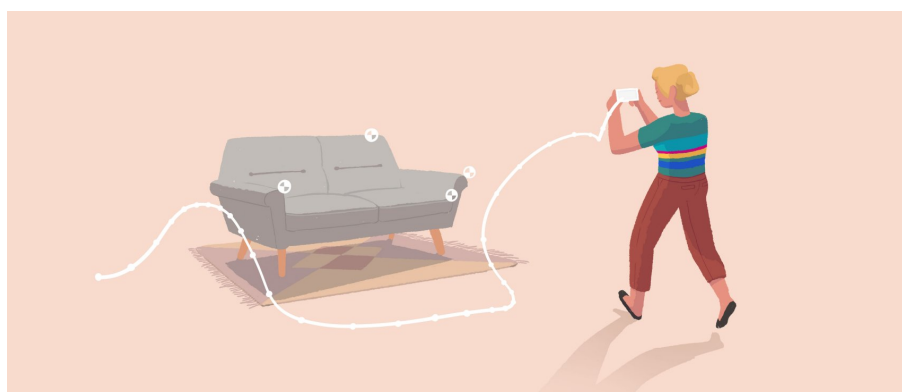


## 0.1 Motion tracking

ARCore usa un processo chiamato *Simultaneous localization and mapping (SLAM)* per determinare lo stato del dispositivo che si trova all'interno di un ambiente sconosciuto. Questo stato è descritto dalla sua posa (posizione e orientazione) che viene stimata attraverso prestazioni di odometria eccezionali e rilevazione di punti caratteristici. Con odometria si intende l'uso di dati ricavati da sensori di movimento che permettono di valutare il cambiamento della posizione nel tempo. Nel caso degli smartphone viene utilizzato il sensore IMU che rileva misure inerziali (dati non visuali) come la velocità, accelerazione e posizione. La rilevazione di punti caratteristici è l'individuazione di immagini con caratteristiche differenti (dati visuali) consentono al dispositivo di calcolare la sua posizione relativa. Questi punti di riferimento insieme alle misurazioni ricavate dai sensori permettono di avere una buona stima della posa e di ricavare la rappresentazione di una mappa dell'ambiente circostante. Tuttavia, il movimento sequenziale stimato dallo SLAM include un certo margine di errore che si accumula nel tempo causando una notevole deviazione dai valori reali. Una soluzione che può essere adottata per risolvere questo problema consiste nel considerare come punto di riferimento un luogo visitato in precedenza di cui si sono memorizzate le sue caratteristiche. Grazie alle informazioni di questo luogo è possibile minimizzare l'errore nella stima della posa.

I contenuti virtuali possono essere renderizzati nella giusta prospettiva allineando la posa della telecamera virtuale con quella calcolata da ARCore. Il contenuto virtuale sembra reale perché è sovrapposto all'immagine ottenuta dalla fotocamera del dispositivo.

Nella figura 1 si può notare come la posizione dell'utente è tracciata in relazione ai punti caratteristici identificati nel divano reale.



Fonte: <https://developers.google.com/ar/develop/fundamentals>

Figura 1: Motion Tracking

Nella figura 2 nella pagina successiva sono rappresentati i moduli principali della SLAM che si differenziano in quattro categorie:

- *Sensori*: nel caso degli smartphone fotocamera, sensore IMU (accelerometro, giroscopio); potrebbero essere inclusi altri sensori per migliorare la precisione come GPS, sensori di luce, profondità.
- *Front-end*: riceve i dati (visuali e non visuali) dai quali ricava una stima della posa. Dai dati *visuali* identifica i punti caratteristici dai quali vengono estratti i descrittori di features. Questi descrittori descrivono orientazione, scala, gravità, direzione e altri aspetti. La stima della posa si basa sulle features del frame precedente. Inoltre, questo modulo mantiene solo una storia limitata dei movimenti tracciati e fissa i punti caratteristici rilevati precedentemente. I dati *non visuali* aiutano a correlare le caratteristiche spaziali osservate in un frame

con quelle osservate in sequenze di frame per determinare la posa efficientemente. Questo modulo include il riconoscimento dei luoghi già visitati se le stesse features vengono associate molte volte ai punti caratteristici.

- *Back-end*: elabora l'output del front-end dal quale crea una rappresentazione 3d dell'ambiente sulla base di una pluralità memorizzata di mappe, descrittori di features, e stima della posa del dispositivo. Successivamente passa questa ricostruzione geometrica alla SLAM estimate.
- *SLAM estimate*: calcola una posa localizzata cercando di ridurre al minimo le discrepanze tra i descrittori di features estratti e memorizzati.

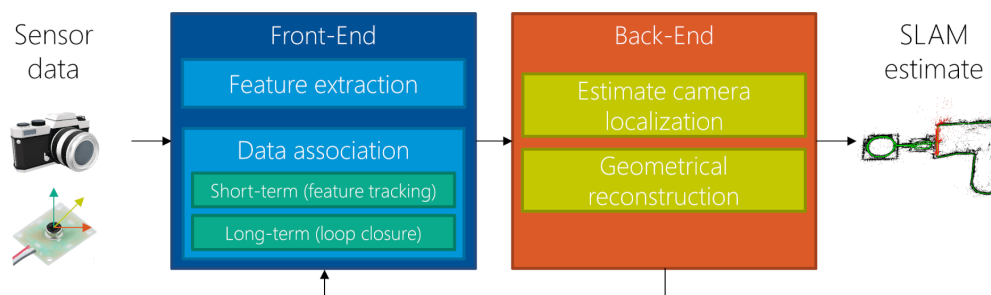
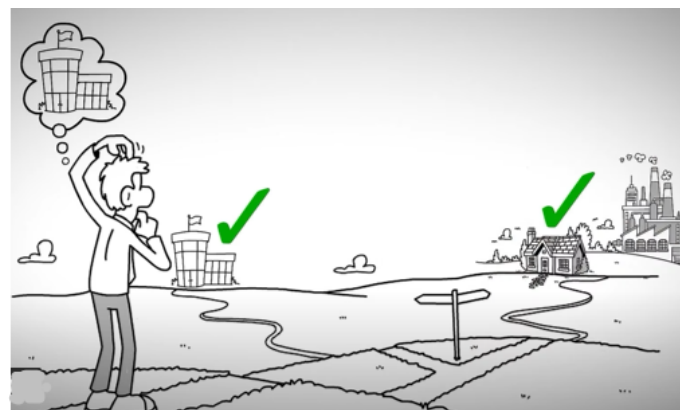


Diagram based on: Cadena, Cesar, et al. "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age." *IEEE Transactions on Robotics* 32.6 (2016): 1309-1332.  
 SLAM estimate: R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, Oct. 2015.  
 3D Model of motion sensor by gerduleib: <https://www.remix3d.com/details/G009SVNP3rSM>  
 3D Model of camera by Microsoft: <https://www.remix3d.com/details/G009SKQ93TH9>

Fonte: <https://www.andreasjakl.com/basics-of-ar-slam-simultaneous-localization-and-mapping/>

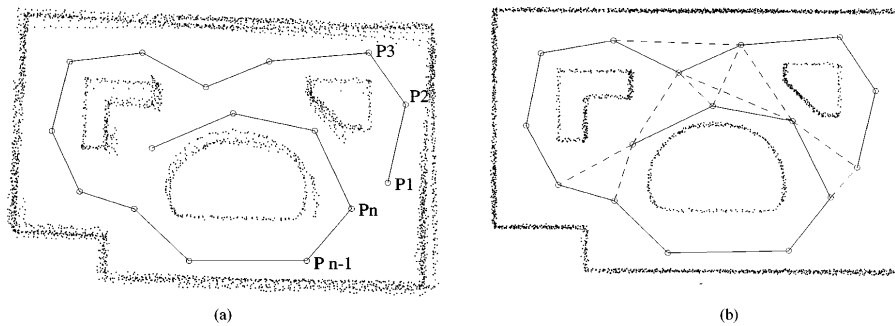
Figura 2: Struttura dei moduli SLAM



Fonte: <https://www.youtube.com/watch?v=UkIcup0TJrY&t=170s>

Figura 3: Punti caratteristici

Nell'esempio (a) in figura 4 a fronte si può vedere come il margine di errore si accumula nel tempo portando ad una rappresentazione inconsistente della mappa dell'ambiente. In (b) si può notare un miglioramento della mappa allineando le varie scansioni in base ai vincoli imposti dalle pose relative.



Fonte: <https://www.andreasjakl.com/basics-of-ar-slam-simultaneous-localization-and-mapping/>

Figura 4: Correzione degli errori da una mappa

Per eseguire dei test sulle **performance** del Motion Tracking si devono tenere in considerazione alcuni aspetti:

- *Angolazione*: l'abilità di rilevare punti caratteristici o superfici può dipendere dall'angolazione del dispositivo. Con alcuni angoli si potrebbero ottenere risultati migliori.
- *Movimento*: questo test potrebbe essere iniziato con un movimento lento e con un aumento progressivo. A seguito di un movimento più rapido il dispositivo dovrebbe essere in grado di comprendere l'ambiente circostante e tracciare eventuali contenuti virtuali. Altri test possono essere effettuati con movimenti rapidi improvvisi e la copertura della fotocamera.