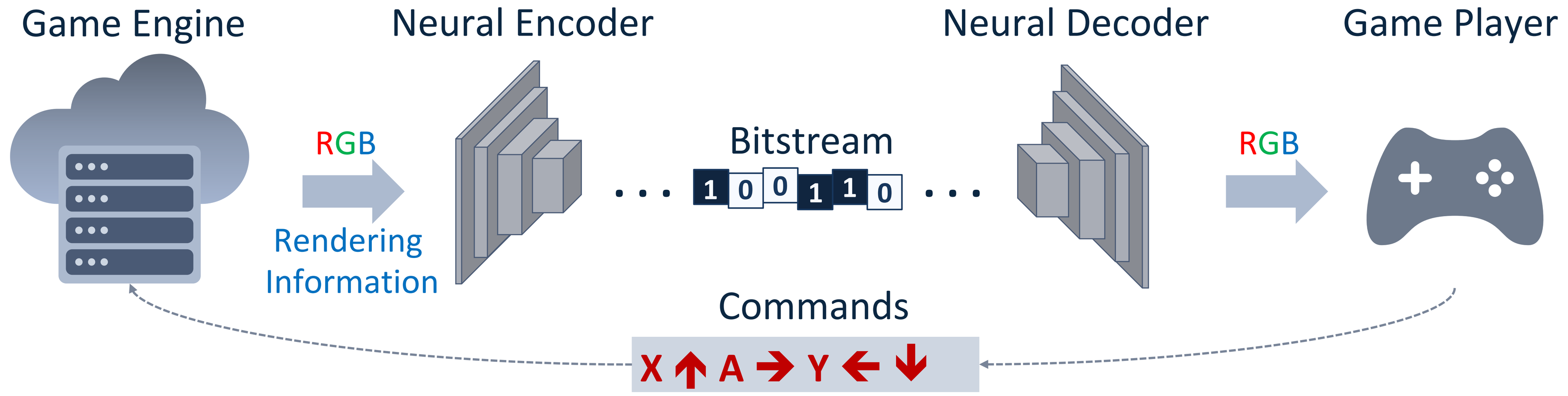
GameCodec: Neural Cloud Gaming Video Codec

Hoang Le, Reza Pourreza, Amir Said, Guillaume Sautiere, Auke Wiggers
Qualcom AI Research
{hoanle,pourreza,gsautie,auke}@qti.qualcomm.com

Cloud Gaming

Motivation

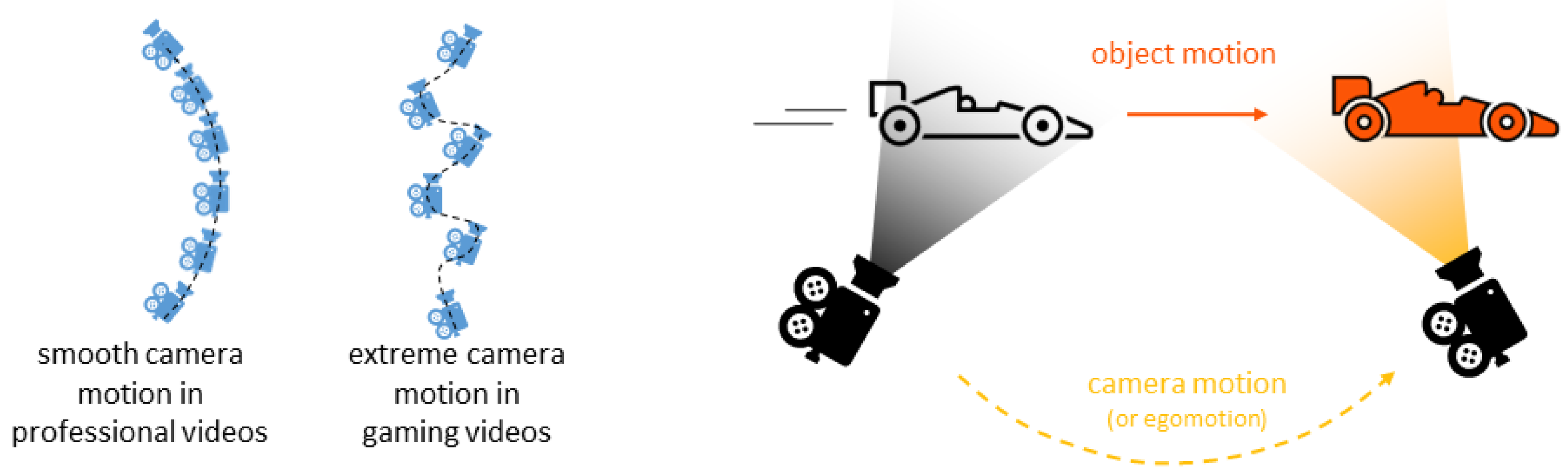
- Modern video games require powerful hardware to render frames in high quality
- Cloud gaming is a type of online gaming that aims to address this issue



Challenges of video codecs in the cloud gaming setting

- Requires high quality frames under extremely low-latency constraints
- Rich textures and visual effect
- Extreme camera and object motions

Camera motion in cloud gaming setting

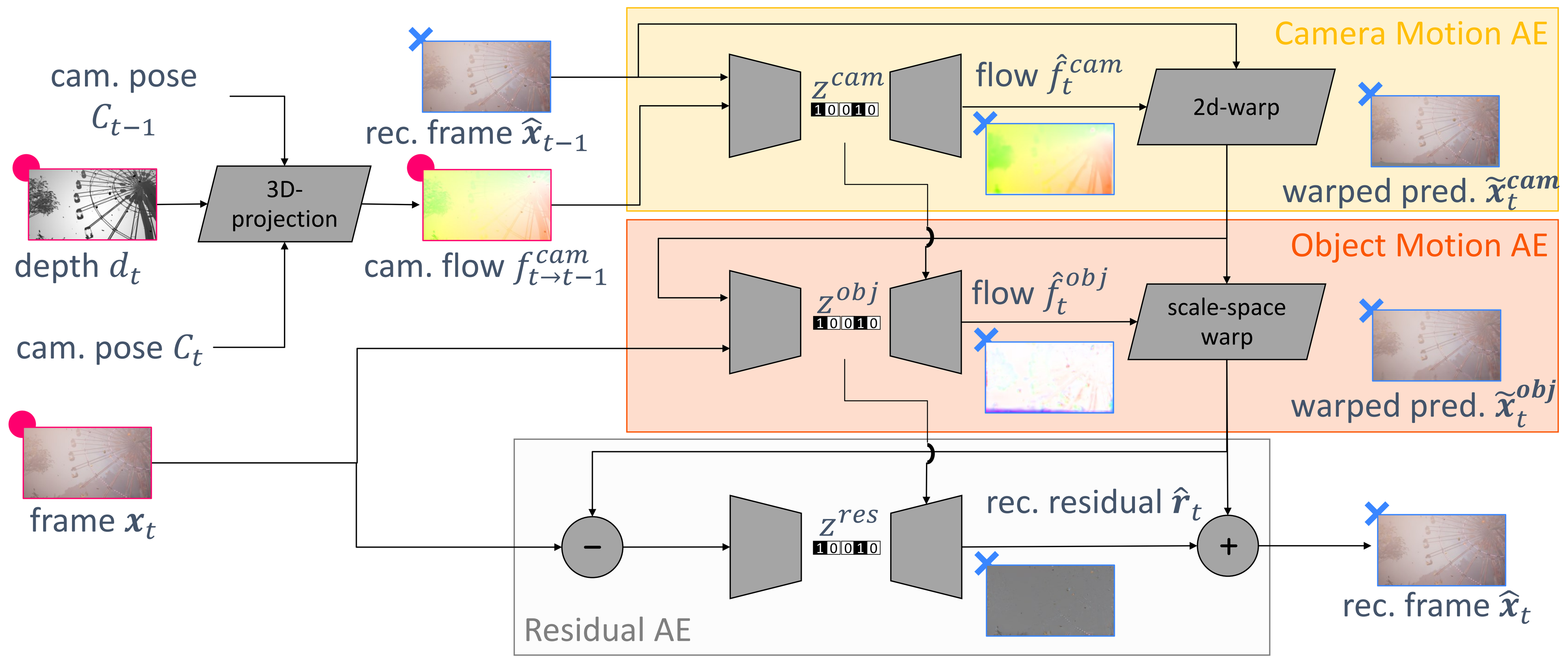


- **In natural videos or animations:**
  - content creator controls the camera movement, not the viewer
  - predictable camera movements to allow the viewer to follow and understand the content
- **In cloud gaming videos:**
  - player controls the camera movement → higher tolerance for abrupt camera motion
  - motion compensation becomes more challenging
  - video codecs likely rely more on residual coding, increasing rate consumption

Our solution: GameCodec

- First end-to-end neural video codec designed for cloud gaming applications
- Proposed motion compensation method addressing abrupt motion typical in gaming
- Proposed complete pipeline from data collection to development and evaluation of neural video codecs on rendered content videos with auxiliary rendering information.

Method: decomposed motion compensation



Camera Motion Compensation

- Leverage camera pose and depth to enable egomotion compensation
  - Use 3D projection to estimate pixel correspondence of the camera planes
  - Use a separated Camera Motion AE to encode and transfer the camera motion

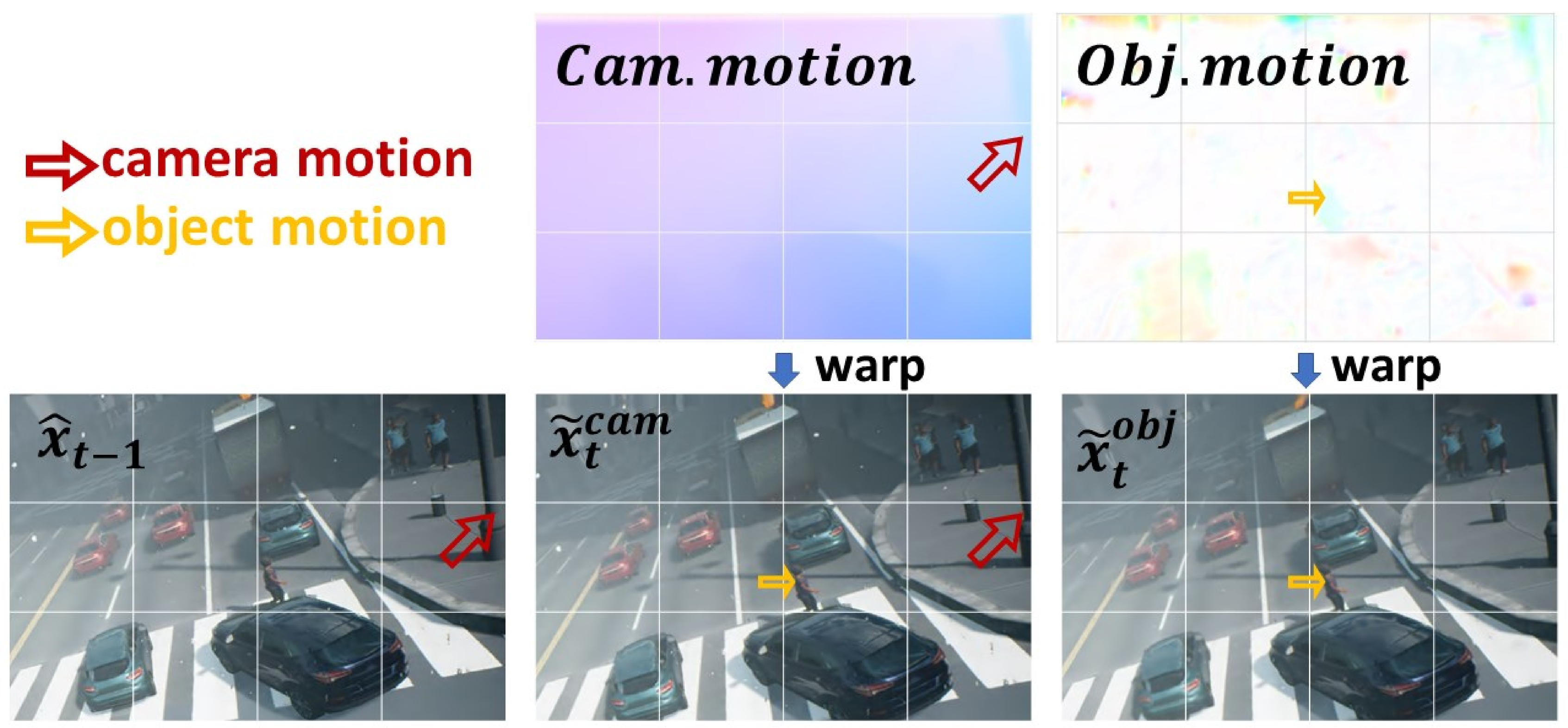
Object Motion Compensation

- Object Motion AE based on the P-frame flow AE in the Scale-Space Flow model
- Mean-Scale Hyperprior AE estimates the object motion scale-space vector field $\hat{f}^{obj}_t$

Rate-Distortion objective

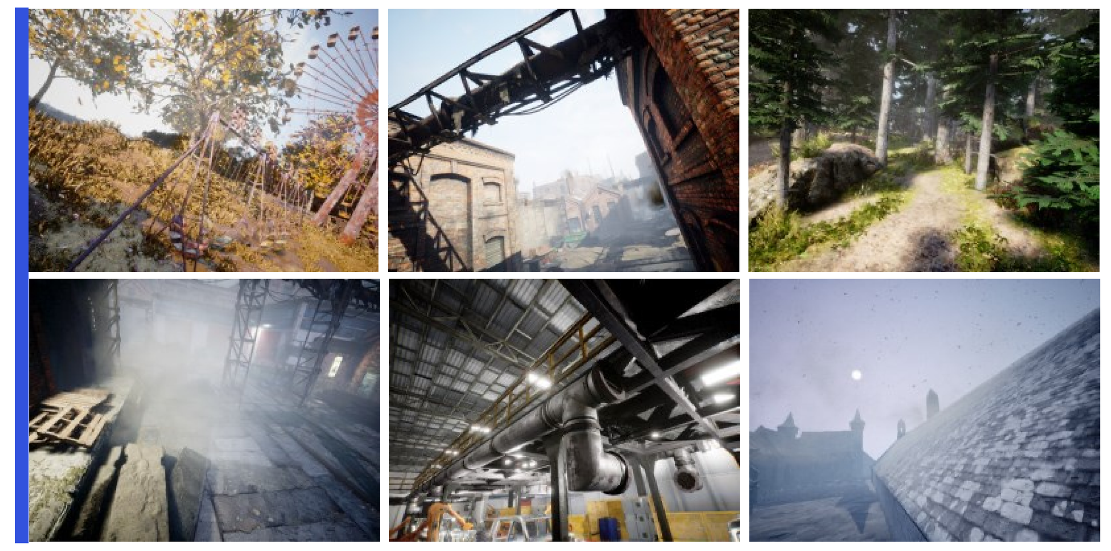
$$\mathcal{L}_{RD}(x) = \beta \cdot \mathcal{L}_R + L_D = \mathbb{E}\left[\beta \cdot ||x - \hat{x}||_2^2 - \log p_\theta(z)\right]$$

Visualization of decomposed motion compensation

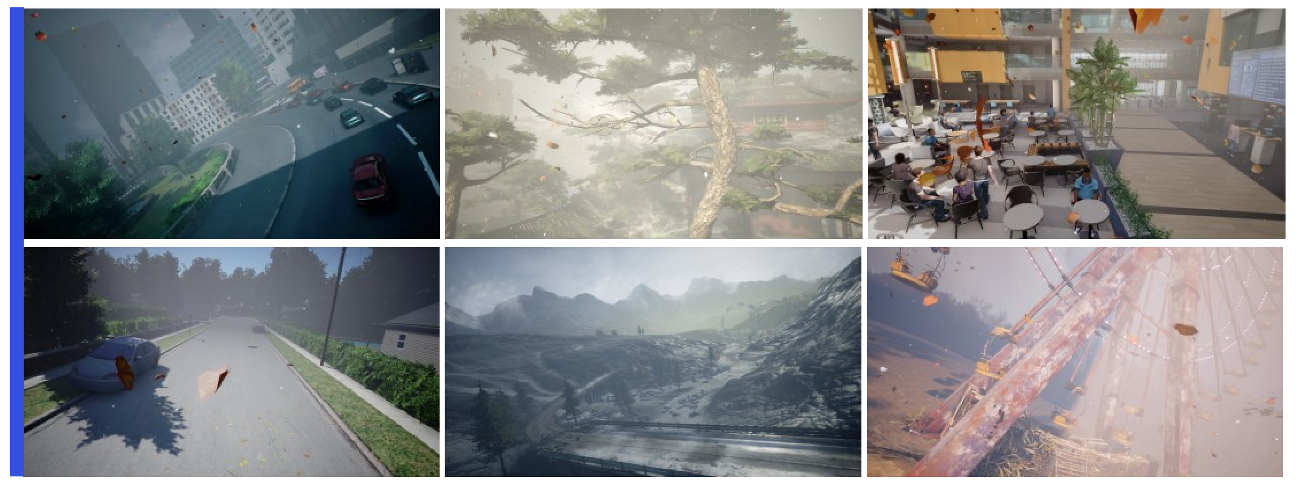


Datasets

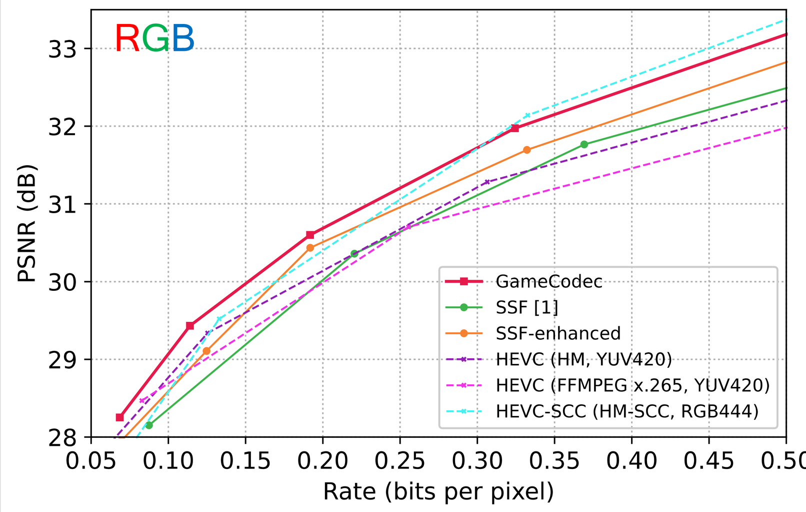
TartanAir



AirSim

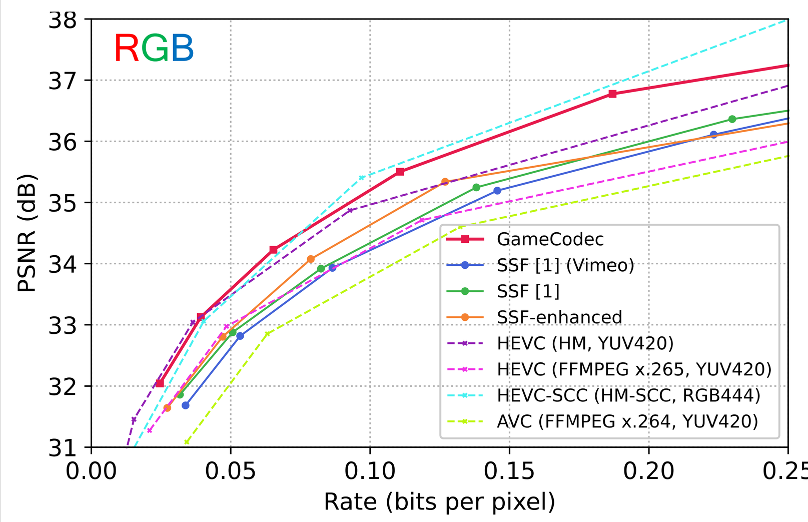


Results



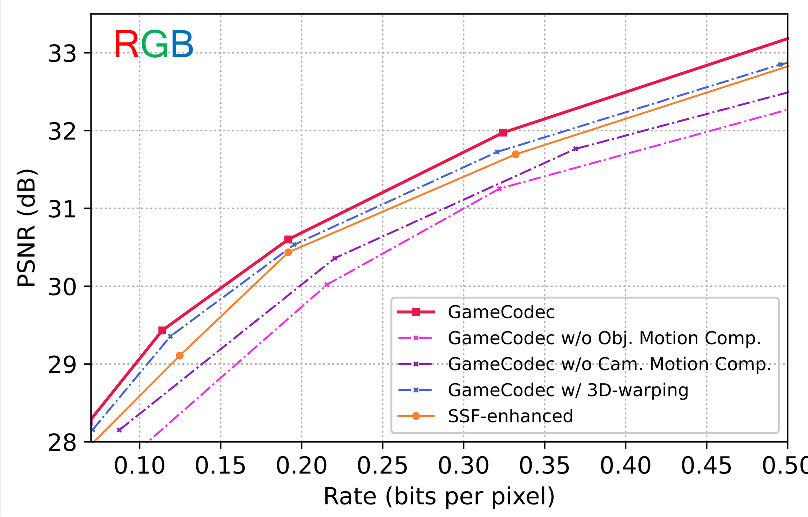
Comparison to literature on TartanAir

- \+ 26.7% BD-rate savings compared to SSF
- \+ outperforms HEVC-SCC in low-rate
- \- underperforms HEVC-SCC in high-rate



Generalization study on AirSim

- \+ maintains edge over SSF and HEVC-SCC
- \- gap closes with HEVC-SCC on low-rate



Ablation study on TartanAir

- \+ shows importance of both object and camera motion AE
- \+ confirms result from SSF where scale-space warping improves performance