Object-Centric Capability Evolution Table

This table aligns:

- Core object-centric abilities (Identity / Pose / Trajectory+Location / Interaction)
- Path 1 (3DGS Reconstruction → Real-Scene Automatic Pipeline)
- Path 2 (Tracking → ReID → Full System)
- Evolution from Now → Mid Stages → Final System

Capability Dimension	**Stage Now**	**Path 1 (3DGS Real-Scene Pipeline)**	**Path 2 Phase 2 (Tracking)**	**Path 2 Phase 3 (ReID)**	**Path 2 Phase 4 (Full System)**	**Final Stage**
Identity Recognition	Class-level only (no individual identity)	Provides high-quality consistent labels to train identity models on real scenes	—	Instance-level identity via ReID embeddings	Multi-camera identity linking across time	Long-term stable identity persistence in any camera setup
State / Pose / Activity	No pose, no action recognition	3DGS can generate pose labels once articulated objects are added	Temporal smoothing allows stable pose pipelines	ReID features support pose-sensitive embedding	Pose tracking integrated with full MOT system	Full human/objects pose + action understanding
Trajectory & Location	Frame-by-frame, unstable; approximate location via affine mapping	Real-scene 3D reconstruction → true camera poses → accurate mapping	LSTM/Transformer → stable continuous trajectories	ReID prevents ID-switch → stable long-term trajectory	Full MCMOT: detection + LSTM + association + ReID	Accurate global trajectory in real scenes with zone-level semantics
Interaction (Object–Object / Object–Zone)	None	Provides full 3D context (walls, zones, obstacles)	Basic relations (approaching, near)	Identity-stable interaction chains	Event pipeline (enter-zone, follow, collide)	Full interaction graph + causal reasoning

Interpretation

Path 1 (3DGS):

- Phases 1–3: preparation
- **Phase 4:** the real value
    - Real-scene reconstruction
    - Automatic labeling
    - Customer-specific training pipeline
    - Reduces sim2real gap
    - Enables deploying a full automated dataset builder with only camera input

Path 2:

- Phase 2: tracking stability
- Phase 3: identity-level consistency
- Phase 4: full multi-camera multi-object perception system

Final Capability

A complete object-centric system capable of:

- identifying who is who
- understanding pose/action
- tracking long-term trajectories
- locating objects in a unified real-scene 3D map
- detecting interactions or events
- supporting customer-specific pipelines with no manual annotation