ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion

Supplementary Material

Appendix

In this supplementary material, we first provide precise definitions of the evaluation metrics (Sec. A), then we describe the color point cloud dataset we generated (Sec. B), the details of the network architecture (Sec. C.1), and the implementation details (Sec. C.2). We also provide additional experiments, including the choice of color system (Sec. D.1), the noise resistance of ColorPCR (Sec. D.2), runtime analysis (Sec. D.3), details of low overlap registration results (Sec. D.4), and hyperparameter ablation (Sec. D.5). Finally, we provide visualizations of the registration results on the datasets (Sec. E.1) and the registration results in real-world scenarios (Sec. E.2).

A. Evaluation Metrics

Inlier Ratio (IR): The Inlier Ratio measures the fraction of point correspondences $(\mathbf{p_i}, \mathbf{q_j}) \in \mathbf{C_e}$ that are within a certain residual threshold under the ground truth transformation $\mathbf{T}_{\mathcal{P}}^{\mathcal{Q}}$. Here, $\mathbf{C_e}$ denotes the estimated correspondence set between point clouds $\mathcal{P}$ and $\mathcal{Q}$, and $\mathbf{T}_{\mathcal{P}}^{\mathcal{Q}}$ represents the ground truth transformation from $\mathcal{P}$ to $\mathcal{Q}$. A pair of correspondences is deemed to be matched if the Euclidean Norm of their residual is less than a threshold $\tau_1 = 10cm$. The Inlier Ratio for a pair of point clouds, $\mathcal{P}$ and $\mathcal{Q}$, can be computed using the following formula:

$$IR(\mathcal{P}, \mathcal{Q}) = \frac{1}{|\mathbf{C_e}|} \sum_{(\mathbf{p_i}, \mathbf{q_j}) \in \mathbf{C_e}} [\|\mathbf{T}_{\mathcal{P}}^{\mathcal{Q}}(\mathbf{p_i}) - \mathbf{q_j}\| < \tau_1], \quad (1)$$

where $[\cdot]$ is an indicator function that counts the number of correspondences with residuals within the threshold τ_1.

Feature Matching Recall (FMR): The Feature Matching Recall [2] is a metric that quantifies the proportion of point cloud pairs for which the Inlier Ratio surpasses a specified threshold, denoted as $\tau_2 = 5\%$. This metric quantifies the likelihood of obtaining the optimal transformation based on the estimated correspondence set $\mathbf{C_e}$, and this process is typically carried on through a robust pose estimator such as RANSAC. Given a dataset $\mathcal{D}$ that comprises $|\mathcal{D}|$ pairs of point clouds, the Feature Matching Recall can be calculated using the following formula:

$$FMR(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(\mathcal{P}, \mathcal{Q}) \in \mathcal{D}} [IR(\mathcal{P}, \mathcal{Q}) > \tau_2]. \quad (2)$$

In this formula, $[\cdot]$ is an indicator function that counts the number of point cloud pairs for which the Inlier Ratio exceeds the threshold τ_2.

Registration Recall (RR): Registration Recall [1] directly assesses the performance of point cloud registration. It quantifies the proportion of point cloud pairs for which the Root Mean Square Error (RMSE) falls within a specified threshold, denoted as $\tau_3 = 0.2m$. Given a dataset $\mathcal{D}$ that comprises $|\mathcal{D}|$ pairs of point clouds, Registration Recall is defined using the following formula:

$$RR(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(\mathcal{P}, \mathcal{Q}) \in \mathcal{D}} [RMSE(\mathcal{P}, \mathcal{Q}) < \tau_3]. \quad (3)$$

For each pair $(\mathcal{P}, \mathcal{Q}) \in \mathcal{D}$, the RMSE is computed as follows:

$$RMSE(\mathcal{P}, \mathcal{Q}) = \sqrt{\frac{1}{|\mathbf{C}|} \sum_{(\mathbf{p_i}, \mathbf{q_j}) \in \mathbf{C}} \|\mathcal{T}_{\mathcal{P}}^{\mathcal{Q}}(\mathbf{p_i}) - \mathbf{q_j}\|^2}, \quad (4)$$

where $\mathbf{C}$ is the ground truth correspondences and $\mathcal{T}_{\mathcal{P}}^{\mathcal{Q}}$ is the estimated transformation .

Patch Inlier Ratio (PIR): The Patch Inlier Ratio [8] is a metric that quantifies the proportion of superpoint (patch) matches that exhibit actual overlap under the ground-truth transformation. This metric provides an indication of the quality of the putative superpoint correspondences. The Patch Inlier Ratio can be computed using the following formula:

$$PIR = \frac{1}{|\hat{\mathbf{C}}|} \sum_{(\hat{\mathbf{p}}_i, \hat{\mathbf{q}}_j) \in \hat{\mathbf{C}}} [\exists \tilde{\mathbf{p}} \in \tilde{\mathcal{U}}(\hat{\mathbf{p}}_i), \tilde{\mathbf{q}} \in \tilde{\mathcal{U}}(\hat{\mathbf{q}}_j) \text{ s.t. } \|\tilde{\mathbf{p}} - \tilde{\mathbf{q}}\|_2 < \tau_3], \quad (5)$$

where the matching radius is $\tau_3 = 5cm$. Here, $\hat{\mathbf{C}}$ denotes the estimated superpoint correspondences, and function $\tilde{\mathcal{U}}$ calculates the up-sampling point set belonging to the input superpoint (*e.g.*, $\tilde{\mathcal{U}}(\hat{\mathbf{p}}_i) \subset \tilde{\mathcal{P}}$).

Relative Translation and Rotation Errors (RTE and RRE): Given the estimated transformation $\mathcal{T}_{\mathcal{P}}^{\mathcal{Q}} \in SE(3)$, which is composed of a translation vector $\mathbf{t_e} \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R_e} \in SO(3)$, the RTE Error and RRE from the ground truth pose $\mathbf{T}_{\mathcal{P}}^{\mathcal{Q}}$ can be calculated as follows:

$$RTE = \|\mathbf{t_e} - \mathbf{t}\| \quad (6)$$

and

$$RRE = \arccos\left(\frac{\text{trace}(\mathbf{R_e^T R}) - 1}{2}\right). \quad (7)$$

In these formulas, $\mathbf{t}$ and $\mathbf{R}$ represent the ground truth translation and rotation in $T_{\mathcal{P}}^{\mathcal{Q}}$, respectively.

B. Dataset Preprocessing

3DMatch [14] combines earlier data Analysis-by-Synthesis [11], 7Scenes [9], SUN3D [12], RGB-D Scenes v.2 [4], etc. Most of these scenes are obtained by taking depth images and RGB images with an RGBD camera, but the data collected by Analysis-by-Synthesis [11] only contains depth images. The official benchmark of the 3DMatch dataset divides the dataset into 54 scenes for training and 8 scenes for testing. The Analysis-by-Synthesis data only accounts for 8 scenes used for training, and the amount of data is very small. In order to generate color point cloud datasets, we delete the corresponding training data. Previous work [3] processed the 3DMatch dataset and generated the 3DLoMatch dataset, where the 3DMatch dataset contains point cloud pairs with an overlap of more than 30%, and the 3DLoMatch dataset contains point cloud pairs with an overlap of 10%-30%. To facilitate performance comparison with previous work, we follow their preprocessing data division, eliminate the Analysis-by-Synthesis part, and colorize the point cloud pairs in the datasets.

The specific method is as follows: We first use the original 3DMatch data to perform three-dimensional reconstruction for each scene, and then project the point clouds in the preprocessed data into the three-dimensional scene to colorize them. By doing this, we generate the Color3DMatch (C3DM) and Color3DLoMatch (C3DLM) datasets. We also provide a more detailed data division. To verify the performance of point cloud registration methods under extremely low overlap, we further divide the C3DLM test data into four parts with overlaps of 0.1-0.15, 0.15-0.2, 0.2-0.25, and 0.25-0.3 for future research use.

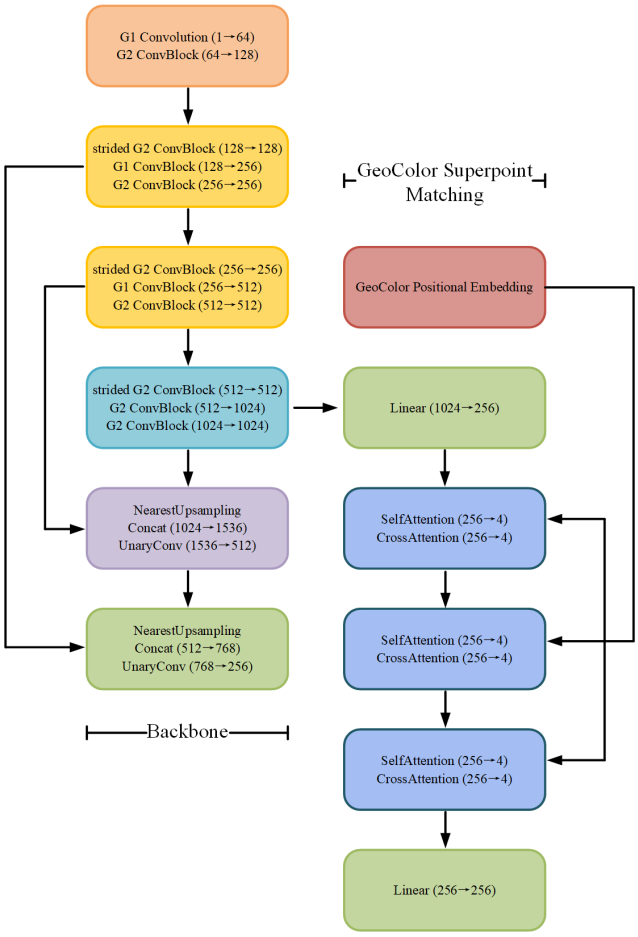


Figure 1. Network architecture of ColorPCR.

C. Implementation

C.1. Network Architecture

CEFE. We propose Color Enhanced Feature Extraction (CEFE) backbone for hierarchical point-wise feature extraction. CEFE is built upon the foundation of KPConv-FPN [5, 10]. Specifically, we employ the down-sampling strategy proposed in [10] to obtain hierarchical down-sampling points, with a total of four down-sampling operations. The voxel size used in the first level down-sampling is 2.5cm, and for each subsequent down-sampling, the voxel size is doubled. The specific execution process of CEFE is shown in the left half of Fig. 1. We encapsulate the G1 and G2 convolutions in a block to carry out the feature extraction process proposed by FPN. After performing the convolution operations, we apply group normalization with 8 groups, ultimately obtaining high-specificity hierarchical point features.

GeoColor Superpoint Matching. Upon the completion of CEFE, we match the superpoints down-sampled from it. The process of extracting superpoint features by GeoColor Superpoint Matching is shown in the right half of Fig. 1. We first use geo-color information to represent global context through GeoColor Positional Embedding. The superpoint features obtained by CEFE are fed into a linear projection to adjust the feature dimensions. Then, the embeddings and features are repeatedly processed three times in the self-attention and cross-attention modules proposed by [8], which thoroughly integrates color information and geometric information. After the final linear projection, the model generates superpoint features rich in semantic information.

C.2. Implementation Details

The training and evaluation processes of ColorPCR are conducted on a CPU Intel (R) Xeon (R) CPU E5-2640 v4 and GPU Tesla V100-PCIE-32GB, with PyTorch [7] as the implementation tool. ColorPCR is trained using the Adam optimizer, with an initial learning rate of 10^{-4} that decreases by 5% per epoch. The batch size is set to 1 and the weight decay is set to 10^{-6}. We set the matching radius to 5cm, consistent with the voxel size of a single down-

Feature extraction method	color system	RR (%) C3DM	C3DLM
GeoTransformer [8]	/	91.5	74.0
GeoTransformer [8]+rgb	rgb	93.5	73.9
CEFE (1)	rgb	97.0	85.5
CEFE (1)	hsv	**97.3**	**87.1**

Table 1. The effect of network structure under different color systems.

sampling, and point pairs smaller than the matching radius are considered points in the overlapping area. Our data augmentation method is consistent with that of [3]. During training, we randomly sample 128 ground-truth superpoint matches, while in testing, we sample 256 ones.

D. Additional Experiments

D.1. Color System

Since color noise has a more significant impact on the performance of the registration network compared to geometric noise, the method of introducing color is crucial to the performance of the registration network. Inspired by [6], we consider the Hue-Saturation-Value (HSV) color space and choose it to incorporate color information after conducting a series of experiments (Tab. 1).

We opt to introduce color information by performing multilevel color-enhanced feature extraction (CEFE) rather than simply using it as an extension of the spatial dimension (e.g., (x, y, z, h, s, v) or (x, y, z, r, g, b)). The simpler approach is not effective and may even perform worse, as this incorrect way of introducing color completely disrupts the spatial structure of the point cloud (as shown in the first two lines of the table). Additionally, experiments have shown that the RGB system is less effective compared to HSV, resulting in a performance gap in registration.

HSV can be easily computed from RGB values but offers a more stable color representation compared to the RGB color space. Experimental verification has shown that the HSV color space is more stable and better at integrating with geometric information (as shown in the last two lines of the table). In the HSV color space, H (Hue) can be computed as follows:

$$H = \begin{cases} 0^\circ, & \text{if } \Delta = 0 \\ 60^\circ \cdot \left(\frac{G-B}{\Delta} \mod 6\right), & \text{if } C_{\max} = R \\ 60^\circ \cdot \left(\frac{B-R}{\Delta} + 2\right), & \text{if } C_{\max} = G \\ 60^\circ \cdot \left(\frac{R-G}{\Delta} + 4\right), & \text{if } C_{\max} = B \end{cases} \tag{8}$$

and S can be computed as:

$$S = \begin{cases} 0, & \text{if } C_{\max} = 0 \\ \frac{\Delta}{C_{\max}}, & \text{if } C_{\max} \neq 0 \end{cases} \tag{9}$$

V is the simplest:

$$V = C_{\max}. \tag{10}$$

Here, R, G, B have been scaled to $[0, 1]$ and $C_{\max} = \max(R, G, B)$, $C_{\min} = \min(R, G, B)$ and $\Delta = C_{\max} - C_{\min}$.

D.2. ColorPCR is stable against color noise.

The acquisition of point clouds is often accompanied by noise. Intuitively, the impact of geometric noise on registration performance is relatively minor, as slight positional deviations of individual points often result in changes in local point density. Conversely, color noise has a far greater impact on registration performance than geometric noise, as color deviations can significantly affect feature extraction of points and disrupt the spatial structure information of the point cloud.

However, ColorPCR demonstrates stability against color noise. During the feature extraction phase, the Color Enhanced Feature Extraction (CEFE) employs a hierarchical feature extraction approach. The entire registration process does not involve noise-prone dense points, but instead uses noise-diluted down-sampled points. The way CEFE introduces color is not through geo-color feature extraction, but through color enhanced feature extraction, where color serves only as a factor to enhance geometric features, without disrupting the spatial structure of the point cloud. Moreover, the object of GeoColor Superpoint Matching is low-noise superpoints, further achieving geo-color fusion, thereby greatly diluting the effect of noise.

To verify the noise resistance of ColorPCR, we add noise with different standard deviations to the color of dense points (with RGB values constrained within [0, 1]). Experimental results show that ColorPCR's registration performance is quite stable under different noise influences.

In the first experiment, minuscule noise of normal distribution is added to the dense points, and the new RGB values falling below 0 or exceeding 1 were truncated, employing the color values with superimposed noise as the input for ColorPCR (Tab. 2). Under such circumstances, the performance of ColorPCR is quite stable. Only when the standard deviation of the noise escalates to 0.05, does the performance of the registration experience a marginal downturn.

Then we modify the method of noise addition: a certain proportion of points are randomly selected, their color values are randomly assigned within the range of [0, 1], and the newly generated dense points are then fed into ColorPCR for the execution of registration (Tab. 3). The registration performance of the network remains remarkably stable when noise points are relatively scarce. It is only when a substantial quantity of noise points are chosen (*e.g.*, 10%), that the registration efficacy experiences a slight decline.

The two types of noise described above emulate the influence of factors such as lighting conditions on the overall

Noise	C3DM				C3DLM			
	PIR	FMR	IR	RR	PIR	FMR	IR	RR
Without noise	89.2	99.6	73.5	96.4	63.1	96.5	51.0	88.1
Noise-0.0001	89.2	99.6	73.4	96.5	63.0	96.8	51.0	87.5
Noise-0.0005	89.2	99.6	73.4	96.2	63.0	96.8	51.0	87.9
Noise-0.001	89.2	99.6	73.3	96.2	63.0	96.8	50.9	87.8
Noise-0.002	89.2	99.5	73.3	96.6	63.0	97.0	50.9	87.2
Noise-0.005	89.2	99.5	73.1	96.2	62.9	96.8	50.6	88.2
Noise-0.01	89.0	99.6	72.6	96.2	62.5	95.9	50.0	87.4
Noise-0.05	87.6	99.2	68.1	94.4	58.2	93.1	43.7	80.9
Noise-0.1	85.3	99.3	63.6	92.8	52.9	88.9	37.7	73.8

Table 2. The registration results of ColorPCR under different levels of noise.

Noise	C3DM				C3DLM			
	PIR	FMR	IR	RR	PIR	FMR	IR	RR
Without noise	89.2	99.6	73.5	96.4	63.1	96.5	51.0	88.1
Noise-0.1%	89.2	99.6	73.3	97.0	63.0	96.6	50.8	88.4
Noise-0.2%	89.1	99.5	73.2	96.7	62.9	96.7	50.7	87.9
Noise-0.5%	89.1	99.6	72.7	97.1	62.7	96.7	50.2	87.4
Noise-1.0%	88.8	99.4	72.1	95.7	62.3	96.4	49.4	87.5
Noise-2.0%	88.7	89.5	71.0	96.7	61.9	96.1	48.4	86.4
Noise-5.0%	88.3	99.5	69.0	96.4	60.4	95.1	45.8	84.2
Noise-10.0%	87.4	99.0	66.6	95.2	58.5	94.7	43.0	81.7
Noise-20.0%	86.1	99.4	63.6	92.3	55.6	93.5	39.6	77.7

Table 3. The registration outcome of ColorPCR subsequent to the random attribution of a certain proportion of dense points colors.

color collection of the point cloud, as well as the occurrence of individual point cloud color inaccuracies due to low device precision. ColorPCR demonstrates exceptional stability under these two types of noise, accomplishing robust point cloud registration.

D.3. Running Time

The primary constraints on the registration time for the majority of preceding methodologies are the procurement of priors and the convergence of RANSAC. The registration procedure of ColorPCR necessitates no priors and employs a local-global registration strategy as an alternative to the more time-intensive RANSAC, resulting in a quicker execution time. Preceding methodologies, such as GeoTransformer [8], have considered both execution time and registration precision, introducing a robust registration network. Tab. 4 juxtaposes the RR and execution time of ColorPCR and GeoTransformer. Both methodologies are identical during the transformation estimation phase, hence the table solely contrasts the execution time of the backbone network, which extracts hierarchical point features, and the transformer network, which extracts superpoint features. As evidenced by the data in the table, the procedure of color extraction by CEFE barely augments the operational time. The GeoColor Superpoint Matching process, which entails $\Delta\mathcal{H}$ computation, marginally elevates the calculation time. Despite the nearly equivalent total time, ColorPCR has accomplished a substantial enhancement in RR.

Model	RR (%)		Time (s)		
	C3DM	C3DLM	Points	Superpoints	Total
GeoTransformer [8]	91.5	74.0	0.027	0.052	0.079
ColorPCR	96.5	88.3	0.027	0.092	0.119

Table 4. The registration recall and running time of ColorPCR and GeoTransformer. Within the table, *Points Time* denotes the duration required for the extraction of hierarchical point features, whereas *Superpoints Time* signifies the operational time of the transformer architecture that further extracts superpoint features.

D.4. Detailed Results With Different Overlaps

We have conducted a detailed validation of the Feature Matching Recall (FMR), Inlier Ratio (IR), and Registration Recall (RR) of ColorPCR and other geo-only methods under various low overlaps, with the data listed in Tab. 5. ColorPCR achieves the best results in all metrics listed in the table and realizes a breakthrough improvement under extremely low overlap. In the challenging registration environment with an overlap of 0.1-0.15, ColorPCR achieves the highest FMR of 94.1%, IR of 54.3%, and RR of 78.2%, while the best results of other geo-only methods are only 77.6% for FMR, 43.1% for IR, and 55.1% for RR. ColorPCR, by utilizing color information to guide the registration process, obtains precise hierarchical point-wise features, thereby identifying points in the overlap areas and achieving a higher IR. Moreover, it achieved relatively accurate registration in most cases of the dataset, resulting in higher FMR and RR.

D.5. Hyperparameter Ablation

We adjust the hyperparameters introduced in ColorPCR. For the convenience of parameter effect verification, we use CEFE (1) for the experiment and adjust the parameter σ_c for color enhancement in the first level. The experimental results are shown in Tab. 6. The performance of ColorPCR is very stable under different σ_c, so in the experiments described in the main text, we choose the most intuitive hyperparameter $\sigma_c = 1.00$ for color enhancement. In addition, the embedding hyperparameter σ_{HD} introduced in the GeoColor Superpoint Matching Module is set to 0.1 in our experiments.

E. Visualizations

E.1. Qualitative Results

We provide additional qualitative results on the Color3DLoMatch dataset in Fig. 2, where we present visualizations of the input, ground truth, ColorPCR, and GeoTransformer [8] results. Both GeoTransformer and our ColorPCR utilize LGR for registration. As mentioned in Geo-

#Samples	0.1-0.15					0.15-0.2					0.2-0.25					0.25-0.3				
	5000	2500	1000	500	250	5000	2500	1000	500	250	5000	2500	1000	500	250	5000	2500	1000	500	250
Feature Matching Recall (%)↑																				
Predator [3]	59.1	59.9	60.3	60.7	60.0	73.9	75.3	73.5	73.0	69.8	81.2	80.9	80.7	80.5	81.3	88.8	89.1	89.5	88.5	88.4
Cofinet [13]	69.6	69.5	69.2	67.8	68.3	83.5	83.5	84.5	84.6	83.2	91.9	91.0	90.8	91.1	89.8	95.5	95.5	95.5	95.5	95.2
GeoTransformer [8]	76.0	76.1	77.6	76.5	76.1	89.1	89.2	89.0	88.5	88.1	92.6	92.8	92.8	92.6	92.3	95.4	95.5	95.2	96.2	95.1
ColorPCR (ours)	**92.7**	**92.7**	**94.1**	**93.5**	**93.0**	**96.8**	**97.8**	**97.1**	**97.2**	**96.8**	**99.4**	**99.4**	**99.5**	**99.4**	**99.3**	**97.5**	**97.5**	**97.5**	**97.5**	**97.6**
Inlier Ratio (%)↑																				
Predator [3]	13.5	14.7	15.2	15.1	14.0	21.0	22.5	22.7	22.1	20.9	25.6	26.7	27.3	26.2	24.9	32.3	33.6	34.3	33.1	31.0
Cofinet [13]	16.4	17.7	18.3	18.6	18.7	24.4	25.8	26.5	26.9	26.8	28.0	29.6	30.4	30.8	30.8	33.4	35.1	35.9	36.2	36.6
GeoTransformer [8]	29.2	33.4	39.1	41.5	43.1	42.9	48.6	54.7	57.4	59.1	48.6	55.1	61.3	64.5	65.4	55.8	62.7	69.0	71.7	73.7
ColorPCR (ours)	**38.8**	**42.5**	**49.1**	**52.1**	**54.3**	**49.6**	**57.8**	**61.7**	**64.7**	**66.8**	**54.7**	**60.4**	**67.0**	**69.8**	**72.1**	**60.2**	**66.0**	**72.2**	**75.0**	**77.0**
Registration Recall (%)↑																				
Predator [3]	38.2	38.4	41.9	42.4	35.7	52.0	54.7	54.8	53.3	49.1	68.9	68.2	66.6	61.7	59.8	75.5	75.3	74.9	76.2	72.0
Cofinet [13]	47.1	49.6	50.4	47.7	46.0	65.8	64.9	65.2	66.6	63.8	72.6	73.8	72.3	70.6	70.8	81.6	83.0	83.1	82.9	81.1
GeoTransformer [8]	50.6	55.1	53.4	54.1	54.5	72.3	70.0	72.4	69.9	68.9	75.4	74.7	76.3	75.3	75.0	82.9	83.9	82.5	80.7	83.1
ColorPCR (ours)	**76.2**	**77.9**	**78.2**	**75.9**	**74.5**	**85.3**	**87.8**	**87.4**	**85.2**	**84.9**	**95.0**	**93.4**	**95.4**	**92.1**	**92.8**	**94.3**	**93.1**	**93.8**	**93.2**	**93.2**

Table 5. A detailed comparison of the registration effects of ColorPCR and geo-only methods under different overlaps. #Samples in the table represents the number of correspondences selected by RANSAC.

σ_c	C3DM				C3DLM			
	PIR	FMR	IR	RR	PIR	FMR	IR	RR
1.25	88.9	99.6	73.0	97.4	62.2	95.6	50.2	86.4
2.00	88.7	99.2	73.6	95.8	61.9	96.0	50.8	85.7
5.00	88.9	99.6	74.3	95.2	60.5	95.8	49.6	85.3
1.00	87.9	99.6	72.1	97.3	60.6	96.2	49.2	87.1

Table 6. Ablation experiments on the hyperparameter σ_c, which is used for color enhancement in CEFE.

Transformer, it severely fails in challenging registration scenarios with extremely low overlap and geometric distinctiveness. Due to the low geometric uniqueness of the overlapping area, it cannot effectively identify the points in the overlapping area, leading to misalignment. However, ColorPCR, guided by color information, takes advantage of both the geometric shape and color features of the overlapping area, thereby successfully achieving registration in such challenging scenarios.

E.2. Real World Registration

ColorPCR achieves high-precision registration under challenging scenarios, thus performing well in difficult registration situations in the real world. To validate this conclusion, we sample some color point clouds from real scenes using the Intel RealSense d455 RGBD camera and register them. The specific registration results are shown in Fig. 3, where we present two input point clouds and the registration results of ColorPCR. The colorless point clouds are presented to distinguish two point clouds.

References

[1] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5556–5565, 2015. 1

[2] Haowen Deng, Tolga Birdal, and Slobodan Ilic. PPF-FoldNet: Unsupervised learning of rotation invariant 3D local descriptors. In *Eur. Conf. Comput. Vis.*, pages 602–618, 2018. 1

[3] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3D point clouds with low overlap. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4267–4276, 2021. 2, 3, 5

[4] Kevin Lai, Liefeng Bo, and Dieter Fox. Unsupervised feature learning for 3D scene labeling. In *IEEE International Conference on Robotics and Automation*, pages 3050–3057. IEEE, 2014. 2

[5] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2117–2125, 2017. 2

[6] Hao Men, Biruk Gebre, and Kishore Pochiraju. Color point cloud registration with 4D ICP algorithm. In *IEEE International Conference on Robotics and Automation*, pages 1511–1516. IEEE, 2011. 3

[7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 2

[8] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11143–11152, 2022. 1, 2, 3, 4, 5

[9] Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in RGB-D images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2930–2937, 2013. 2

[10] Hugues Thomas, Charles Ruizhongtai Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and

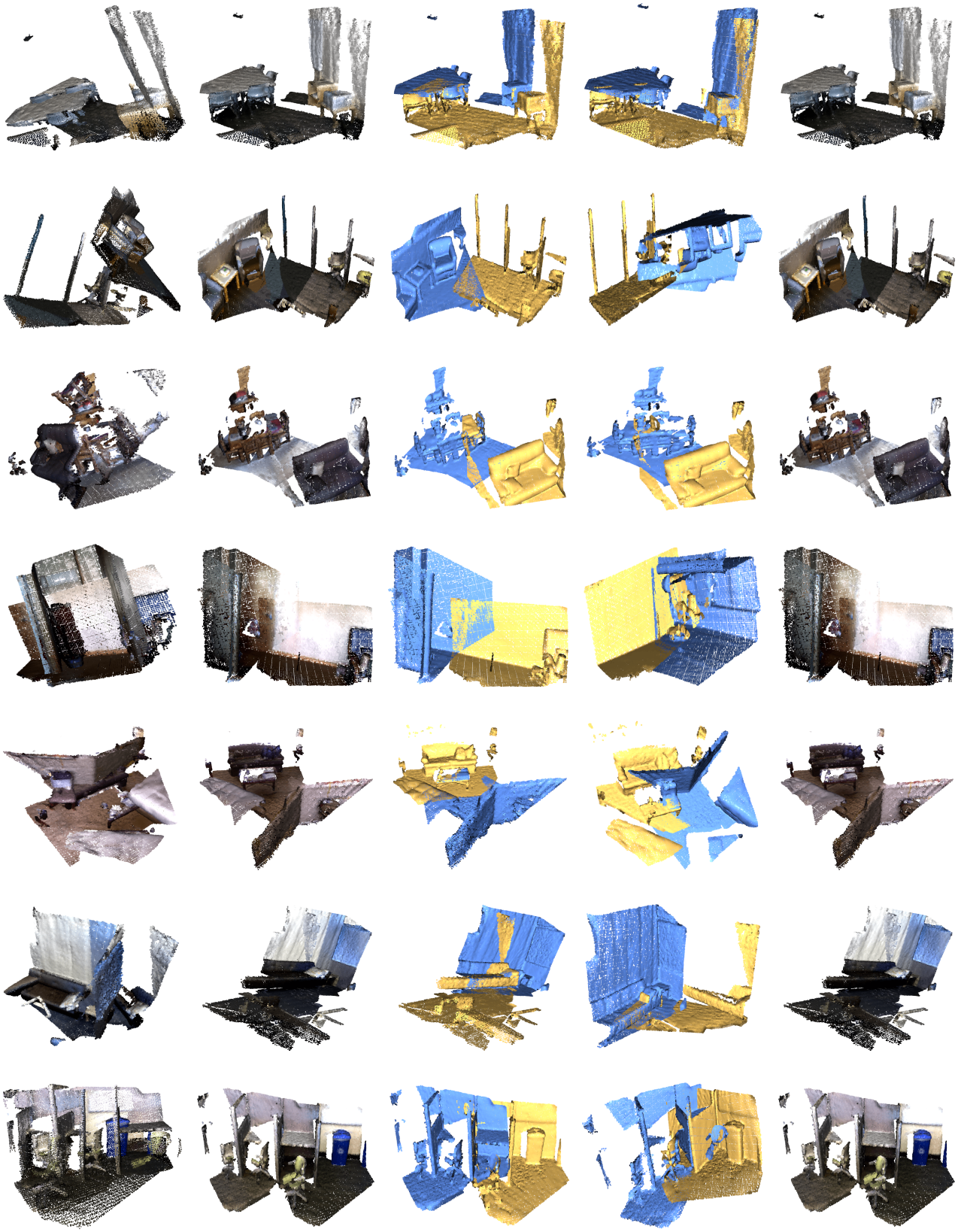
Leonidas J Guibas. KpConv: Flexible and deformable convolution for point clouds. In *Int. Conf. Comput. Vis.*, pages 6411–6420, 2019. 2

[11] Julien Valentin, Angela Dai, Matthias Nießner, Pushmeet Kohli, Philip Torr, Shahram Izadi, and Cem Keskin. Learning to navigate the energy landscape. In *International Conference on 3D Vision*, pages 323–332. IEEE, 2016. 2

[12] Jianxiong Xiao, Andrew Owens, and Antonio Torralba. Sun3D: A database of big spaces reconstructed using sfm and object labels. In *Int. Conf. Comput. Vis.*, pages 1625–1632, 2013. 2

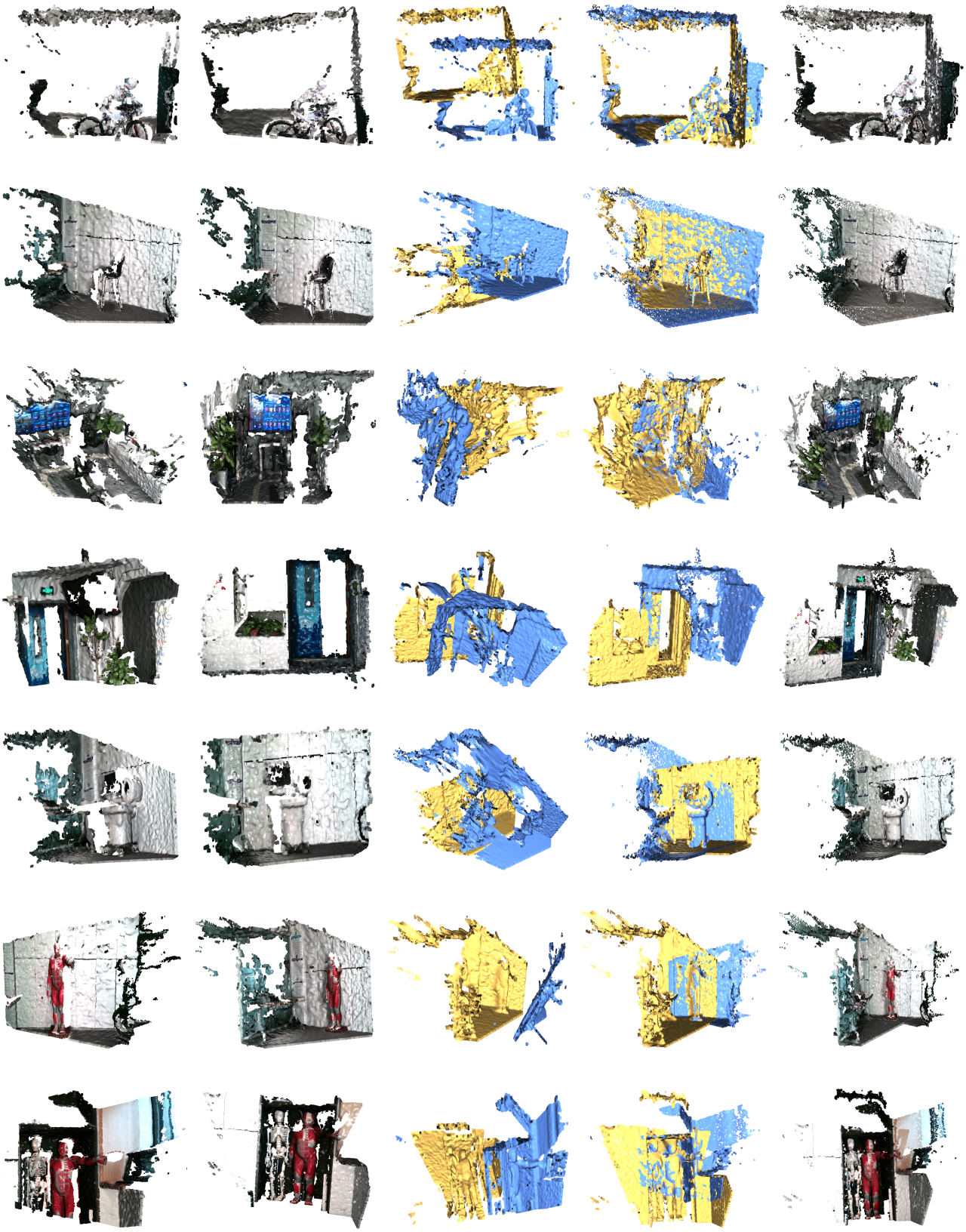
[13] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. CofiNet: Reliable coarse-to-fine correspondences for robust pointcloud registration. *Advances in Neural Information Processing Systems*, 34:23872–23884, 2021. 5

[14] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1802–1811, 2017. 2



(a) input (b) colorized ground truth (c) ground truth (d) GeoTransformer (e) ColorPCR

Figure 2. Registration results on Color3DMatch and Color3DLoMatch.



(a) point cloud one (b) point cloud two (c) input (d) ColorPCR (e) ColorPCR

Figure 3. Real world registration results of ColorPCR. The colorless point clouds are presented to distinguish two point clouds.