

Online Resource 1

Does Enabling the Depth API Improve Anchor Stability? A Controlled Cross-Platform Evaluation of Mobile AR

Journal: *Virtual Reality*

Deokhwan Park, Woojae Kim, Hoyeon Ahn

Corresponding author: Hoyeon Ahn, College of Software Convergence, Soonchunhyang University, Asan
31538, South Korea

E-mail: hoyeon@sch.ac.kr

This resource contains the robustness analyses referred to in Sects. 3.4 and 4.2 of the article (Sect. S1), the per-cell protocol-fidelity checks summarised in Sect. 4.1 (Sect. S2), the per-cell values of the secondary post-block endpoint reported in aggregate in Sect. 4.4 (Sect. S3), the run-level relation between depth availability and anchor error (Sect. S4), and the list of the 13 recorded runs that were set aside, with the criterion applied to each (Sect. S5). Cell names follow the article: E1 is the texture-rich room, E2 the texture-sparse room; *low* and *normal* are the two illuminance levels; ON/OFF is the depth-API setting. Abbreviations follow the article: API, application programming interface; ANOVA, analysis of variance; ToF, time-of-flight; LiDAR, light detection and ranging; IQR, interquartile range. All values were computed from the deposited per-run endpoint table and per-frame logs with the released analysis script.

S1 Factorial ANOVA on $\log_{10} \delta_{\text{term}}$ and stratified permutation test

The pre-declared analysis is rank based (Mann–Whitney U , Cliff’s δ with bootstrap intervals). As a robustness check, a full-factorial ordinary-least-squares model was fitted to $\log_{10} \delta_{\text{term}}$ with sum-to-zero contrasts and Type-III sums of squares, separately for each device (texture $\times$ illuminance $\times$ depth, $n = 200$ each, Tables S1–S3) and for all devices together (device $\times$ texture $\times$ illuminance $\times$ depth, $n = 600$, Table S4). The log transform removes most of the skew (Shapiro–Wilk rejects normality in 9 of the 24 cells, against 20 of 24 on the raw scale) but the residual variance still differs across cells (Levene’s test, $p < 10^{-16}$), so these tables are a robustness check and not the primary analysis. Because of the heteroscedasticity, the depth main effect within each device was additionally tested with a stratified permutation test: the ON/OFF labels were shuffled within each texture $\times$ illuminance stratum (5 000 permutations) and the difference in mean $\log_{10} \delta_{\text{term}}$ between the ON and OFF arms was recomputed each time (Table S5). A positive difference means that the depth-ON arm had the larger error.

Table S1: Galaxy S10 5G (ToF): texture $\times$ illuminance $\times$ depth ANOVA on $\log_{10} \delta_{\text{term}}$, $n = 200$

Effect	df	F	p	partial η^2
Texture	1	33.76	<0.001	0.150
Illuminance	1	20.38	<0.001	0.096
Depth	1	1.38	0.241	0.007
Texture $\times$ Illuminance	1	0.35	0.555	0.002
Texture $\times$ Depth	1	2.39	0.124	0.012
Illuminance $\times$ Depth	1	0.01	0.933	0.000
Texture $\times$ Illuminance $\times$ Depth	1	2.06	0.153	0.011

On every device the ON arm had the larger mean $\log_{10} \delta_{\text{term}}$, matching the direction of the rank-based within-cell results (article, Table 3 and Fig. 6). The permutation p -values agree with the ANOVA p -values to within the precision expected from 5 000 permutations, and neither approach finds a depth $\times$ texture or depth $\times$ illuminance interaction.

S2 Protocol fidelity per cell

Table S6 lists, for each of the 24 primary cells, the cell median of three per-run quantities taken from the per-frame logs: the platform-estimated camera-to-marker distance during P0 (median over the frames in which the marker was tracked before the cover was applied), the absolute yaw rate during the four pans (P1–P4,

Table S2: Galaxy S24+ (passive): texture $\times$ illuminance $\times$ depth ANOVA on $\log_{10} \delta_{\text{term}}$, $n = 200$

Effect	df	F	p	partial η^2
Texture	1	42.21	<0.001	0.180
Illuminance	1	4.86	0.029	0.025
Depth	1	3.58	0.060	0.018
Texture $\times$ Illuminance	1	0.04	0.846	0.000
Texture $\times$ Depth	1	0.10	0.748	0.001
Illuminance $\times$ Depth	1	2.65	0.105	0.014
Texture $\times$ Illuminance $\times$ Depth	1	0.56	0.456	0.003

Table S3: iPhone 14 Pro Max (LiDAR): texture $\times$ illuminance $\times$ depth ANOVA on $\log_{10} \delta_{\text{term}}$, $n = 200$

Effect	df	F	p	partial η^2
Texture	1	48.08	<0.001	0.200
Illuminance	1	3.35	0.069	0.017
Depth	1	5.79	0.017	0.029
Texture $\times$ Illuminance	1	3.33	0.070	0.017
Texture $\times$ Depth	1	2.71	0.102	0.014
Illuminance $\times$ Depth	1	0.82	0.365	0.004
Texture $\times$ Illuminance $\times$ Depth	1	0.89	0.347	0.005

Table S4: All devices: device $\times$ texture $\times$ illuminance $\times$ depth ANOVA on $\log_{10} \delta_{\text{term}}$, $n = 600$

Effect	df	F	p	partial η^2
Device	2	8.33	<0.001	0.028
Texture	1	107.11	<0.001	0.157
Illuminance	1	25.51	<0.001	0.042
Depth	1	7.74	0.006	0.013
Device $\times$ Texture	2	4.90	0.008	0.017
Device $\times$ Illuminance	2	4.64	0.010	0.016
Texture $\times$ Illuminance	1	0.07	0.787	0.000
Device $\times$ Depth	2	0.32	0.727	0.001
Texture $\times$ Depth	1	0.59	0.441	0.001
Illuminance $\times$ Depth	1	0.83	0.362	0.001
Device $\times$ Texture $\times$ Illuminance	2	0.64	0.525	0.002
Device $\times$ Texture $\times$ Depth	2	1.70	0.183	0.006
Device $\times$ Illuminance $\times$ Depth	2	1.57	0.208	0.005
Texture $\times$ Illuminance $\times$ Depth	1	1.43	0.233	0.002
Device $\times$ Texture $\times$ Illuminance $\times$ Depth	2	1.16	0.313	0.004

Table S5: Depth main effect within each device: stratified permutation test on the difference in mean $\log_{10} \delta_{\text{term}}$ (ON minus OFF, in dex; labels shuffled within texture $\times$ illuminance, 5 000 permutations), with the corresponding ANOVA F and p from Tables S1–S3 for comparison

Device	n	Δ mean $\log_{10} \delta_{\text{term}}$ (dex)	p_{perm}	$F_{1,192}$	p_{ANOVA}
Galaxy S10 5G (ToF)	200	+0.102	0.241	1.38	0.241
Galaxy S24+ (passive)	200	+0.169	0.061	3.58	0.060
iPhone 14 Pro Max (LiDAR)	200	+0.091	0.014	5.79	0.017

median over 5–60 s), and the run duration. The distance column confirms that the gimbal was placed 1.0 m from the marker wall in every cell; the yaw-rate column confirms that the gimbal programme, and not the operator, set the motion.

Table S6: Protocol-fidelity checks per primary cell (cell medians of per-run values; the distance column also gives the range of per-run medians across the 25 runs)

Device	Texture	Light	Depth	n	Camera–marker distance in P0 (m)		Yaw rate ($^{\circ}$ /s)	Duration (s)
					median	range	median	median
Galaxy S10 5G	E1	low	OFF	25	1.04	1.03–1.05	22.9	80.0
Galaxy S10 5G	E1	low	ON	25	1.04	1.03–1.04	22.9	80.0
Galaxy S10 5G	E1	normal	OFF	25	1.03	1.03–1.05	23.0	80.0
Galaxy S10 5G	E1	normal	ON	25	1.03	1.02–1.05	23.1	80.0
Galaxy S10 5G	E2	low	OFF	25	1.04	1.03–1.05	22.5	80.0
Galaxy S10 5G	E2	low	ON	25	1.04	1.03–1.05	22.4	80.0
Galaxy S10 5G	E2	normal	OFF	25	1.04	0.97–1.05	22.8	80.0
Galaxy S10 5G	E2	normal	ON	25	1.04	1.03–1.07	22.9	80.0
Galaxy S24+	E1	low	OFF	25	1.04	1.03–1.06	22.8	80.0
Galaxy S24+	E1	low	ON	25	1.05	1.03–1.05	22.8	80.0
Galaxy S24+	E1	normal	OFF	25	1.04	1.02–1.04	22.9	80.0
Galaxy S24+	E1	normal	ON	25	1.04	1.03–1.06	22.9	80.0
Galaxy S24+	E2	low	OFF	25	1.05	1.02–1.06	22.6	80.0
Galaxy S24+	E2	low	ON	25	1.05	1.04–1.08	22.6	80.0
Galaxy S24+	E2	normal	OFF	25	1.04	1.01–1.05	23.0	80.0
Galaxy S24+	E2	normal	ON	25	1.03	1.02–1.06	23.0	80.0
iPhone 14 Pro Max	E1	low	OFF	25	1.03	1.03–1.04	23.1	80.0
iPhone 14 Pro Max	E1	low	ON	25	1.03	1.03–1.04	23.1	80.0
iPhone 14 Pro Max	E1	normal	OFF	25	1.03	1.03–1.04	23.1	80.0
iPhone 14 Pro Max	E1	normal	ON	25	1.03	1.03–1.04	23.1	79.9
iPhone 14 Pro Max	E2	low	OFF	25	1.04	1.02–1.04	23.2	80.0
iPhone 14 Pro Max	E2	low	ON	25	1.04	1.04–1.04	23.1	80.0
iPhone 14 Pro Max	E2	normal	OFF	25	1.04	1.04–1.04	23.2	80.0
iPhone 14 Pro Max	E2	normal	ON	25	1.04	1.03–1.04	23.2	80.0

S3 Post-block endpoint per cell

Table S7 gives the per-cell distribution of the secondary endpoint δ_{block} , the largest anchor error over the valid marker frames within 2 s of tracking resuming after the 5 s camera block (P6), the number of runs in which it reached 0.5 m, and the recovery latency from unblocking to the first such frame. The article reports the per-device medians (0.28, 0.23 and 0.06 m) and the two E2/low cells; this table supplies the remaining cells. δ_{block} is undefined in three iPhone 14 Pro Max E2/normal runs in which no valid marker frame was logged during the block window, so the denominators in the two E2/normal rows are 24 and 23 rather than 25. Recovery latency is likewise undefined in a run in which no valid marker frame was logged within the 2 s window; the last column gives the number of runs with a defined latency. Latency is quantised by the logging rate, which is a nominal 10 Hz and 9.5 Hz as achieved on the iPhone 14 Pro Max (article, Sect. 4.1.3). Because the uncovering instant is estimated from the log rather than instrumented, a run in which the marker was re-acquired before that estimate yields a small negative latency; 33 runs, all on the iPhone 14 Pro Max, are affected, and they do not change any median reported here.

S4 Depth availability and anchor error at the run level

The article reports that availability at neither stage separated stable anchors from drifting ones (Sect. 4.3). This section gives the run-level evidence for that statement, for the eight ARCore depth-ON cells in which delivery varies (on the iPhone 14 Pro Max every frame carries a complete depth image, so the correlation is undefined).

Within a cell, neither the proportion of occluded-pan frames carrying a depth image nor the proportion of valid pixels within those images is related to δ_{term} : the eight within-cell Spearman correlations for delivery span -0.28 to $+0.19$ and those for fill -0.27 to $+0.19$, and none reaches the 5% level (Table S8). Fitting $\log_{10} \delta_{\text{term}}$ on delivery with the cell as a fixed effect gives $\beta = -0.0016$ per percentage point of delivery ($p = 0.43$); the same fit on fill gives $\beta = +0.00094$ per percentage point of fill ($p = 0.58$).

Table S7: Post-block anchor error δ_{block} and recovery latency per primary cell (median and interquartile range over the runs of the cell)

Device	Texture	Light	Depth	δ_{block} (m)		$\delta_{\text{block}} \geq 0.5$ m	Recovery latency (s)	
				median	[IQR]		runs	median
Galaxy S10 5G	E1	low	OFF	0.385	[0.110, 0.607]	7/25	0.40	24
Galaxy S10 5G	E1	low	ON	0.320	[0.120, 0.449]	5/25	0.40	25
Galaxy S10 5G	E1	normal	OFF	0.101	[0.055, 0.196]	1/25	0.40	23
Galaxy S10 5G	E1	normal	ON	0.178	[0.088, 0.229]	0/25	0.40	24
Galaxy S10 5G	E2	low	OFF	1.283	[1.100, 1.842]	25/25	0.40	22
Galaxy S10 5G	E2	low	ON	1.468	[1.062, 1.873]	23/25	0.40	25
Galaxy S10 5G	E2	normal	OFF	0.237	[0.123, 0.358]	1/25	0.40	21
Galaxy S10 5G	E2	normal	ON	0.187	[0.131, 0.298]	1/25	0.40	23
Galaxy S24+	E1	low	OFF	0.198	[0.078, 0.386]	3/25	0.40	24
Galaxy S24+	E1	low	ON	0.292	[0.203, 0.456]	6/25	0.40	25
Galaxy S24+	E1	normal	OFF	0.078	[0.055, 0.197]	1/25	0.40	25
Galaxy S24+	E1	normal	ON	0.131	[0.090, 0.147]	0/25	0.40	25
Galaxy S24+	E2	low	OFF	1.313	[1.101, 1.892]	21/25	0.40	25
Galaxy S24+	E2	low	ON	1.121	[0.871, 1.475]	24/25	0.40	24
Galaxy S24+	E2	normal	OFF	0.183	[0.131, 0.266]	2/25	0.40	25
Galaxy S24+	E2	normal	ON	0.259	[0.147, 0.369]	5/25	0.40	25
iPhone 14 Pro Max	E1	low	OFF	0.057	[0.036, 0.070]	0/25	0.05	24
iPhone 14 Pro Max	E1	low	ON	0.090	[0.062, 0.103]	0/25	0.00	21
iPhone 14 Pro Max	E1	normal	OFF	0.035	[0.019, 0.054]	0/25	0.10	25
iPhone 14 Pro Max	E1	normal	ON	0.046	[0.033, 0.085]	0/25	0.12	25
iPhone 14 Pro Max	E2	low	OFF	0.065	[0.044, 0.240]	1/25	0.10	22
iPhone 14 Pro Max	E2	low	ON	0.082	[0.058, 0.130]	0/25	0.12	22
iPhone 14 Pro Max	E2	normal	OFF	0.073	[0.051, 0.126]	0/24	0.10	24
iPhone 14 Pro Max	E2	normal	ON	0.112	[0.072, 0.152]	0/23	0.10	22

Pooling the eight cells instead gives $\rho = -0.33$ ($p = 1.5 \times 10^{-6}$), which would suggest that runs receiving more depth drift less. That association is an artefact of aggregating over the texture contrast rather than an effect of depth: the textured room has both the higher delivery (median 46% against 5%) and the lower error (median 50 mm against 270 mm), so texture induces the correlation across cells while none exists within them. We therefore report the within-cell analysis, in which the cell fixes texture, illuminance and device.

Table S8: Run-level Spearman correlation between depth availability and δ_{term} within each ARCore depth-ON cell. n_f is the number of runs with a defined fill value, which is smaller than n in cells where some runs received no depth image at all

Device	Texture	Illum.	Delivery in W vs δ_{term}		Fill vs δ_{term}		n_f	ρ	p
			n	ρ	p				
Galaxy S10 5G	E1	low	25	-0.01	0.95	25	+0.08	0.70	
Galaxy S10 5G	E1	normal	25	+0.02	0.92	24	+0.15	0.48	
Galaxy S10 5G	E2	low	25	+0.01	0.95	25	+0.19	0.37	
Galaxy S10 5G	E2	normal	25	+0.07	0.73	21	-0.27	0.23	
Galaxy S24+	E1	low	25	+0.19	0.36	24	+0.01	0.97	
Galaxy S24+	E1	normal	25	-0.27	0.20	22	+0.15	0.49	
Galaxy S24+	E2	low	25	-0.16	0.45	24	-0.03	0.89	
Galaxy S24+	E2	normal	25	-0.28	0.17	23	+0.14	0.53	

S5 Runs set aside

Of the 708 recorded runs, 600 form the 24 primary cells, 95 form the eight supplementary temporal-smoothing cells and 13 were set aside before any endpoint was computed. Table S9 lists the 13 runs with the criterion applied and the logged quantity on which the decision was based. All criteria are procedural (protocol completion, logging integrity, marker exposure) and none refers to δ_{term} . Marker exposure is the fraction of logged frames in which the marker was tracked; the cell range quoted for comparison is over the retained runs of the same cell. The per-frame logs of all 13 runs are included in the data deposit in a separate folder.

Table S9: The 13 recorded runs that were set aside, with the criterion applied to each (the log ID is the date and local time of recording, as in the deposited file names)

Device	Texture	Light	Depth	Log ID	Criterion	Logged basis
Galaxy S10 5G	E1	normal	OFF	20260822.161508	Protocol not completed	Duration 67.2 s (below 79 s)
Galaxy S10 5G	E1	normal	OFF	20260822.163712	Logging fault	FeatureCount constant at 131 over 796 frames
Galaxy S10 5G	E1	normal	ON	20260822.163953	Logging fault	FeatureCount constant at 131
Galaxy S10 5G	E1	normal	ON	20260822.164314	Logging fault	FeatureCount constant at 131
Galaxy S10 5G	E1	normal	ON	20260822.171457	Marker exposure	Marker visible in 20% of frames against 29–33% in the cell
Galaxy S10 5G	E2	low	ON	20260810.202048	Protocol not completed	FirstReacq = 0; marker visible in 7% of frames, 0% of the final 5 s
Galaxy S10 5G	E2	low	ON (smoothed)	20260811.115128	Protocol not completed	FirstReacq = 0; duration 74.8 s; marker visible in 9% of frames
Galaxy S10 5G	E2	normal	ON (smoothed)	20260809.175959	Pre-protocol	Smoothing protocol not yet finalised; marker visible in 6% of the final 5 s
Galaxy S10 5G	E2	normal	ON (smoothed)	20260809.190255	Pre-protocol	Smoothing protocol not yet finalised
Galaxy S10 5G	E2	normal	ON (smoothed)	20260809.190420	Pre-protocol	Smoothing protocol not yet finalised
Galaxy S10 5G	E2	normal	ON (smoothed)	20260809.190546	Pre-protocol	Smoothing protocol not yet finalised
iPhone 14 Pro Max	E2	normal	ON	20260814.143152	Final window	Marker visible in 4% of the final 5 s (75–100% in the cell)
iPhone 14 Pro Max	E2	normal	OFF	20260814.151040	Final window	Marker visible in 6% of the final 5 s (80–100% in the cell)