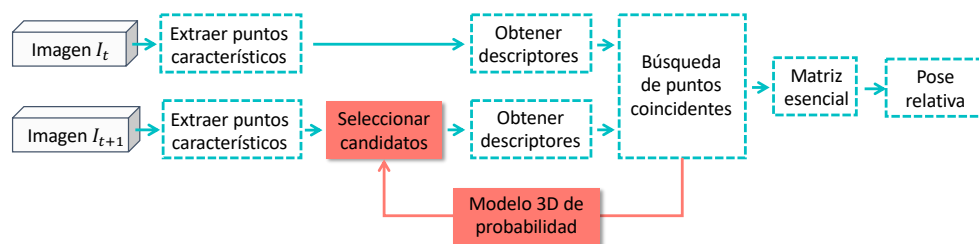


Figura 1: Diagrama del algoritmo del APOFM. Los bloques de APOFM añadidos al SVOM se han resaltado con otro color (rojo).

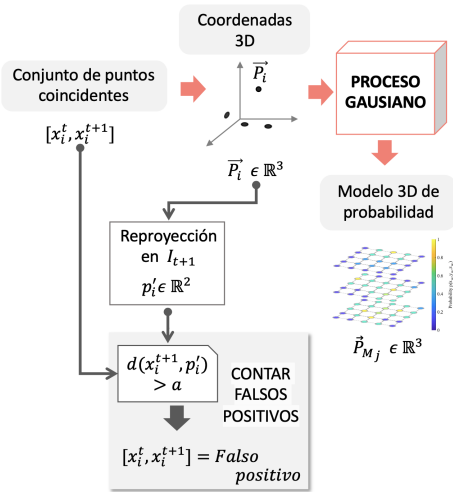


Figura 2: Crear del modelo 3D y contar falsos positivos.

Después de resolver el problema de triangulación, el siguiente paso es crear el modelo de la escena. Para realizar esta tarea se ha recurrido al GP. El bloque GP recibe datos de entrada de entrenamiento que corresponden al conjunto de coordenadas 3D y datos de salida de entrenamiento que son un vector de unos que indican que la proyección de estos puntos 3D en el par de imágenes ha sido considerado correspondientes. Además, hay un conjunto de puntos de test que son los puntos 3D que definen el modelo. La salida del GP es la media y la covarianza de la distribución condicional predicha para los puntos de prueba. Como el objetivo es crear un modelo de probabilidad de la escena, la predicción debe tomar valores entre cero y uno. Para este fin, se emplea una función logística (sigmoidea). Finalmente, el mapa global se actualiza usando Bayesian Committee Machine (BCM).

3.2. SELECCIÓN DE PUNTOS CANDIDATOS

Una vez el modelo de escena 3D con probabilidad es creado ($t = 1$) o actualizado ($t > 1$), el robot móvil se mueve a una nueva posición, y se toma una imagen I_{t+1} , de la cual se extraen los puntos característicos. Después, se proyecta la información de probabilidad 3D en esta nueva imagen.

Para proyectar de 3D a 2D y viceversa, se deben conocer tanto los parámetros intrínsecos como extrínsecos. Los primeros se han obtenido previamente con el proceso de calibración. Por contra, los segundos se estiman empleando el modelo de vehículo y aplicando la transformación de la estructura del robot móvil al sistema de referencia

de la cámara (se conoce porque la cámara está en todo momento instalada en la misma posición a bordo del robot móvil). Estos datos de odometría solo se utilizan para el mapeo de puntos 2D a 3D y de 3D a 2D.

En esta etapa del algoritmo, se tienen dos conjuntos de puntos en coordenadas de píxel: uno con información de imagen (puntos característicos) y otro con probabilidad basada en correspondencias (proyección del modelo 3D de la escena). El paso de selección de puntos candidatos viene dado por una búsqueda del punto más cercano en el segundo conjunto a cada punto del primero. Esta búsqueda se basa en una medida métrica, concretamente la distancia City-block, y la técnica para encontrar al vecino más cercano es el algoritmo del árbol kd. Un punto característico se considerará candidato si la distancia calculada es menor que un umbral específico (χ) cuyo valor está dado por la inversa de la función de distribución acumulativa de chi-cuadrado. Si el punto característico se clasifica como candidato para encontrar una coincidencia en la imagen I_t , a dicho punto se le asocia la probabilidad del punto proyectado más cercano. De esta forma, los puntos candidatos se pueden filtrar de acuerdo con el valor de probabilidad asociado, obteniendo un conjunto de puntos característicos candidatos cuya probabilidad de encontrar una coincidencia es mayor que una probabilidad mínima (ρ_{min}). En la Figura 3, se muestra el esquema de selección de puntos candidatos.

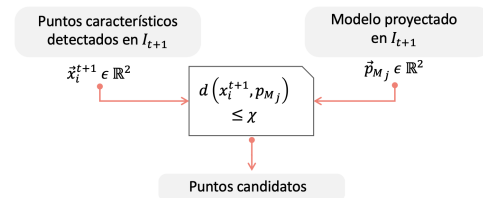


Figura 3: Clasificación de los puntos característicos como candidatos.

Los siguientes pasos son buscar correspondencias, obtener la matriz esencial y extraer la pose relativa, que corresponden al algoritmo de odometría visual estándar. Sin embargo, la diferencia radica en que de la segunda imagen únicamente se extraen los descriptores de los puntos clasificados como candidatos, ya que solo estos puntos buscarán sus puntos coincidentes en la primera imagen.

4. EXPERIMENTOS

Como se indica en la Sección 3, APOFM estima la pose relativa a partir de puntos de características locales. Teniendo en cuenta esto, los experimen-

tos realizados en este trabajo tienen dos objetivos principales: (a) evaluar el comportamiento del APOFM con varios tipos de características locales para determinar cuál de ellos proporciona una estimación más precisa de la pose relativa; y (b) realizar una comparación entre APOFM y SVOM con el fin de evaluar la mejora lograda con el método APOFM. Por tanto, se han realizado un total de ocho pruebas como resultado de la combinación de los dos métodos y las cuatro características locales.

Para evaluar la influencia del tipo de características, hemos realizado un estudio sobre los siguientes aspectos: el número de características detectadas con cada tipo de puntos característicos, cuántos de ellos han encontrado coincidencia en la otra imagen y los falsos positivos (Sección 4.1); el error al resolver el problema de localización (Sección 4.2); y el tiempo de cálculo (Sección 4.3). En cada figura, el resultado que se muestra es el promedio de los valores obtenidos con cada par de imágenes. Todos los experimentos se han realizado con un PC con CPU Intel Core i7-10700 R a 2.90GHz y Matlab 2019 como software.

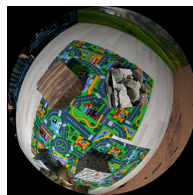


Figura 4: Ejemplo de una de las imágenes disponibles en la base de datos.

Con respecto a las imágenes, hemos utilizado un conjunto de datos público [19], que proporciona, a su vez, un conjunto de imágenes tomadas con una cámara de lente de ojo de pez (ver Figura 4) y un archivo de salida con las posiciones de la cámara en las que ha capturado cada imagen (ground truth). Se compone de un total de 200 imágenes con una resolución de 640x480 píxeles.

4.1. NÚMERO DE PUNTOS Y CORRESPONDENCIAS

Como se ha mencionado, el algoritmo de APOFM, así como el de SVOM, a partir de un par de imágenes, detectan puntos característicos locales en cada una y obtienen un conjunto de correspondencias entre ellos que se emplean para resolver el problema de localización. Por lo tanto, esta parte del algoritmo será lo que estudiaremos en esta subsección.

En el eje X de la Figura 5, se muestra la cantidad de puntos característicos detectados en función del método empleado para su detección y descripción

(es independiente al método escogido para resolver el problema de odometría). El eje Y de esta misma figura representa el número de estos puntos que finalmente han encontrado correspondencia en la otra imagen. El número de coincidencias sí depende del método de odometría empleado, por lo que el número obtenido con cada uno está representado con un color distinto. Tras analizar la Figura 5, podemos concluir que ORB y KAZE ofrecen el mayor número de correspondencias. Por el contrario, SURF proporciona el menor número de coincidencias con ambos métodos. En el caso de KAZE, se obtienen más correspondencias de puntos, aunque la cantidad de puntos característicos detectados no sea tan elevada como al emplear ORB o FAST. Por lo que respecta a los métodos de odometría, podemos observar que el algoritmo SVOM encuentra más coincidencias que APOFM. Esto último era predecible, ya que el segundo método no utiliza todos los puntos característicos detectados en este paso, sino solo aquellos que se han considerado candidatos a encontrar sus coincidentes debido a su probabilidad de existencia.

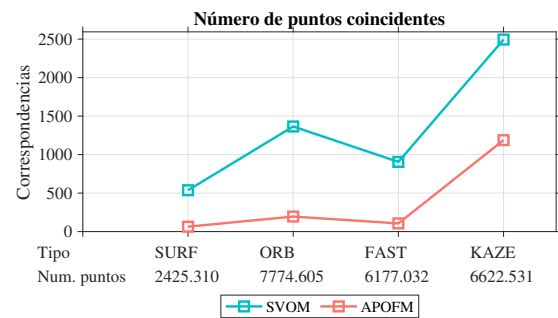


Figura 5: El número medio de puntos detectados se muestra en la parte inferior del eje X y en el eje Y se representa el número medio de ellos que han encontrado correspondencia.

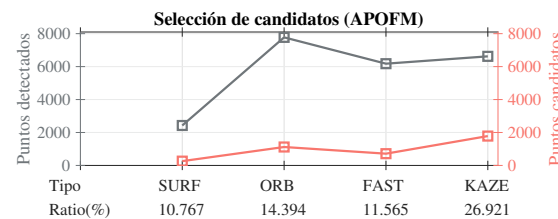


Figura 6: Número medio de puntos clasificados como candidatos.

En la Figura 6, se muestra el número de puntos característicos detectados y cuántos de ellos se han considerado como candidatos, así como el ratio entre dichos valores, únicamente para el algoritmo APOFM. De esta forma, podemos extraer que el número de puntos característicos que se consideran durante la búsqueda de correspondencia es menor que el conjunto de características inicial ya

que se ha filtrado por probabilidad de existencia.

Además de lo mencionado, es importante analizar cuántos de estos pares de correspondencias son verdaderos positivos y cuántos son falsos positivos. En este sentido, la Figura 7 muestra el valor medio y la desviación estándar del número de falsos positivos en cada caso. Con APOFM, el valor medio y la desviación estándar es menor en todos los casos excepto con KAZE. De esta forma, podemos decir que con este método se consigue un menor número de valores atípicos en la búsqueda de correspondencias en ambas imágenes.

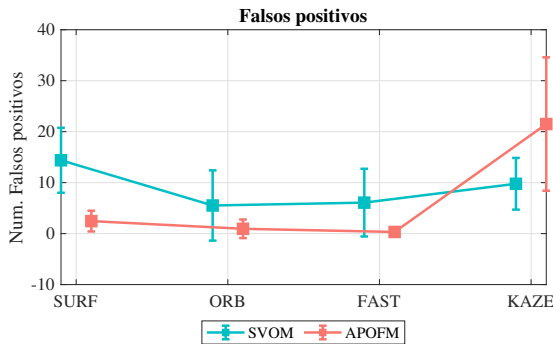


Figura 7: Falsos positivos.

4.2. ESTIMACIÓN DE LA POSE RELATIVA

No se puede obviar que el objetivo es estimar la pose relativa con la mayor precisión posible. Por lo tanto, se debe estudiar el error obtenido con cada combinación y así ver cómo funciona el método de localización dependiendo del tipo de puntos característicos utilizados como entrada. Para calcular el error, se ha utilizado el ground truth proporcionado por la base de datos. Como se ha comentado con anterioridad, el movimiento relativo se descompone en una traslación definida por dos parámetros angulares ϕ y β (Figura 8(a) y Figura 8(b), respectivamente) y una rotación que viene dada por θ , γ y α (Figura 9(a), Figura 9(b) y Figura 9(c)).

Analizando las figuras correspondientes a la traslación, podemos observar que el error es mayor con SURF, como cabía esperar considerando los resultados obtenidos en la anterior subsección, es decir, menor precisión y menor número de puntos detectados y coincidencias. En cuanto a la Figura 8(a), el método APOFM presenta un error menor con respecto a SVOM en todos los casos, excepto cuando se utiliza KAZE, aunque la diferencia de error entre ambos métodos es pequeña. No obstante, en relación con la desviación estándar, APOFM presenta mejores resultados. De esto se puede concluir que la mejor solución de localiza-

ción se ha obtenido con la combinación del método APOFM y ORB, estando el error en ϕ en torno a 4° . En cambio, APOFM proporciona el menor error, con independencia del tipo de característica, estimando el parámetro β . En cuanto a la rotación, APOFM estima la orientación con más precisión que SVOM. Del mismo modo que en la traslación, la mejor solución de localización, esto es, el menor error, se obtiene con ORB.

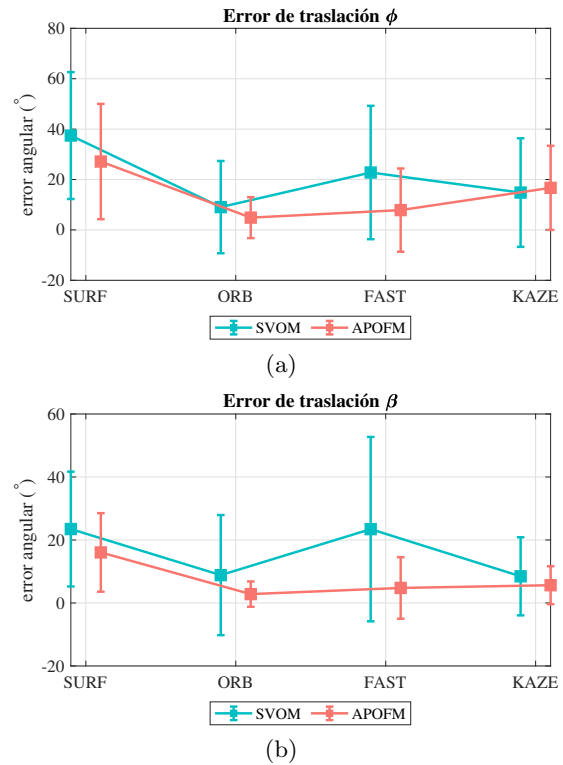


Figura 8: Error calculando (a) ϕ y (b) β .

4.3. TIEMPO DE CÁLCULO

Finalmente, también es necesario analizar el tiempo de cálculo computacional requerido en cada una de las combinaciones estudiadas en las subsecciones anteriores. En el Cuadro 1 se muestra la media de del tiempo de estimación de la pose relativa a partir de cada par de imágenes.

Cuadro 1: Tiempo medio de cálculo en segundos.

	SURF	ORB	FAST	KAZE
SVOM	2,34s	14,66s	13,64s	2,04s
APOFM	1,26s	3,51s	2,68s	12,05s

El tiempo transcurrido durante la estimación de la pose relativa es menor para APOFM cuando las características locales son SURF, ORB y FAST. Esto se debe a que, en el caso de APOFM, el número de puntos característicos correspondientes

a I_{t+1} que tiene que encontrar una coincidencia es menor que en el caso de SVOM, pues los puntos característicos han sido filtrados (candidatos). De esta manera, el tiempo asociado con este paso y las características SURF, ORB y FAST también es más reducido, salvo cuando se usa KAZE.

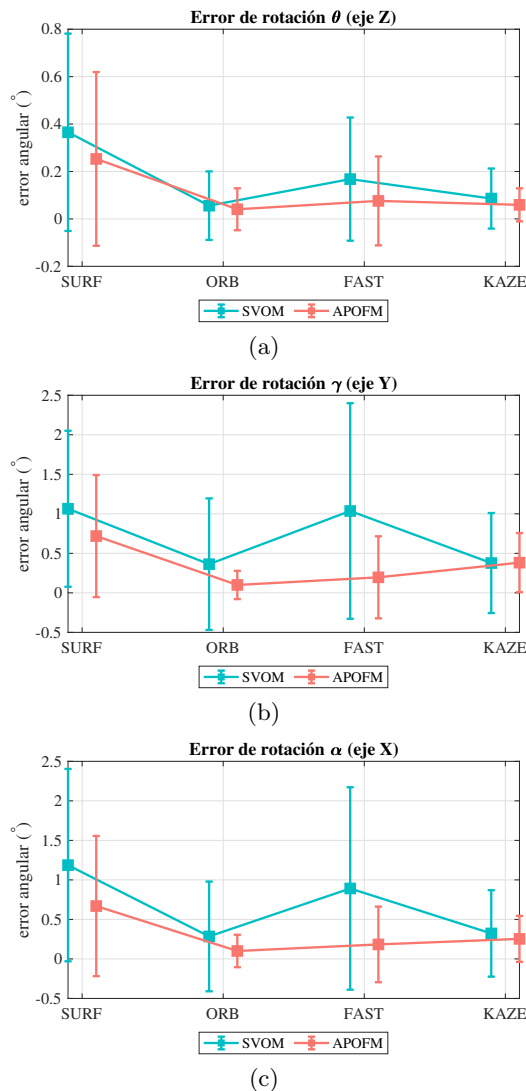


Figura 9: Error calculando (a) θ , (b) γ y (c) α .

5. CONCLUSIONES

En este trabajo, el problema de localización se resuelve utilizando una base de datos con imágenes ojo de pez disponibles públicamente. Hemos estudiado el comportamiento del método APOFM al utilizar distintos detectores y descriptores de características locales. APOFM modela el entorno de forma dinámica con puntos 3D que presentan cierta probabilidad de existencia de puntos característicos. En este sentido, las correspondencias de características se pueden encontrar en áreas determinadas de las imágenes. Dicho modelo dinámico

se obtiene empleando una técnica de aprendizaje automático, el GP.

En [15], los autores evalúan este método utilizando imágenes tomadas por un sistema de visión catadióptrico y características SURF únicamente. Por el contrario, en el presente análisis se ha examinado el comportamiento de la técnica de localización mediante un total de ocho pruebas como resultado de combinar un algoritmo (el SVOM o el APOFM) y un tipo de característica local (SURF, ORB, FAST o KAZE). Para cada caso, hemos estudiado varios aspectos.

Tras analizar los resultados, podemos concluir que el método APOFM ha superado considerablemente al SVOM con respecto a la solución de localización y el tiempo de cálculo cuando las características locales son SURF, ORB y FAST. La diferencia entre ambos métodos cuando usan KAZE es insignificante, y el error al usar el método APOFM es un poco mayor. La combinación del método APOFM y ORB proporciona una localización más precisa (el error se encuentra alrededor de los 4° para el parámetro de traslación ϕ), así como un tiempo de cálculo más reducido en comparación con SVOM.

En resumen, el problema de localización que se resuelve usando el método APOFM se ha mejorado empleando otros tipos de puntos característicos, concretamente ORB. Como trabajo futuro, sería interesante evaluar este método con otras características locales, como ASIFT [18], que es invariante a transformaciones afines. Además, otra investigación futura podría consistir en ampliar estos resultados comparativos a otros modelos de imágenes no lineales. Dado que este método se basa en un modelo de la escena, sería interesante también evaluar este método en exteriores, ya que podría presentar algunas limitaciones en cuanto a la creación del modelo 3D (el entorno no está tan acotado como en interiores).

Agradecimientos

Este trabajo ha sido apoyado por la Generalitat Valenciana mediante la beca con referencia ACIF/2020/141.

English summary

EVALUATION OF LOCAL DESCRIPTORS IN VISUAL LOCALIZATION WITH FISHEYE IMAGES

Abstract

To navigate autonomously, a mobile robot must be able to localize itself in the environment.

To that end, in this work, a method called Adaptive Probability-Oriented Feature Matching (APOFM) is used. This algorithm obtains the relative pose from visual information (a pair of images), as a visual odometry algorithm. The main feature of this method is that it performs a filtering of the feature points using the information provided by a 3D model of the environment, in which each point is assigned a probability of existence of correspondence. This results in fewer false positives. The main objectives of this work are to evaluate this method with different types of local features and compare its effectiveness with the standard visual odometry method, and also to analyze the results using images taken by a fisheye lens camera. After the different experiments, the most precise solution estimating the relative pose was obtained with the APOFM method, extracting the visual information with ORB.

Keywords: Adaptive Probability-Oriented Feature Matching, visual odometry, localization, local features.

Referencias

- [1] ALATISE, M. B., AND HANCKE, G. P. A review on challenges of autonomous mobile robot and sensor fusion methods. *IEEE Access* 8 (2020), 39830–39846.
- [2] ALCANTARILLA, P. F., BARTOLI, A., AND DAVISON, A. J. KAZE features. In *Computer Vision – ECCV 2012* (Berlin, Heidelberg, 2012), A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., Springer Berlin Heidelberg, pp. 214–227.
- [3] AMORÓS, F., PAYÁ, L., MAYOL-CUEVAS, W., JIMÉNEZ, L. M., AND REINOSO, O. Holistic descriptors of omnidirectional color images and their performance in estimation of position and orientation. *IEEE Access* 8 (2020), 81822–81848.
- [4] AQEL, M. O. A., MARHABAN, M. H., SARIPAN, M. I., AND ISMAIL, N. B. Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus* 5, 1 (2016), 1897.
- [5] BAY, H., ESS, A., TUYTELAARS, T., AND VAN GOOL, L. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110, 3 (2008), 346–359. Similarity Matching in Computer Vision and Multimedia.
- [6] JOSHI, K., AND PATEL, M. I. Recent advances in local feature detector and descriptor: a literature survey. *International Journal of Multimedia Information Retrieval* 9, 4 (2020), 231–247.
- [7] MATSUKI, H., VON STUMBERG, L., USENKO, V., STÜCKLER, J., AND CREMERS, D. Omnidirectional DSO: Direct Sparse Odometry With Fisheye Cameras. *IEEE Robotics and Automation Letters* 3, 4 (2018), 3693–3700.
- [8] ROMÁN, V., PAYÁ, L., CEBOLLADA, S., AND REINOSO, Ó. Creating incremental models of indoor environments through omnidirectional imaging. *Applied Sciences* 10, 18 (2020).
- [9] ROSTEN, E., AND DRUMMOND, T. Machine learning for high-speed corner detection. In *Computer Vision – ECCV 2006* (Berlin, Heidelberg, 2006), A. Leonardis, H. Bischof, and A. Pinz, Eds., Springer Berlin Heidelberg, pp. 430–443.
- [10] RUBLEE, E., RABAUD, V., KONOLIGE, K., AND BRADSKI, G. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision* (2011), pp. 2564–2571.
- [11] SCARAMUZZA, D. *Omnidirectional Camera*. Springer US, Boston, MA, 2014, pp. 552–560.
- [12] SCARAMUZZA, D., AND FRAUNDORFER, F. Visual odometry [tutorial]. *IEEE Robotics Automation Magazine* 18, 4 (2011), 80–92.
- [13] TUYTELAARS, T., AND MIKOLAJCZYK, K. Local invariant feature detectors: A survey. *Foundations and Trends® in Computer Graphics and Vision* 3, 3 (2008), 177–280.
- [14] VALIENTE, D., FERNÁNDEZ, L., GIL, A., PAYÁ, L., AND REINOSO, O. Visual odometry through appearance-and feature-based method with omnidirectional images. *Journal of Robotics* 2012 (2012).
- [15] VALIENTE, D., PAYÁ, L., JIMÉNEZ, L. M., SEBASTIÁN, J. M., AND REINOSO, Ó. Visual information fusion through bayesian inference for adaptive probability-oriented feature matching. *Sensors* 18, 7 (2018).
- [16] WILLIAMS, C. K., AND RASMUSSEN, C. E. *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.
- [17] WIRTH, S., CARRASCO, P. L. N., AND CODINA, G. O. Visual odometry for autonomous underwater vehicles. In *2013 MTS/IEEE OCEANS-Bergen* (2013), IEEE, pp. 1–6.
- [18] YU, G., AND MOREL, J.-M. ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line* 1 (2011), 11–38.
- [19] ZHANG, Z., REBECQ, H., FORSTER, C., AND SCARAMUZZA, D. Benefit of large field-of-view cameras for visual odometry. In *2016 IEEE International Conference on Robotics and Automation (ICRA)* (2016), pp. 801–808.



© 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).