

R3PM-Net: Real-time, Robust, Real-world Point Matching Network

Supplementary Material

A.1. Datasets

ModelNet40 Dataset. [31] is a collection of synthetic 3D CAD models from 40 object categories. To generate point clouds, 2,000 points are randomly sampled from the mesh surfaces of each model and normalized to a unit sphere.

The Sioux-Cranfield Dataset. is a diverse collection of 13 objects designed to evaluate model robustness across varying data qualities. The CAD models are presented in Fig. 6 while Table 8 provides a structured breakdown of the composition of this dataset.

Sioux-Scan Data. represents the core challenge of this work: the registration of raw event-camera scans against digital models. Unlike the simulated datasets described above, these pairs exhibit a genuine domain gap. The **Target** point clouds were derived from 3D CAD models of seven small objects (also present in the Sioux-Cranfield dataset). The **Source** point clouds were acquired by scanning the physical objects using an event camera within a 3D Quality Control Setup [24] developed by Sioux Technologies (Fig. 5). Before processing, gross outliers were filtered using *Cloud-Compare* [12]. As Fig. 7 indicates, these scans exhibit severe unavoidable flaws which are not present in synthetic benchmarks, including high sparsity, sensor noise, and significant occlusions, particularly on object undersides and sharp edges hidden from the camera’s field of view.

Category	Source Type	Qty
Sioux (Reconstructed)	Photogram.	4
Sioux (Synthetic)	CAD Models	3
Cranfield [6]	Pristine	6
Total	—	13

Table 8. Composition of the Sioux-Cranfield Dataset. Sioux (Reconstructed) and Sioux (Synthetic) are also used in Sioux-Scans dataset to produce Target point clouds.

A.2. Evaluation Metrics

Following the literature and based on the nature of the data, this paper uses several metrics to capture different aspects of registration quality, from geometric accuracy to computational efficiency. Table 9 presents a summary of these metrics, their types, strengths, limitations, and ground truth requirements.

Relative Rotation Error (RRE): Similar to [26], [27], and [34], relative rotational error is defined as the deviation



Figure 5. 3D Quality Control Setup developed by Sioux Technologies [24]. This setup leverages event-camera technology to provide fast, high-fidelity 3D scanning suitable for in-line real-time quality assurance for specialized products. It aims to deliver optimized process control in applications including 3D metal printing, integrated electronics in molded parts, and dental prosthetics.

from the ground truth rotation matrices:

$$\text{RE} = \arccos\left(\frac{\text{Tr}(R) - 1}{2}\right), \quad (8)$$

Where Tr indicates the trace of a matrix and R is the relative rotation matrix, calculated as

$$R = R_{gt}R_{est}^\top, \quad (9)$$

where R_{gt} is the ground truth and R_{est} is the estimated rotation matrices.

This rotation error is not specific to any axes and captures the overall misalignment of the point clouds in the 3D space caused by the difference of the estimated and ground-truth rotation matrices. However, a significant limitation is its inability to account for object symmetries (like a cylinder or a cube). It treats only the labeled ground truth as valid, and incorrectly penalizes equivalent rotations that result in physically identical alignments.

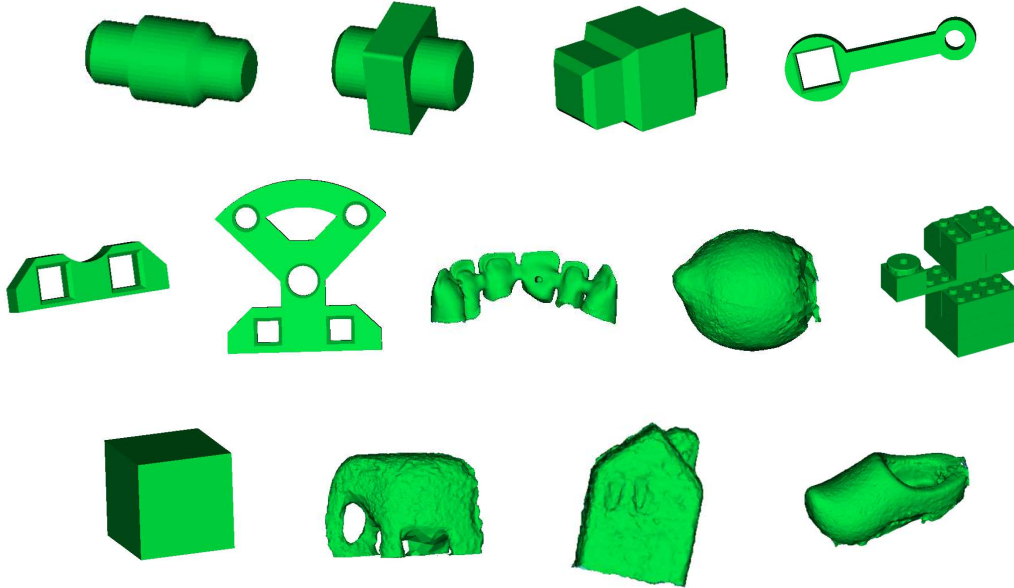


Figure 6. CAD models of the Sioux-Cranfield dataset. The first six belong to the Cranfield Assembly benchmark [6] and the rest are contributions of this paper (Sioux dataset).

Object	Image	Target	Source	Object	Image	Target	Source
Teeth				Elephant			
Lego				House			
Lime				Shoe			
Cube							

Figure 7. Sioux-Scans point cloud data overview. Target (blue) and Source (yellow) point clouds for seven distinct objects.

Relative Translation Error (RTE): Following [26], [27], and [34], the relative translation error is considered the Euclidean distance between the ground truth and the estimated translation vectors.

$$TE = \|t_{gt} - t_{est}\|_2. \quad (10)$$

Although rotation and translation errors are commonly re-

ported in research, they do not provide an intuitive understanding of registration quality. Furthermore, they do not support an easy or fair comparison; one approach might get a lower rotation error but a higher translation error than another, or vice versa. As a result, relying only on these two metrics does not provide an objective way to assess the performance of a method [9, 11]. In addition, computing these errors requires information about the ground-truth

Metric	Type	Pros	Cons	GT-Free
RRE	Geodesic dist.	Scale-invariant	Less intuitive, not always objective	✗
RTE	Euclidean dist.	Scale-invariant	Less intuitive, not always objective	✗
Chamfer Dist.	Euclidean dist.	Suitable for symmetric shapes	Scale-variant	✓
Fitness	Overlap ratio	Intuitive	Scale-variant	✓
Inlier RMSE	Euclidean dist.	Practical	Scale-variant, can be misleading with poor alignment	✓
Time	Seconds	Important for real-time use	Does not reflect accuracy	✓

Table 9. Summary of Evaluation Metrics: Types, strengths, limitations, and ground-truth (GT) requirements. Scale-invariant metrics do not depend on the point cloud size. GT-free metrics are suitable for real-world scenarios whereas GT-based metrics are used in simulations.

transformation, which is not available in real cases.

Chamfer Distance (CD): This metric calculates the average distance between pairs of nearest neighbors of the resulting \tilde{X} and the target Y point cloud [3] [29]:

$$\begin{aligned} \text{CD}(\tilde{X}, Y) = & \frac{1}{N} \sum_{i=1}^N \|\tilde{x}_i - \text{NN}(\tilde{x}_i, Y)\|_2 \\ & + \frac{1}{M} \sum_{j=1}^M \|y_j - \text{NN}(y_j, \tilde{X})\|_2. \end{aligned} \quad (11)$$

Our implementation of the Chamfer distance based on [29] does not require exact correspondences, as it calculates the distances from each point in a point cloud to its nearest neighbor in the other, rather than its exact corresponding point. Additionally, unlike RRE, it does not unfairly penalize alternative alignments for symmetric objects [34]. However, outliers, oclusions, and high point cloud sparsity result in high error values as the distances between points increase. Additionally, this metric is scale-variant, meaning that the same transformation applied to a larger point cloud would result in a different error than if the point cloud were smaller [11].

Fitness: Fitness measures overlapping areas of two point clouds. The better the alignment, the higher the fitness score. This score is defined and implemented as the ratio of the number of inlier correspondences to the total number of points in the target point cloud [41].

$$\text{Fitness} = \frac{|\mathcal{I}|}{M}, \quad (12)$$

where inlier correspondences (\mathcal{I}) refer to pairs of nearest neighbor points whose Euclidean distances are below a pre-defined threshold τ :

$$\mathcal{I} = \{(i, j) | \tilde{x}_i \in \tilde{X}, y_j \in Y, \|R^* \tilde{x}_i + t^* - y_j\|_2 < \tau\}, \quad (13)$$

where \tilde{X} and Y are the result and target point clouds, respectively. $R^* \in \mathbb{R}^{3 \times 3}$ and $t^* \in \mathbb{R}^{3 \times 1}$ are the estimated rotation

matrix and translation vector applied to the source to create \tilde{X} .

As Eq.12 implies, the maximum achievable fitness score is 1, indicating perfect alignment where every point in the target has a corresponding inlier in the transformed source. This metric is scale-variant because it relies on the number of points.

Inlier RMSE: This metric, often reported with Fitness, measures the average alignment error for all inlier correspondences. This is computed as the root mean square of the Euclidean distances between the inlier correspondences [41]:

$$\text{Inlier RMSE} = \sqrt{\frac{1}{|\mathcal{I}|} \sum_{(i,j) \in \mathcal{I}} \|R^* \tilde{x}_i + t^* - y_j\|_2}. \quad (14)$$

Inlier RMSE should be interpreted in the context of other metrics such as Chamfer Distance and fitness. The reason is that in the case of a failed registration where no inliers are detected ($\mathcal{I} = \emptyset$), the RMSE value becomes zero, which can misleadingly suggest high-quality alignment.

Similar to Chamfer distance and fitness, the inlier RMSE is scale-dependent. However, these three metrics do not require ground-truth transformations and, therefore, can serve as useful guidelines for the real-world data. Nevertheless, qualitative analysis is necessary to confirm success.

A.3. Results

The performance of the baseline methods on real-life data is presented in detail in Table 10. The results are averaged across seven independent runs for each of the seven test objects. In this evaluation, a method is considered successful on a case only if it achieves accurate registration in at least four of the seven runs. As mentioned in the paper, the failure or success of the registration is decided based on visual inspection.

Method	Data	Teeth	Lime	Cube	Lego	Eleph.	House	Shoe	SR (%)	Time
RPMNet [34]	CD	.205	.270	.290	.218	.173	.264	.191	28.6	0.042s
	Fit.	1.00	1.00	1.00	1.00	1.00	1.00	.968		
	RMSE	.059	.047	.023	.065	.064	.132	.106		
	Status	failed	success	success	failed	failed	failed	failed		
Predator [14]	CD	.185	.270	.289	.210	.158	.284	.092	28.6	0.038s
	Fit.	1.00	1.00	1.00	1.00	1.00	.992	1.00		
	RMSE	.048	.048	.023	.055	.061	.089	.059		
	Status	failed	success	success	failed	failed	failed	failed		
GeoTrans. [18]	CD	.324	.260	.295	.259	.184	.260	.183	28.6	0.042s
	Fit.	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
	RMSE	.053	.041	.024	.068	.067	.132	.101		
	Status	failed	success	success	failed	failed	failed	failed		
RegTR [35]	CD	.196	.270	.292	.261	.190	.267	.105	28.6	0.038s
	Fit.	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
	RMSE	.059	.047	.023	.055	.068	.132	.067		
	Status	failed	success	success	failed	failed	failed	failed		
LoGDesc [22]	CD	.186	.366	.292	.207	.164	.222	.092	28.6	0.043s
	Fit.	1.00	.989	1.00	1.00	1.00	1.00	1.00		
	RMSE	.045	.081	.024	.055	.069	.052	.059		
	Status	failed	failed	success	failed	failed	success	failed		
R3PM-Net (ZS) (ours)	CD	.178	.326	.510	.381	.169	.295	.104	28.6	<u>0.041s</u>
	Fit.	1.00	1.00	.912	1.00	1.00	1.00	1.00		
	RMSE	.047	.060	.102	.107	.070	.095	.066		
	Status	success	failed	success	failed	failed	failed	failed		
R3PM-Net (FT) (ours)	CD	.144	.288	.735	.398	.167	.222	.148	42.9	0.045s
	Fit.	1.00	1.00	.795	1.00	1.00	1.00	1.00		
	RMSE	.034	.042	.172	.124	.075	.052	.096		
	Status	success	success	failed	failed	failed	success	failed		

Table 10. Complete performance comparison of baseline methods on real-life industrial data. Results are averaged over seven independent runs across seven test objects. A successful registration is defined as achieving accurate alignment, verified visually, in at least four out of seven trials.