

Realidad Aumentada Egocéntrica basada en Panoramas e Interacción Gestual

Propuesto por: Dr. Javier Flavio Viguera Gómez

Áreas: Visión Computacional en Tiempo Real, Realidad Aumentada

El objetivo de la tesis es implementar un método de localización visual en tiempo real basado en la observación de imágenes panorámicas 360°, ya sea en interiores o exteriores, para aplicaciones de Realidad Aumentada (RA) (ver Fig. 1). Una imagen panorámica de la escena real será adquirida y sus puntos de interés serán detectados y descritos antes del proceso de localización. El sistema de referencia inicial será elegido por el usuario en la primera imagen al inicio del proceso de localización. Ya en línea, todas las imágenes adquiridas por la cámara serán registradas en tiempo real con respecto al sistema de referencia inicial mediante la imagen panorámica.

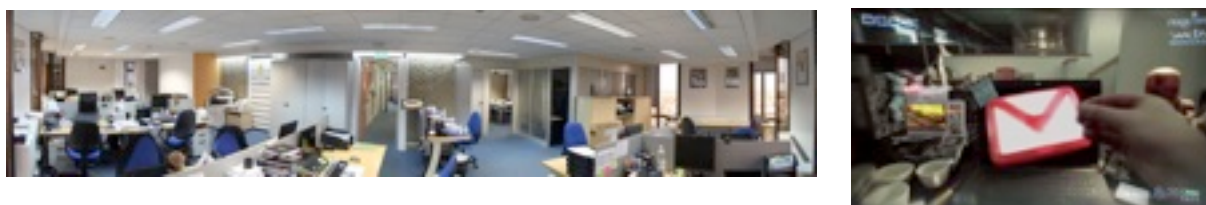


Fig. 1. A la izquierda: ejemplo de una imagen panorámica en el interior de una oficina. A la derecha: ejemplo de una imagen aumentada [MagicLeap] y la aparición de la mano del usuario para la interacción gestual.

Uno de los retos del proyecto es partir del hecho de que los parámetros de calibración de la cámara pudieran no ser conocidos (o correctos) y deberán entonces ser adquiridos y/o corregidos automáticamente conforme el procedimiento continúa su ejecución, incluida la distorsión radial [1].

Los sistemas egocéntricos [2] actuales asumen que el centro óptico de la cámara está permanentemente en la posición exacta desde la que fue adquirida la imagen panorámica, y que el usuario realiza rotaciones puras con la cámara. No obstante, esta condición es crítica, restrictiva y muy difícil de mantener en la práctica, lo que hace al sistema de RA altamente sensible a errores de aumentación incluso si el usuario se desplazó mínimamente de la posición inicial donde se adquirió el panorama. Sin embargo, aún es posible determinar las correspondencias entre las imágenes actuales y las de referencia panorámicas, por lo que se busca cumplir dos objetivos específicos respecto a la etapa de localización:

(1) Analizar la sensibilidad de la localización visual y de las imágenes aumentadas en términos de cada uno de los parámetros intrínsecos y extrínsecos de la cámara y del mapa de profundidad de la escena (usando un modelo conocido o una cámara RGB-D para estimar el *ground truth*) bajo la hipótesis de que el centro óptico se mantuvo constante a pesar de traslaciones de éste mediante homografías rotacionales.

(2) En base al análisis anterior, corregir en línea la posición del centro óptico y proponer un modelo de localización basado en homografías planas por pedazos que reduzcan localmente los errores.

Considerando que el centro óptico de la cámara es coincidente con la vista del usuario (por ejemplo, con un casco o lentes de realidad aumentada o un dispositivo móvil) la interacción gestual se realizará mediante el movimiento de las manos del usuario frente al campo de visión de la cámara. Esto es, en la segunda etapa, se desarrollará un método de segmentación capaz de segmentar las manos del usuario respecto al escenario y las imágenes aumentadas deberán ser consistentes con las oclusiones generadas por éstas, además de reconocer un diccionario básico de gestos de interacción.

Puesto que se busca que ambas tareas se efectúen en tiempo real, la segmentación se realizará mediante un método piramidal para acelerar el proceso y, la localización será híbrida entre el empleo de puntos de interés acelerados (FAST) sobre la imagen original y super-píxeles [3] si predominan regiones homogéneas grandes, decisión que debe ser evaluada en línea y de forma automática.

[1] Galego, Ricardo, Alexandre Bernardino, and José Gaspar. "Auto-Calibration of Pan-Tilt Cameras Including Radial Distortion and Zoom." *Advances in Visual Computing*. Springer Berlin Heidelberg, 2012. 169-178.

[2] Lovegrove, Steven, and Andrew J. Davison. "Real-time spherical mosaicing using whole image alignment." *Computer Vision—ECCV 2010*. Springer Berlin Heidelberg, 2010. 73-86.

[3] Concha, Alejo, and Javier Civera. "Using superpixels in monocular SLAM." *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014.