

Método para detección y seguimiento de objetos con aplicaciones en Realidad Aumentada

Christian Nicolás Pfarher

Punto de control N° 1 - Recopilación Bibliográfica

Director

Dr. Enrique Marcelo Albornoz

Codirector

Dr. César Martínez

29 de junio de 2012



Ingeniería Informática
Facultad de Ingeniería y Ciencias Hídricas
UNIVERSIDAD NACIONAL DEL LITORAL

1. Conceptos claves de Realidad Aumentada

La Realidad Aumentada permite enriquecer la perspectiva sobre imponiendo objetos virtuales (2D o 3D) en el mundo real, con el objetivo de lograr persuadir al observador y haciéndole creer que el objeto virtual forma parte del ambiente real. De esta forma, se puede interpretar la Realidad Aumentada, como una mezcla entre un mundo real y virtual tal como se visualiza en el diagrama Reality-Virtuality (RV) continuum diagram (Diagrama continuo de Realidad-Virtualidad) mostrado en la Fig. 1.

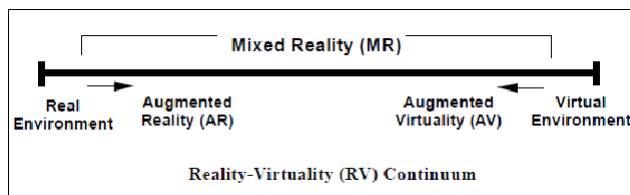


Figura 1: Reality-Virtuality Continuum (Diagrama continuo de Realidad-Virtualidad), Milgram, Takemura, Itsumi and Kishino [MTUK94]

Algunos de los conceptos que se manejan en Realidad Aumentada son los siguientes:

Reality View (Vista Real) Se refiere a la secuencia o stream de video producido por la cámara. La aplicación de Realidad Aumentada captura imágenes desde esta secuencia de video, aumentándolas con objetos virtuales para crear una “vista aumentada”. [But11]

Registration & Tracking (Identificación y seguimiento) Describe los métodos disponibles para alinear un objeto virtual con un sistemas de coordenadas en tres dimensiones en una vista real. En el seguimiento o Tracking, se pueden distinguir entre diferentes métodos: **location based tracking (seguimiento basado en localización)** y **optical tracking (seguimiento basado en visión)** o una combinación de ambos. [H07], [But11]

Point of interest-POI (Punto de interés) Es un elemento de información asociado con una ubicación geográfica (longitud, latitud, altura) o un patrón visual (marcador, tapa de libro, etc) que se puede representar de alguna forma por la aplicación de Realidad Aumentada. [But11], [ABB⁺01]

Virtual Object (Objeto Virtual) Es algún tipo de contenido digital que es dibujado por la aplicación de Realidad Aumentada y es sobre impuesto en la vista real. Por lo general, incluye modelos 3D, imágenes 2D, iconos, textos, entre otros. [H07], [But11]

Marker Based and Markerless AR (AR basados/no basados en marcadores) Cuando es utilizado el reconocimiento de imágenes para alinear un objeto virtual (optical tracking (seguimiento basado en visión)), hay una distinción entre sistemas: los que identifican un **marcador artificial** (código matricial 2D o diodo emisor de luz) colocado en un lugar determinado o sobre algún objeto del mundo real; los que utilizan la **detección de características naturales de los objetos**, para identificar objetos del mundo real inalterados (como tapas de libros, posters, etc.) los cuales no poseen marcas artificiales que asistan en el reconocimiento del objeto. En ambos casos, el objeto virtual 2D o 3D, aparece “pegado” al marcador o característica natural cuando se ve a través de la ventana gráfica de AR o Viewport. [H07], [But11]

Location Based Tracking (Reconocimiento basados en localización) Se refiere al reconocimiento basado en información de geo-localización, obtenida a partir de dispositivos o sensores de localización. El término, es usado para hacer una distinción entre sistemas que usan

sensores de localización, en contraste con los que hacen tracking (seguimiento) del objeto usando técnicas de reconocimiento de imagen. El reconocimiento basado en localización, posee por lo general menos precisión que los métodos ópticos y es aplicado sobretodo en ambientes exteriores. [But11], [ABB⁺01]

Six Degrees of Freedom (6DoF) (Seis grados de libertad) Se refiere a la capacidad del sistema de seguimiento para mantener la alineación de un objeto del mundo real en un espacio tridimensional. Esto es, la capacidad de moverse hacia delante/atrás, arriba/abajo, izquierda/derecha (traslación en tres ejes perpendiculares), combinados con la rotación sobre tres ejes perpendiculares (guiñada, cabeceo, alabeo).

Reconstruir la posición de la cámara con 6DoF es el objetivo más importante en AR, ya que de esta manera se determina la posición de la cámara y en consecuencia se obtiene la posibilidad de poder renderizar o dibujar los objetos virtuales en una perspectiva correcta. [H07], [But11], [ABB⁺01]

Referencias de la sección: [MTUK94], [But11], [H07], [ABB⁺01], [LLLC11].

2. Puntos característicos y Descriptores

La detección de puntos característicos (del inglés: keypoints o feature points) y descriptores de imágenes, es una característica ampliamente usada en problemas de reconocimiento de objetos, detección de imágenes, seguimiento visual, reconstrucción 3D entre otras aplicaciones. El objetivo recae en la idea de seleccionar ciertos puntos claves en una imagen, de forma que estos permitan distinguirla de otras sin la necesidad de tener que analizar la totalidad de la imagen. Esta aproximación funciona correctamente, en la medida que se detecten la cantidad suficiente de puntos de interés que sean “distinguibiles” y a su vez estables para que puedan ser nuevamente localizados en otra escena.

Existen diversos detectores de puntos característicos, descriptores y seguidores. Entre los algoritmos y métodos más destacados se pueden nombrar:

- Harris Corner Detector (Detector de esquinas de Harris) [BK08], [Lag11], [TM07],
- FAST: Features From Accelerated Segment Test (Detección rápida de puntos característicos) [Lag11],
- SIFT: Scale Invariant Feature Transform (Transformación de características invariante a la escala) [BK08], [Lag11], [TM07], [MS05], [WRM⁺10], [SL04], [OBG11], [LLLC11],
- PCA-SIFT: Principal Components Analysis - Scale Invariant Feature Transform (Análisis de componentes principales - Transformación de características invariante a la escala) [BK08], [MS05],
- SURF: Speeded Up Robust Features (Detector rápido de características robustas) [Lag11], [BETV08], [TM07], [BETV08], [OBG11],
- KLT: Kanade–Lucas–Tomasi feature tracker (Seguidor de características Kanade-Lucas-Tomasi) [BK08], [TM07], [SL04], [LLLC11],
- L-K optical flow (Flujo óptico LK) [BK08], [Lag11], [SL04],
- SLAM: Simultaneous Localization And Mapping (Localización y Mapeado Simultáneos) [hCMW08].

Otras técnicas ampliamente usadas en diferentes líneas de investigaciones junto a los métodos nombrados arriba son: K-D tree (árbol k-dimensional) [Lag11] y RANSAC: RANdom SAmple Consensus (muestras aleatorias consensuadas) [Lag11].

Referencias de la sección: [BK08], [Lag11], [BETV08],[TM07], [MS05], [WRM⁺10], [LS09], [hCMW08],[BETV08], [SL04], [OBG11],[LLLC11].

3. Operaciones morfológicas en imágenes

Los filtros morfológicos son operadores que transforma una imagen de una forma predefinida (de acuerdo a la forma de la máscara aplicada). La interacción entre el filtro y los vecinos al píxel sobre el cuál se está aplicando el procesamiento, dan el resultado final de la transformación.

Técnicas como erosión, dilatación, detección de bordes y esquinas, umbralización, entre otras, son algunas clases de este tipo de operaciones que a veces son usadas como parte de un pre-procesamiento sobre las imágenes, para su posterior tratamiento.

Referencias de la sección: [BK08], [Lag11], [JH00].

4. Homografía y proyecciones

Las imágenes generalmente son producidas usando una cámara, que captura la escena mediante la proyección de luz en un sensor a través de una lente. El hecho de que una imagen se forma a través de la proyección de una escena 3D en un plano 2D, impone la existencia de una importante relación entre la escena, la imagen y diferentes imágenes de la misma escena. La geometría proyectiva, es la herramienta usada para describir y caracterizar, en términos matemáticos, el proceso de formación de la imagen.

Algunos de los conceptos fundamentales en las relaciones proyectivas involucran temas tales como:

- el modelo pin-hole de la cámara [BK08], [Lag11],
- la calibración de la cámara que arroja como resultado los parámetros de distorsión y la matriz que la caracteriza [BK08], [Lag11], [ABB⁺01],
- Epipolos y líneas epipolares [BK08], [Lag11],
- Matriz fundamental [BK08], [Lag11], [ABB⁺01],
- Matching de imágenes usando RANSAC [BK08], [Lag11],
- Computo de homografía entre imágenes: una homografía, permite determinar la posición relativa de la cámara respecto a una escena. Es un mapeo uno a uno entre dos imágenes, el cual es definido por ocho parámetros. [BK08], [Lag11], [CD10].

Referencias de la sección: [BK08], [Lag11], [CD10], [ABB⁺01], [LLLC11].

Referencias

- [ABB⁺01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Comput. Graph. Appl.*, 21(6):34–47, November 2001.
- [BETV08] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding: CVIU*, 110(3):346–359, June 2008.
- [BK08] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O’Reilly Media, 1st edition, October 2008.
- [But11] Ben Butchart. Augmented reality for smartphones - a guide for developers and content publishers. In *JISC Observatory*, 2011.
- [CD10] Daniel Conrad and Guilherme N. DeSouza. Homography-based ground plane detection for mobile robot navigation using a modified EM algorithm. In *ICRA* [CD10], pages 910–915.
- [H07] Taehee Lee 0002 and Tobias Höllerer. Handy AR: Markerless inspection of augmented reality objects using fingertip tracking. In *ISWC* [H07], pages 83–90.
- [hCMW08] Ian Y h Chen, Bruce Macdonald, and Burkhard Wünsche. Markerless augmented reality for robots in unprepared environments, 2008.
- [JH00] B. Jahne and H. W. Haussecker. *Computer Vision and Applications: A Guide for Students and Practitioners*. Academic Press, April 2000.
- [Lag11] Robert Laganière. *OpenCV 2 Computer Vision Application Programming Cookbook*. Packt Publishing, June 2011.
- [LLLC11] Ahyun Lee, Jae-Young Lee, Seok-Han Lee, and Jong-Soo Choi. Markerless augmented reality system based on planar object tracking. In *Frontiers of Computer Vision (FCV), 2011 17th Korea-Japan Joint Workshop on*, pages 1–4, feb. 2011.
- [LS09] T. W. R. Lo and J. P. Siebert. Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding*, 113(12):1235–1250,, December 2009.
- [MS05] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.
- [MTUK94] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. pages 282–292, 1994.
- [OBG11] Òscar Boullosa García. Estudio comparativo de descriptores visuales para la detección de escenas cuasi-duplicadas. In *Proyecto Fin De Carrera, Universidad Autónoma De Madrid Escuela Politécnica Superior*, 2011.
- [SL04] Iryna Skrypnyk and David G. Lowe. Scene modelling, recognition and tracking with invariant image features. In *ISMAR* [SL04], pages 110–119.
- [TM07] Tinne Tuytelaars and Krystian Mikolajczyk. Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2007.

- [WRM⁺10] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. Real-time detection and tracking for augmented reality on mobile phones. *IEEE Trans. Vis. Comput. Graph*, 16(3):355–368, 2010.