



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.

Keypoint Correspondence Learning for CBCT–IOS Registration

Jeonglok Park^{1†}, SeungKwan Jo^{1†}, Jonghoon An¹, and Minyoung Chung^{2*}

¹ Soongsil University, Seoul, Republic of Korea

² Osstem Implant, Seoul, Republic of Korea
chungmy.contact@gmail.com

Abstract. Accurate CBCT–IOS registration is essential for dental diagnosis and implant planning, yet the modality gap between CBCT volumetric intensity and IOS surface geometry substantially increases correspondence uncertainty. Conventional segmentation-based pipelines reduce the modality gap by converting CBCT into a surface-like representation, but their accuracy is highly sensitive to unstable segmentation and does not explicitly enforce reliable correspondences. We reformulate CBCT–IOS registration as *direct keypoint correspondence estimation in feature space*, and propose a correspondence-centric end-to-end framework. Our method learns modality-aligned embeddings, selects registration-relevant keypoints by favoring points with *peaked* cross-modal matching distributions (i.e., unambiguous matches) while regularizing intra-modality distinctiveness, and predicts correspondences jointly with their validity. To further suppress residual near-tie mismatches in repetitive dental geometry, we refine the rigid transform by iterative SVD while annealing the matching criterion from feature similarity to geometric consistency. Experiments on a clinical CBCT–IOS dataset demonstrate that our approach consistently outperforms previous segmentation-based approaches.

Keywords: Registration · Keypoint correspondence · Cross-modality

1 Introduction

Accurate registration between cone-beam CT (CBCT) and intraoral scans (IOS) is a key component in digital orthodontics, enabling precise 3D dental modeling and clinically grounded treatment planning [1–3]. CBCT provides volumetric information that captures internal anatomical structures of the teeth and surrounding tissues, whereas IOS offers high-resolution surface geometry of the dental crowns. Aligning and integrating these complementary modalities [4, 5] in a common coordinate system is therefore essential for constructing high-quality anatomical tooth models. However, CBCT–IOS registration is still predominantly performed manually in clinical practice, leading to high labor costs and

* Corresponding author

† These authors contributed equally.

poor time efficiency, which motivates the need for a reliable automatic registration system.

Reliable correspondences are critical for rigid registration, as accurate point-to-point matches directly support precise estimation of the transformation. A widely adopted paradigm is feature-based registration [6–8], which detects distinctive keypoints [9, 10], encodes local context around each keypoint as a descriptor [11–13], and establishes correspondences by comparing keypoint similarity in feature space. While such a correspondence-driven strategy is also desirable for CBCT–IOS registration, the substantial modality discrepancy makes direct similarity comparison between keypoints difficult. Therefore, a method is needed to reduce the modality gap while ensuring accurate and stable correspondences.

In practice, CBCT–IOS registration typically follows a segmentation-based pipeline [14–16]: teeth are segmented in each modality, CBCT is converted to a surface-like representation [17] to match IOS modality, and final alignment is obtained via distance-based optimization [18–20], such as ICP [18]. While this strategy reduces modality gap, distance-based optimization does not guarantee reliable correspondences and is sensitive to outliers. Moreover, CBCT metal artifacts often destabilize segmentation, further limiting robustness. As a result, existing pipelines remain limited in reliability and correspondence-based dental registration at the feature-space level has not yet been sufficiently explored.

In this work, we propose a CBCT–IOS registration framework that estimates keypoint correspondences directly in feature space. Instead of converting both modalities into a common surface representation, we learn reliable cross-modal correspondences end-to-end even under appearance gaps between CBCT and IOS. Specifically, (1) We first learn a modality-aligned feature space in which cross-modal similarity becomes meaningful. (2) We then select registration-relevant keypoints using intra-modality saliency and a cross-modal confidence signal derived from the *sharpness* of the matching distribution (peaked matches indicate high confidence, while flat matches indicate ambiguity). (3) Given these keypoints, we predict correspondences by jointly learning inlier existence and pairwise matching correctness. All modules are trained under a single objective and thus mutually reinforce each other. At inference, we iteratively estimate the rigid transform via SVD while progressively shifting from feature similarity to geometric consistency to resolve residual ambiguity caused by repetitive tooth patterns and metal artifacts.

In summary, our main contributions are as follows:

- We reformulate CBCT–IOS registration as direct keypoint correspondence estimation in feature space, avoiding modality conversion to a common representation.
- We propose an end-to-end learning scheme that jointly optimizes embedding alignment, keypoint selection, and correspondence prediction under modality discrepancies.
- We introduce an inference-time refinement that iteratively estimates the rigid transform via SVD while annealing the matching criterion from feature similarity to geometric consistency.

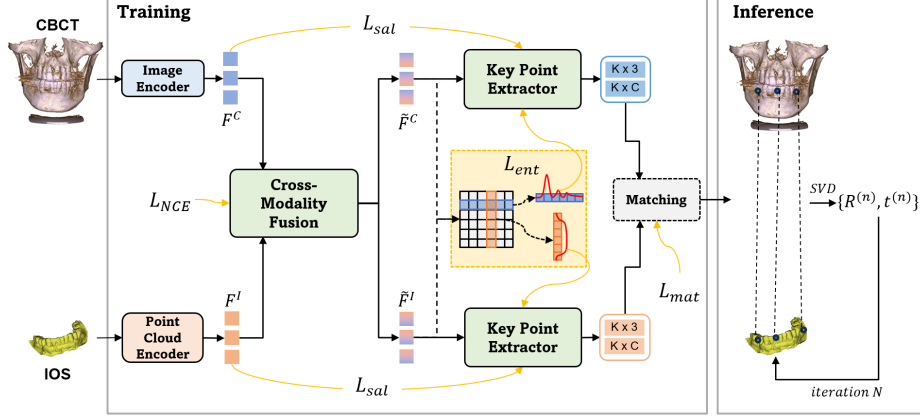


Fig. 1. Overall proposed framework. We first extract modality-specific features from CBCT and IOS. We then define modality-aligned fused features via a cross-modality fusion and modality contrastive learning (L_{NCE}). Using the fused features, we compute keypoint scores supervised by the entropy of the cross-modal matching distribution (L_{ent}) and an intra-modality saliency term (L_{sal}). After selecting the top- k keypoints, we estimate correspondences by correspondence learning (L_{mat}). During inference, the predicted correspondences are further refined via iterative SVD-based alignment, yielding the final rigid transformation.

2 Method

We propose an end-to-end framework for CBCT-IOS registration that directly estimates keypoint correspondences. An overview of our framework is illustrated in Fig. 1. First, we build a modality-aligned feature embedding to reduce the modality gap and to establish a shared feature space (Sec. 2.1). Second, we select registration-relevant keypoints using the entropy of cross-modal matching distributions together with an intra-modality saliency regularizer to favor structurally distinctive points (Sec. 2.2). Third, given the selected keypoints, we learn correspondence prediction by jointly supervising inlier existence and pairwise matching correctness (Sec. 2.3). Finally, at inference time we iteratively estimate the rigid transform with SVD while shifting the matching criterion from feature similarity to geometric distance (Sec. 2.5).

2.1 Modality-aligned Feature Embedding

We construct a shared latent space in which cross-modal similarity becomes meaningful and can support correspondence formation downstream. Let the CBCT input be a volume $\mathbf{V} \in \mathbb{R}^{D \times H \times W}$ and the IOS input be a point set $\mathbf{P} = \{p_j\}_{j=1}^{N_I}$. For CBCT, we sample a set of query locations $\mathbf{X}^C = \{x_i\}_{i=1}^{N_C}$ within a region of interest (i.e., high intensity ROI) and extract pointwise volumetric features using a 3D encoder E_C [21], yielding $\mathbf{F}^C = E_C(\mathbf{V}, \mathbf{X}^C) \in \mathbb{R}^{N_C \times d}$. For IOS, a point-cloud encoder E_I [21] produces $\mathbf{F}^I = E_I(\mathbf{P}) \in \mathbb{R}^{N_I \times d}$.

To exchange relational cues across modalities, we apply cross-attention fusion [22] so that CBCT features can attend to IOS features and vice versa, producing fused features $\tilde{\mathbf{F}}^C$ and $\tilde{\mathbf{F}}^I$. Finally, to explicitly align the fused representations in the shared space, we adopt an instance-level contrastive objective. We pool fused features to global embeddings $\mathbf{g}^C = \text{pool}(\tilde{\mathbf{F}}^C)$ and $\mathbf{g}^I = \text{pool}(\tilde{\mathbf{F}}^I)$, and apply an InfoNCE [23] loss over a batch so that paired CBCT-IOS embeddings from the same case are pulled together while embeddings from different cases are pushed apart. This embedding stage provides a consistent foundation for keypoint selection and correspondence learning in later stages.

2.2 Keypoint Extraction

The guiding principle is that a good keypoint should be (i) confidently matchable across modalities (low matching entropy) and (ii) structurally distinctive within its own modality (high saliency). Given fused features $\tilde{\mathbf{F}}^C$ and $\tilde{\mathbf{F}}^I$, we compute cross-modal similarities $M_{ij} = \text{sim}(\tilde{\mathbf{f}}_i^C, \tilde{\mathbf{f}}_j^I)$ and matching probabilities by row-wise softmax, e.g., $P_{ij}^{C \rightarrow I} = \text{softmax}_j(M_{ij})$ (symmetrically for $I \rightarrow C$). We define entropy $H_i = -\sum_j P_{ij} \log P_{ij}$, confidence $c_i = 1 - \bar{H}_i$, and predict the keypoint score $s_i = \sigma(h(\tilde{\mathbf{f}}_i))$. We supervise keypoint scores with the entropy-derived confidence,

$$\mathcal{L}_{ent} = \sum_{m \in \{C, I\}} \frac{1}{N_m} \sum_{i=1}^{N_m} \left(s_i^{(m)} - c_i^{(m)} \right)^2. \quad (1)$$

To encourage intra-modality distinctiveness, we compute a deterministic saliency score $\text{sal}_i^{(m)} \in [0, 1]$ directly from each modality-specific feature $\mathbf{f}_i^{(m)}$ using a D3Feat style spatial/channel formulation [13] (no additional saliency predictor). Let $z_i^{(m)}$ be the keypoint logit ($s_i^{(m)} = \sigma(z_i^{(m)})$) and $w_i^{(m)} = \frac{\exp(z_i^{(m)}/\tau_k)}{\sum_j \exp(z_j^{(m)}/\tau_k)}$ be the soft selection weight. We maximize the expected saliency of selected keypoints by

$$\mathcal{L}_{sal} = - \sum_{m \in \{C, I\}} \sum_{i=1}^{N_m} w_i^{(m)} \cdot \text{sg}(\text{sal}_i^{(m)}), \quad (2)$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operator, which prevents trivially increasing saliency by distorting features. We then optimize $L_{\text{kpe}} = \lambda_{\text{ent}} L_{\text{ent}} + \lambda_{\text{sal}} L_{\text{sal}}$. At inference, we select top- K points by s_i in each modality.

2.3 Correspondence Learning

We learn correspondence prediction over the selected K keypoints using a similarity matrix $\mathbf{M} \in \mathbb{R}^{K \times K}$, supervised by two complementary objectives: matching existence and pairwise correspondence correctness. First, we predict whether each keypoint has a valid correspondence (inlier existence). For a CBCT (or IOS) keypoint i , its existence score is defined as the maximum similarity in its row

(or column), $s(i) = \max_j \mathbf{M}_{ij}$. A binary label $\hat{s}_i \in \{0, 1\}$ is assigned using the ground-truth transform $\mathbf{T}_{gt}$: if $\|p_i^C - \mathbf{T}_{gt}(p_{j^*(i)}^I)\| < \delta$, the keypoint is considered an inlier ($\hat{s}_i = 1$), otherwise $\hat{s}_i = 0$. We optimize $s(i)$ using a binary cross-entropy loss:

$$\mathcal{L}_{ME} = \frac{1}{K} \sum_i [-\hat{s}_i \log s(i) - (1 - \hat{s}_i) \log(1 - s(i))]. \quad (3)$$

Second, we supervise pairwise correspondence correctness by encouraging high confidence for the ground-truth match $j^*(i)$:

$$\mathcal{L}_{PC} = \frac{1}{K} \sum_i [-\hat{s}_i \log \mathbf{M}(i, j^*(i))], \quad (4)$$

where $\hat{s}_i$ acts as an inlier mask to suppress the influence of outliers. The final matching loss is defined as $\mathcal{L}_{mat} = \mathcal{L}_{ME} + \mathcal{L}_{PC}$.

2.4 End-to-end Training and Final Loss

Our framework is optimized end-to-end so that embedding alignment, keypoint selection, and correspondence learning mutually reinforce each other under a single objective. This joint optimization mitigates the common failure mode of stage-wise pipelines, where early biases become fixed and propagate to downstream key point extraction and matching. The overall loss is defined as

$$\mathcal{L}_{total} = \lambda_{NCE} \mathcal{L}_{NCE} + \lambda_{kpe} \mathcal{L}_{kpe} + \lambda_{mat} \mathcal{L}_{mat}. \quad (5)$$

2.5 Rigid Transform Estimation and Inference-time Refinement

Given predicted correspondences, we estimate the rigid transform $(\mathbf{R}^{(t)}, \mathbf{t}^{(t)})$ using the closed-form SVD [24]. To reduce bias from near-tie feature matches caused by repetitive tooth geometry and metal artifacts, we refine at inference by iterating SVD with an annealed matching matrix:

$$\mathbf{M}^{(t)} = \alpha_t \mathbf{M}_{feat} + (1 - \alpha_t) \mathbf{M}_{dist}, \quad (6)$$

where $\mathbf{M}_{dist}$ is computed from pairwise distances after warping and $\alpha_t \searrow 0$. We alternate correspondence update and SVD until convergence.

3 Experiments

We evaluate our method on a clinical CBCT–IOS dataset from 46 patients, each with one CBCT volume and two IOS scans (maxilla and mandible), resulting in 92 pairs. The dataset was provided by Osstem Implant after de-identification, and the authors had no access to identifiable patient information. It was split into 64 training and 28 test pairs at the patient level. Although single-source, the dataset contains clinically meaningful variability, including full-arch scans, metal-artifact cases, and cases with reduced remaining teeth.

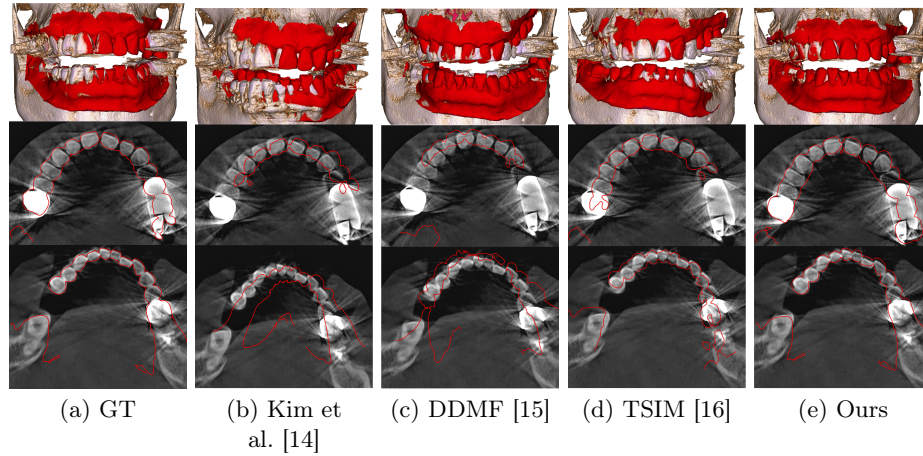


Fig. 2. Qualitative comparison of CBCT–IOS registration results produced by different methods along with the ground truth. The top row shows the 3D registration result, while the middle and bottom rows show 2D cross-sections of the maxilla and mandible, respectively.

Table 1. Quantitative comparison on CBCT–IOS registration.

Method	TRE@mm↓	RTE@mm↓	RRE@°↓
Kim et al. [14]	8.06 ± 14.79	7.59 ± 13.65	6.10 ± 8.22
DDMF [15]	5.06 ± 10.33	5.75 ± 8.92	4.01 ± 6.09
TSIM [16]	2.13 ± 6.67	3.53 ± 5.65	0.50 ± 1.38
Our method	0.30 ± 1.90	0.35 ± 0.75	0.03 ± 0.11

CBCT volumes were resampled to 1 mm^3 isotropic resolution and intensity-normalized; IOS point clouds were downsampled to 24,000 points by FPS [25]. To focus model capacity on regions relevant to dental alignment, we crop a tooth ROI following [26] and construct training inputs centered on this ROI. We used AdamW with learning rate 0.001, batch size 1, and loss weights $\lambda_{\text{NCE}} = \lambda_{\text{kpe}} = \lambda_{\text{mat}} = 1$. Experiments were run on an RTX A6000 (48GB) GPU, with an average inference time of 5.2s per pair. Ground-truth transformations were available for all pairs, and correspondence labels were derived from T_{gt} using the inlier threshold $\delta = 0.05\text{ mm}$, as described in Sec. 2.3. Performance is evaluated by Target Registration Error (TRE) for point-level accuracy, and by Relative Translation Error (RTE) and Relative Rotation Error (RRE) [27] for transformation discrepancy.

3.1 Experimental Results

Table 1 compares our method with segmentation-based CBCT–IOS registration pipelines. Since existing registration methods require both inputs to be represented as comparable voxels or sets of points, the existing CBCT–IOS pipeline

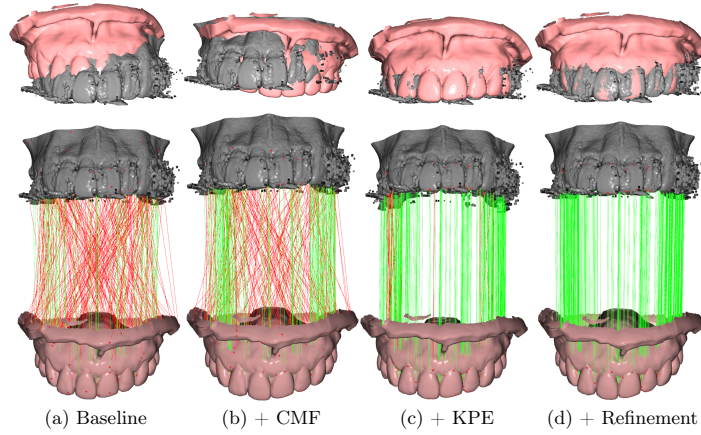


Fig. 3. Qualitative Ablation Results of Modules. Top row is 3D overlay registration results and bottom row is key point correspondence results. The number of points for visualization was set to 300. Green and red lines denote correct and incorrect matches, respectively.

Table 2. Ablation study of the proposed components.

Method	TRE@mm↓	RTE@mm↓	RRE@°↓
Baseline	12.82 ± 18.16	11.22 ± 27.54	17.25 ± 20.68
+ Cross-Modality Fusion	10.04 ± 14.98	9.43 ± 18.38	14.50 ± 15.47
+ Key Point Extractor	1.51 ± 2.10	1.19 ± 1.95	0.50 ± 0.55
+ Refinement	0.30 ± 1.90	0.35 ± 0.75	0.03 ± 0.11

converts to a surface representation via segmentation/surface extraction before registration. Our method avoids this modality-conversion step and achieves the lowest errors across all metrics, reducing TRE from $2.13mm$ (TSIM) to $0.30mm$. Fig. 2 shows that segmentation-based methods are vulnerable to severe metal artifacts, whereas our correspondence-based approach maintains stable alignment. Since metal-induced CBCT artifacts often lack stable IOS surface counterparts, the entropy-based confidence and correspondence supervision reduce reliance on artifact-dominated regions.

3.2 Ablation Study of Modules

We conduct module-wise ablations to quantify how each component contributes to correspondence stabilization and final registration accuracy. The Baseline directly learns feature matching from modality-specific encoder features. Table 2 shows a monotonic improvement in TRE/RTE/RRE when progressively adding components in the order **Baseline** $\rightarrow$ **+CMF (Cross-Modality Fusion)** $\rightarrow$ **+KPE (Key Point Extractor)** $\rightarrow$ **+Refinement**. The largest gain occurs when adding +KPE, indicating that selecting a sparse set of registration-relevant

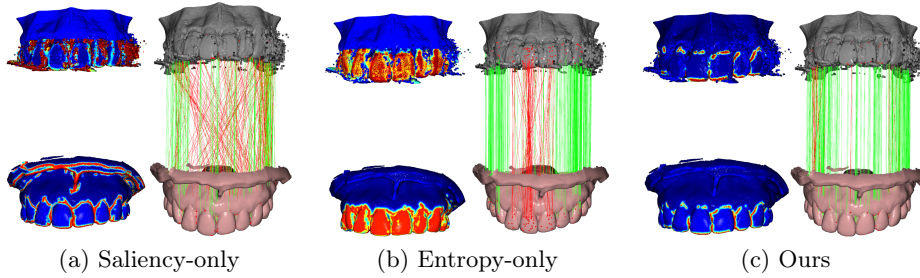


Fig. 4. Qualitative ablation results of keypoints. For each case, the left column shows the keypoint heatmap and the right column shows the keypoint correspondence result.

keypoints plays a critical role in correspondence-driven rigid registration. Qualitative comparisons in Fig. 3 are consistent with this trend. Baseline exhibits widespread mismatches, including failure cases where points on one tooth are incorrectly matched to points on a different tooth. +CMF yields more structured correspondences overall, yet still produces mismatches under repetitive patterns. +KPE concentrates keypoints and correspondence lines on structurally meaningful regions, substantially reducing outliers. Finally, +Refinement further reduces residual local misalignment and yields the most uniform convergence and the best final alignment.

3.3 Ablation Study of Keypoints

We further evaluate the impact of keypoint definitions on correspondence stability by comparing **Saliency-only**, **Entropy-only**, and **Saliency+Entropy (Ours)**. As shown in Fig. 4, Saliency-only concentrates heatmaps on locally distinctive regions (e.g., cusps/ridges or artifact-induced sharp areas) but produces more incorrect correspondences due to the lack of cross-modal matching confidence. Entropy-only yields more consistent matches by favoring confident cross-modal alignments; however, it still suffers from clustered mismatches in structurally ambiguous regions such as near-planar repetitive tooth surfaces. In contrast, the combined definition (Saliency+Entropy) achieves the highest proportion of correct correspondences and effectively suppresses outliers in metal-artifact regions. These results indicate that neither intra-modality saliency nor inter-modality confidence alone is sufficient; their combination is critical for robust correspondence-based CBCT-IOS registration.

4 Conclusion

We proposed a correspondence-driven CBCT-IOS rigid registration framework that avoids segmentation-based modality conversion by learning modality-aligned embeddings, entropy-guided keypoints, and correspondence prediction with iterative refinement. Experiments on a clinical dataset showed consistent improvements over segmentation-based pipelines.

This study follows the standard rigid-registration setting for CBCT–IOS dental alignment, assuming geometrically consistent teeth or hard-tissue-related structures and sufficient anatomical overlap across acquisitions. These assumptions may be weakened by large acquisition time gaps, treatment-related anatomical changes, severe IOS stitching distortion, or insufficient overlap. Although the dataset contains clinically meaningful variability, broader multi-center and multi-scanner validation remains future work.

Acknowledgments. This work was supported by the Materials and Components Technology Development Program funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and Korea Evaluation Institute of Industrial Technology (KEIT) (Project Number: RS-2023-KT224280).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Bianchi, J., Mendonca, G., Gillot, M., Oh, H., Park, J., Al Turkestani, N., Gurgel, M., Cevidanes, L.: Three-dimensional digital applications for implant space planning in orthodontics: A narrative review. *Journal of the World federation of orthodontists* **11**(6), 207–215 (2022)
2. Harrell Jr, W.E.: 3d diagnosis and treatment planning in orthodontics. In: *Seminars in Orthodontics*. vol. 15, pp. 35–41. Elsevier (2009)
3. Alaooffey, A.S., Asiri, M.A., Alhazmi, T.A.A., Alshetaiwi, A.A., Almobarak, A.M., Alqasir, Y.H., Algazlan, S.A., Alouf, H.S., Zakri, H.H.M., Al-Amri, F.G., et al.: Digital dentistry: transforming diagnosis and treatment planning through cad/cam and 3d printing. *Egyptian Journal of Chemistry* **67**(13), 1177–1190 (2024)
4. Shen, L., Lian, X., Zhang, L., Wang, G., Wang, Y., Zou, H., Wang, H.: Fusion modeling of intraoral scanning and cbct images on dental caries. In: *2023 42nd Chinese Control Conference (CCC)*. pp. 7394–7399. IEEE (2023)
5. Lee, S., Huang, G., Cowan, H., Huang, Y.: A novel technique for maxillary regional superimposition using merged cbcts and intra-oral scans. In: *Seminars in Orthodontics*. Elsevier (2025)
6. Qin, Z., Yu, H., Wang, C., Guo, Y., Peng, Y., Xu, K.: Geometric transformer for fast and robust point cloud registration. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 11143–11152 (2022)
7. Yew, Z.J., Lee, G.H.: Regtr: End-to-end point cloud correspondences with transformers. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 6677–6686 (2022)
8. Huang, Q., Guo, X., Wang, Y., Sun, H., Yang, L.: A survey of feature matching methods. *IET Image Processing* **18**(6), 1385–1410 (2024)
9. Jin, H., Shen, Y., Lou, J., Zhou, K., Zheng, Y.: Keypointdetr: an end-to-end 3d keypoint detector. In: *European Conference on Computer Vision*. pp. 374–390. Springer (2024)
10. Barroso-Laguna, A., Mikolajczyk, K.: Key. net: Keypoint detection by handcrafted and learned cnn filters revisited. *IEEE transactions on pattern analysis and machine intelligence* **45**(1), 698–711 (2022)

11. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: 2009 IEEE international conference on robotics and automation. pp. 3212–3217. IEEE (2009)
12. Han, X.F., Feng, Z.A., Sun, S.J., Xiao, G.Q.: 3d point cloud descriptors: state-of-the-art. *Artificial Intelligence Review* **56**(10), 12033–12083 (2023)
13. Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., Tai, C.L.: D3feat: Joint learning of dense detection and description of 3d local features. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6359–6367 (2020)
14. Kim, Y.J., Ahn, J.H., Lim, H.K., Nguyen, T.P., Jha, N., Kim, A., Yoon, J.: Novel procedure for automatic registration between cone-beam computed tomography and intraoral scan data supported with 3d segmentation. *Bioengineering* **10**(11), 1326 (2023)
15. Liu, J., Hao, J., Lin, H., Pan, W., Yang, J., Feng, Y., Wang, G., Li, J., Jin, Z., Zhao, Z., et al.: Deep learning-enabled 3d multimodal fusion of cone-beam ct and intraoral mesh scans for clinically applicable tooth-bone reconstruction. *Patterns* **4**(9) (2023)
16. Jang, T.J., Yun, H.S., Hyun, C.M., Kim, J.E., Lee, S.H., Seo, J.K.: Fully automatic integration of dental cbct images and full-arch intraoral impressions with stitching error correction via individual tooth segmentation and identification. *Medical Image Analysis* **93**, 103096 (2024)
17. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: *Seminal graphics: pioneering efforts that shaped the field*, pp. 347–353 (1998)
18. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: *Sensor fusion IV: control paradigms and data structures*. vol. 1611, pp. 586–606. Spie (1992)
19. Segal, A., Haehnel, D., Thrun, S., et al.: Generalized-icp. In: *Robotics: science and systems*. vol. 2, p. 435. Seattle, WA (2009)
20. Zhang, J., Yao, Y., Deng, B.: Fast and robust iterative closest point. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7), 3450–3466 (2021)
21. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems* **30** (2017)
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
23. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018)
24. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence* (5), 698–700 (1987)
25. Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y.: The farthest point strategy for progressive image sampling. *IEEE transactions on image processing* **6**(9), 1305–1315 (1997)
26. Chung, M., Lee, M., Hong, J., Park, S., Lee, J., Lee, J., Yang, I.H., Lee, J., Shin, Y.G.: Pose-aware instance segmentation framework from cone beam ct images for tooth segmentation. *Computers in Biology and Medicine* **120**, 103720 (2020)
27. Yew, Z.J., Lee, G.H.: Rpm-net: Robust point matching using learned features. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11824–11833 (2020)