

(a) Photo query



Image I

Observation O

Names N_O ; text T_O

Objects B_O ; floor ℓ_O

Context: position $\mathbf{x}_0$, floor, route

(b) Map index

Maps + landmark records

Record k_i

Names N_i ; descriptions T_i

Node v_i ; floor ℓ_i

Source; association

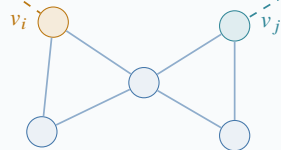
Record k_j

Names N_j ; descriptions T_j

Node v_j ; floor ℓ_j

Source; association

Public-path graph $G = (V, E)$



Solid: mobility edges
Dashed: record associations

(c) Bounded candidate retrieval

Floor consistency + spatial radius r
Character-bigram ranking; keep Top-20

ID	Retrieved record	Anchor
k_i	N_i, T_i, ℓ_i	v_i
k_j	N_j, T_j, ℓ_j	v_j
$\vdots$	$\vdots$	$\vdots$

$C_{20} \subseteq K$ Distance breaks lexical ties.

(e) Evidence and candidate ranking

Observed span $\leftrightarrow$ mapped span

$k_i : o_a[\text{photo}] \leftrightarrow m_a[\text{record}]$

$k_j : o_b[\text{user}] \leftrightarrow m_b[\text{record}]$

Validate declared sources and span length.

$S_k = \text{round}_3(\min\{\gamma, \kappa_k, W_k\})$

$W_k = 0.70L_k + 0.15M_k + 0.10D_k + 0.05R_k$

Rank 1: $k_{(1)}$ $v_{(1)}$ $S_{(1)}$ evidence

Rank 2: $k_{(2)}$ $v_{(2)}$ $S_{(2)}$ evidence

Decision: ready / confirm / no match

Only ready or confirmed anchors proceed.

(d) Evidence-grounded prompt

Structured input

observation: O

candidates: C_{20}

user_description: U (optional)

Use supplied candidate IDs.

Pair observed and mapped text spans.

Preserve each evidence source.

Map-matching model

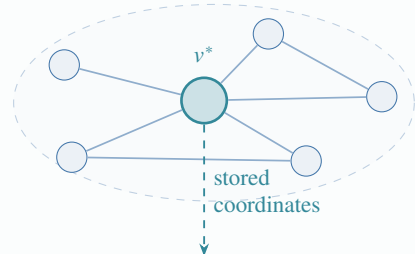
Candidate IDs

+ evidence pairs

matches H

(f) Graph-node coordinate lookup

Resolve the selected record to its stored node.



$v^* \mapsto (\text{lon}(v^*), \text{lat}(v^*), \ell^*)$

Validate ID and journey context

Replan from the new anchor