This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material

Intraoperative AR Visualization of Preoperative (CB)CT via X-Ray Registration

1 Supplementary Scope

This supplement reports analyses referenced but not expanded in the main manuscript. It contains weight sensitivity experiments, subject-level results, paired analysis, and the use of multiple view combinations for result checking. It also expands on the transition from phantom to cadaveric validation.

2 Additional Multi-View Analyses

Equation (3) of the main manuscript leaves the view weights as design parameters. In an ordered group $A/B(/C)$, A is the anchor and the listed weights follow that order. Rows corresponding to the main configurations are marked in the tables.

The tables report anchor-view 3D mTRE for the same view groups used in the main experiments. The 10 mm threshold provides a consistent summary of the large-error tail and is not a clinical safety threshold.

2.1 Femur

Table 1. Femur anchor-view 3D mTRE (mm) across Single, Dual, and Triple strategies. Lower is better.

Method	Weight	Trials	Mean ± SD	Median	P95	Maximum	> 10 mm
XVR (single-view)	1	49	22.200 ± 77.237	9.588	24.246	550.276	24/49
Fixed Dual (main)	0.50/0.50	49	8.486 ± 4.730	7.629	17.641	19.991	13/49
Fixed Dual	0.70/0.30	49	8.865 ± 5.065	7.368	18.784	21.756	15/49
Fixed Dual	0.90/0.10	49	26.432 ± 82.428	9.992	20.859	506.613	24/49
Fixed Dual	0.95/0.05	49	26.297 ± 80.060	9.420	23.891	475.431	23/49
Fixed Dual	0.98/0.02	49	27.791 ± 118.381	9.589	20.557	838.684	24/49
Fixed Triple	1/3, 1/3, 1/3	49	8.592 ± 4.738	7.215	17.166	22.490	16/49
Fixed Triple	0.40/0.30/0.30	49	8.579 ± 4.657	7.766	16.227	21.074	16/49
Fixed Triple	0.45/0.45/0.10	49	8.485 ± 5.278	7.138	18.922	19.718	15/49
Fixed Triple (main)	0.50/0.25/0.25	49	8.934 ± 4.967	7.820	17.838	20.451	18/49
Fixed Triple	0.60/0.20/0.20	49	8.634 ± 4.989	7.693	18.602	22.513	15/49
Fixed Triple	0.75/0.125/0.125	49	15.557 ± 46.844	7.993	20.635	334.710	20/49
Fixed Triple	0.92/0.04/0.04	49	26.684 ± 120.497	7.454	19.670	852.105	20/49

For Femur, the main fixed Dual setting provided the most consistent Dual performance. Balanced Triple settings produced similar results. The anchor-dominant settings identify useful limits for future weight design.

2.2 DeepFluoro

Table 2. DeepFluoro anchor-view 3D mTRE (mm) across Single, Dual, and Triple strategies. Lower is better.

Method	Weight	Trials	Mean ± SD	Median	P95	Maximum	> 10 mm
XVR (single-view)	1	366	4.889 ± 18.321	1.525	14.578	254.157	25/366
Fixed Dual (main)	0.50/0.50	366	4.296 ± 17.885	1.593	5.671	184.552	8/366
Fixed Dual	0.70/0.30	366	4.749 ± 22.594	1.583	6.428	272.906	12/366
Fixed Dual	0.90/0.10	366	5.590 ± 21.053	1.748	15.208	253.921	22/366
Fixed Dual	0.95/0.05	366	5.451 ± 20.537	1.746	13.983	255.065	24/366
Fixed Dual	0.98/0.02	366	5.126 ± 18.218	1.869	13.720	254.024	23/366
Fixed Triple	1/3, 1/3, 1/3	366	3.909 ± 20.105	1.367	3.790	260.562	6/366
Fixed Triple (main)	0.50/0.25/0.25	366	3.188 ± 14.619	1.407	4.031	176.466	7/366
Fixed Triple	0.60/0.20/0.20	366	4.160 ± 20.201	1.492	4.527	255.537	8/366
Fixed Triple	0.75/0.125/0.125	366	3.952 ± 17.616	1.496	6.023	255.435	12/366
Fixed Triple	0.92/0.04/0.04	366	4.880 ± 18.365	1.713	12.926	254.780	23/366

On DeepFluoro, balanced settings generally reduced the typical error and the large-error tail. More anchor-dominant weights approached single-view performance.

3 Subject-Stratified and Paired Evidence

The main manuscript pools 49 Femur and 366 DeepFluoro trials from five and six subjects, respectively. Tables 3 and 4 show the corresponding per-subject results. *Fixed* denotes the main-text Dual and Triple weights.

Table 3. Femur per-subject anchor-view 3D mTRE (mm), mean ± SD.

Femur subject	Trials	XVR	Fixed Dual	Fixed Triple
F01	10	8.119 ± 4.482	6.595 ± 2.377	7.260 ± 2.531
F02	11	7.892 ± 4.128	6.927 ± 2.808	5.884 ± 2.026
F03	10	11.743 ± 6.250	7.973 ± 4.044	9.440 ± 4.884
F04	9	18.115 ± 4.368	15.716 ± 3.627	16.491 ± 2.735
F05	9	71.039 ± 179.772	5.830 ± 3.404	6.402 ± 3.646

Table 4. DeepFluoro per-subject anchor-view 3D mTRE (mm), mean ± SD.

DeepFluoro subject	Trials	XVR	Fixed Dual	Fixed Triple
D01	111	6.464 ± 27.401	3.850 ± 14.304	1.685 ± 1.271
D02	104	3.462 ± 15.026	7.150 ± 29.218	5.977 ± 26.391
D03	24	2.424 ± 3.248	1.955 ± 1.240	1.781 ± 0.940
D04	48	8.275 ± 16.232	4.738 ± 9.926	4.219 ± 9.913
D05	55	1.793 ± 1.850	1.481 ± 0.813	1.369 ± 0.853
D06	24	6.575 ± 9.516	1.905 ± 1.341	1.570 ± 0.711

Femur F04 remained the most challenging subject, while fixed Dual and fixed Triple substantially reduced the large F05 errors. These results show how anatomical heterogeneity and sparse large errors shape the pooled statistics. Future confidence estimation will be designed to remain robust across this variation.

3.1 Paired inference with subject-level clustering

Methods were compared on identical view groups. To account for repeated trials within each subject, paired differences were first averaged by subject. The 95% interval resamples subjects, and p is an exact two-sided sign-flip test over subject means. W/T/L and rescued/new are trial-level counts. Negative differences favor the target.

Table 5. Paired anchor-view 3D mTRE analysis. Intervals and exact p-values use subjects as independent units and are reported without multiplicity adjustment.

Dataset	Target − reference	Trial mean Δ	Subject mean Δ [95% CI]	Exact p	W/T/L	Rescued/new
Femur	Fixed Dual − XVR	-13.715	-14.773 $[-40.085, -1.475]$	0.0625	33/0/16	11/0
Femur	Fixed Triple − XVR	-13.266	-14.286 $[-39.568, -1.319]$	0.0625	34/0/15	8/2
Femur	Fixed Triple − Fixed Dual	$+0.448$	$+0.487$ $[-0.356, +1.128]$	0.3125	20/0/29	1/6
DeepFluoro	Fixed Dual − XVR	-0.592	-1.319 $[-3.377, +0.969]$	0.3125	181/0/185	22/5
DeepFluoro	Fixed Triple − XVR	-1.701	-2.065 $[-4.082, +0.247]$	0.1563	217/0/149	20/2
DeepFluoro	Fixed Triple − Fixed Dual	-1.108	-0.746 $[-1.383, -0.248]$	0.0313	215/0/151	4/3

Fixed Dual moved 11 Femur and 22 DeepFluoro XVR trials from above 10 mm to at or below 10 mm, compared with zero and five transitions in the opposite direction. This demonstrates a meaningful reduction of the large-error tail. With five Femur and six DeepFluoro subjects, these results provide an initial subject-level estimate. Future studies will use larger cohorts, held-out evaluation, and multiplicity-adjusted inference.

4 Using Multiple View Combinations

The results above show that simply adding a view does not always improve the registration. One practical alternative is to run several combinations of the

available X-rays and compare the resulting bone poses. Three X-rays provide four combinations containing at least two views (12, 13, 23, and 123). Four X-rays provide eleven such combinations. These counts are made before choosing the anchor view.

When several combinations return similar poses, their agreement supports use of the result for AR display. When one result differs substantially from the others, it can be rejected or reviewed instead of being displayed directly. The current study has not yet implemented this automatic comparison. Future work will define how closely the candidate poses must agree before a result is accepted. Because every additional X-ray increases radiation dose and acquisition time, extra views should be acquired only when the improvement in reliability justifies that cost.

5 From Phantom to Cadaveric Validation

The phantom experiment intentionally excluded soft tissue so that the operator could directly observe the bone and its corresponding fiducials. This created a clear and controlled setting for evaluating the complete workflow and the perceived alignment of the AR overlay.

Cadaveric validation will introduce additional complexity because soft tissue obscures the bone surface. Even when the workflow correctly recovers the bone pose, the overlay cannot be verified directly while the bone remains covered. Verification would require a surgeon to complete the workflow and then expose the bone for direct comparison. The key challenge is to ensure that the bone does not move during exposure, since any displacement would confound the evaluation. Developing a surgical protocol that preserves the registered pose throughout this process is therefore an important part of future work. Evaluation in collaboration with surgeons will extend the current feasibility study to a more realistic and comprehensive assessment.