SUPPLEMENTARY MATERIALS
GFrames: Gradient-Based Local Reference Frame for 3D Shape Matching

Simone Melzi
University of Verona
simone.melzi@univr.it

Riccardo Spezialetti
University of Bologna
riccardo.spezialetti@unibo.it

Federico Tombari
TU Munich
tombari@in.tum.de

Michael M. Bronstein
Imperial College London / USI
m.bronstein@imperial.ac.uk

Luigi Di Stefano
University of Bologna
luigi.distefano@unibo.it

Emanuele Rodolà
Sapienza University of Rome
rodola@di.uniroma1.it

In these pages we collect additional results and details that due to lack of space were not included in the paper.

1. RGB-D SLAM Dataset

We perform additional experiments on the RGB-D SLAM dataset [3]. In Figures 1 and 2 we visualize the LRF repeatability on two view pairs of different room interiors.

In Figure 3 we visualize an overview of the complete pipeline together with an evaluation of our method in comparison with SHOT [4]. For these tests, we adopt the Euclidean distance from a fixed point (assumed to be in the center of the scene) as the scalar function used in the computation of our gradient-based LRFs. This choice leads to coherent functions and thus stable gradients. We compare the x axes alone (obtained as the gradient of the selected scalar function in our case) as well as the complete LRFs of the two approaches. For the sake of visualization, we show these on a selection of points in the common regions of the two views.

At the bottom of Figure 3 we eventually compare the MeanCos score (encoding the LRF repeatability) and the matching error (encoding the descriptor repeatability); the latter is computed as the Euclidean distance between the matching point provided by each method and the ground truth correspondence.

2. Angel Dataset

In Table 1 we collect all the results on the Angel point clouds [2]. On this data, both **Ours FLARE** and **Ours STED** outperform the competitors on all the proposed metrics, with **Ours STED** providing the best results overall. Note that in the last row ('% Cor.') we are evaluating the *descriptor* repeatability; we do so by injecting our LRFs into the SHOT pipeline for computing local descriptors. For this reason, we only compare with SHOT and denote FLARE as not evaluated (n.e.).

A qualitative comparison is provided in Figure 4, where two views of the *Angel* subset are visualized. Note the large amount of dark red points visible in the first column (SHOT), indicating large deviation from the ground truth correspondence; this contrasts the large amount of white points (small to no error) in the last column (**Ours STED**). The improvement is better seen from the zoom-ins at the bottom of the figure.

	FLARE	SHOT	**Ours FLARE**	**Ours STED**
MeanCos	0.38	0.19	0.49	**0.69**
ThCos	0.12	0.06	0.12	**0.16**
%Cor.	n.e.	0.13	0.17	**0.25**

Table 1. Comparison between FLARE, SHOT and our method on the *Angel* subset. Bold numbers denote best performance.

3. Deformable shape matching

Finally, in Figure 5 we report the error rates for deformable matching on the *centaur* class from the TOSCA dataset. Similarly to the results reported for the other classes (see main manuscript), we demonstrate a big gap in performance in comparison to SHOT and the mean and Gaussian curvature baselines.

References

[1] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer Science & Business Media, 2008. 5

[2] M. Khoury, Q.-Y. Zhou, and V. Koltun. Learning compact geometric features. In *Proc. CVPR*, 2017. 1

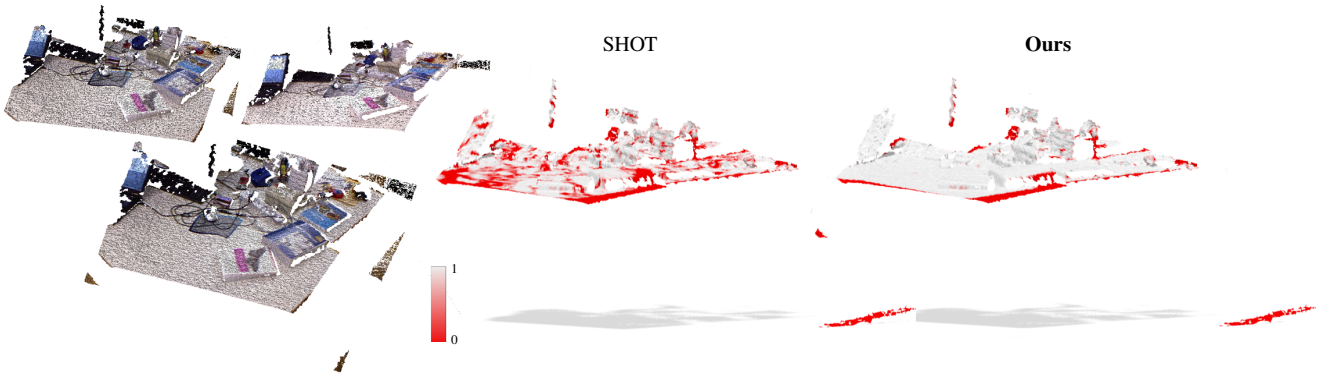

Figure 1. LRF repeatability on a pair of views of a room (depicted on the left; their ground-truth alignment is at the bottom). The MeanCos score is encoded as a heat map, growing from red (gross misalignment) to white (perfect alignment). For our method, the error is only the inevitable one localized on non-overlapping regions between the two views due to their partiality. See also Figure 2.

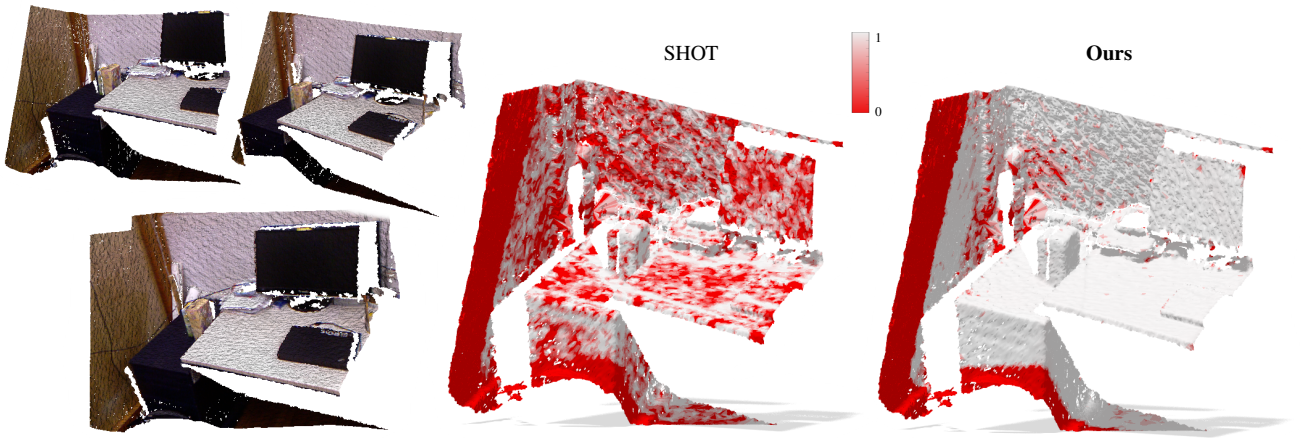

Figure 2. LRF repeatability on a pair of views from the RGB-D SLAM dataset (depicted on the left; their ground-truth alignment is on the bottom). The MeanCos score is encoded as a heat map, growing from red (gross misalignment) to white (perfect alignment). For our method, the error is only the inevitable one localized on non-overlapping regions between the two views due to their partiality. See also Figure 1.

[3] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, Oct. 2012. 1

[4] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *International Conference on Computer Vision (ICCV)*, pages 356–369, 2010. 1

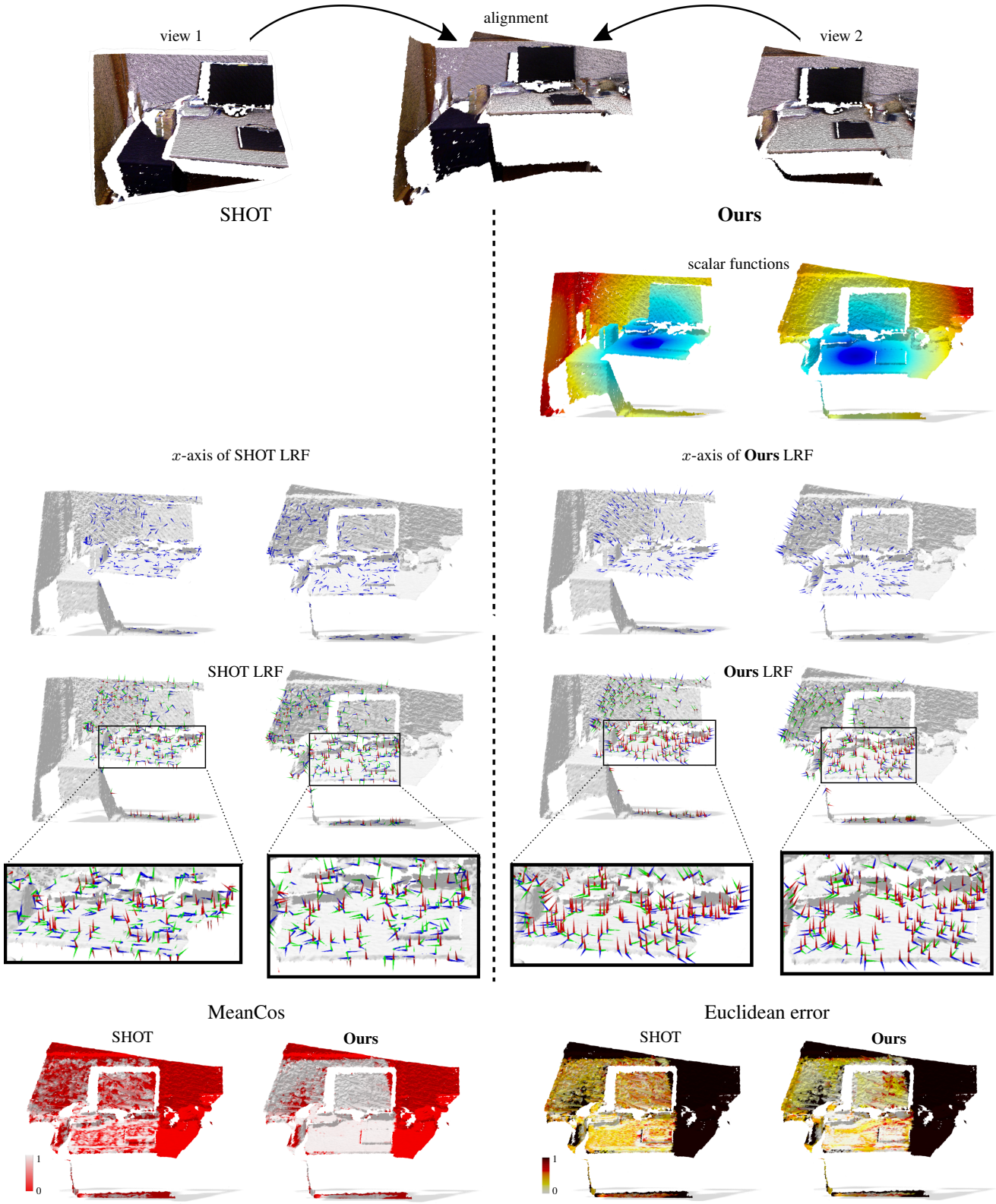

Figure 3. LRF repeatability and descriptor matching comparison on two views of the same room (depicted on the top) from the RGB-D SLAM dataset. Top to bottom: Scalar functions adopted in the computation of GFrames; the x-axis and complete LRF for SHOT (left) and for our method (right); MeanCos score (bottom left) and descriptor matching error (bottom right). The latter is encoded as a heat map growing from white (no matching error) to black (large matching error) via shades of red.

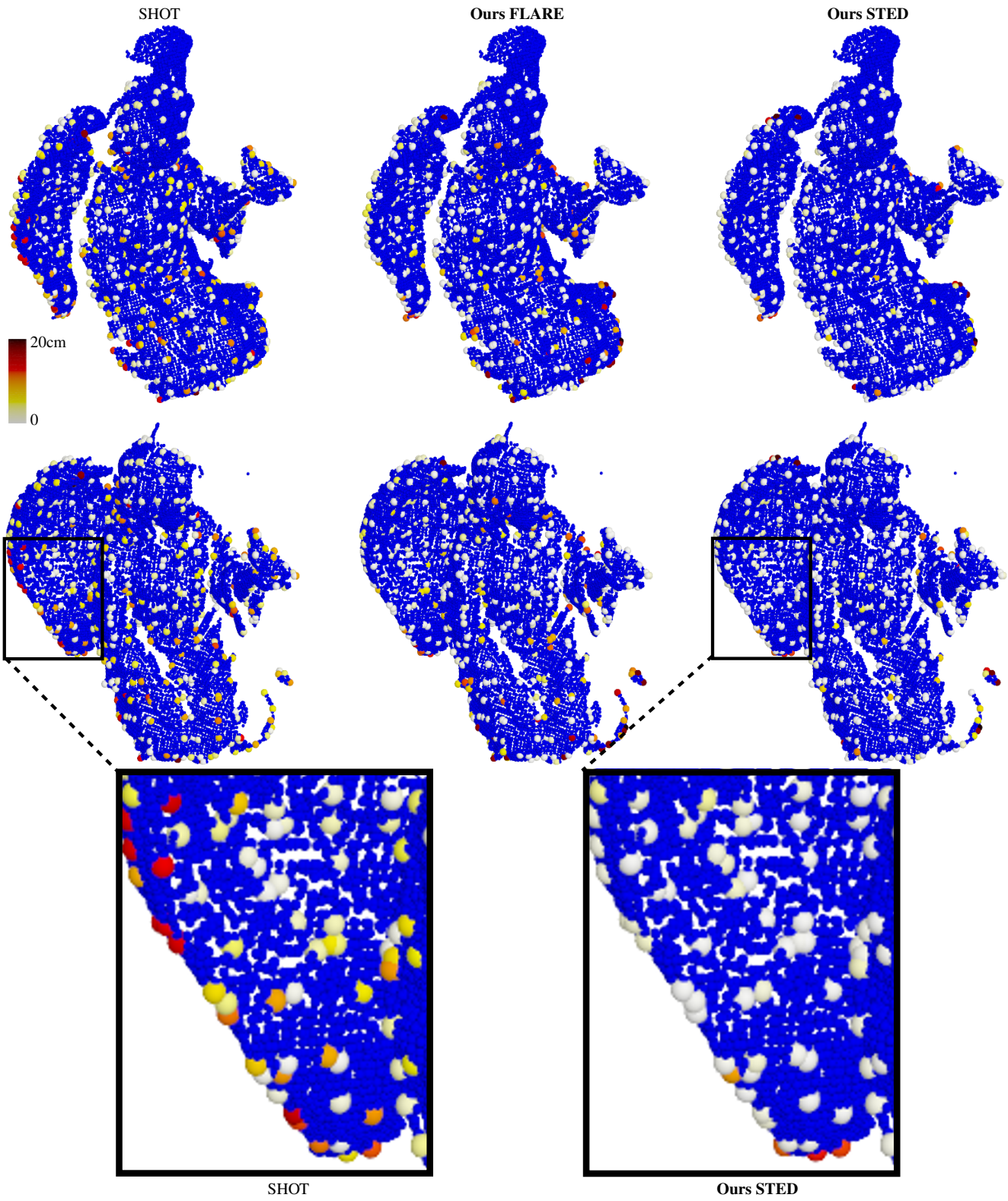

Figure 4. Qualitative comparison on a pair of point clouds from the *Angel* subset. The point-wise matching error (shown on a sparse set of points for visualization purposes) is encoded as a heatmap, growing from white to dark red. The remaining regions of the point clouds are visualized in blue. The error is computed as the Euclidean distance between the matched points and the ground truth correspondence.

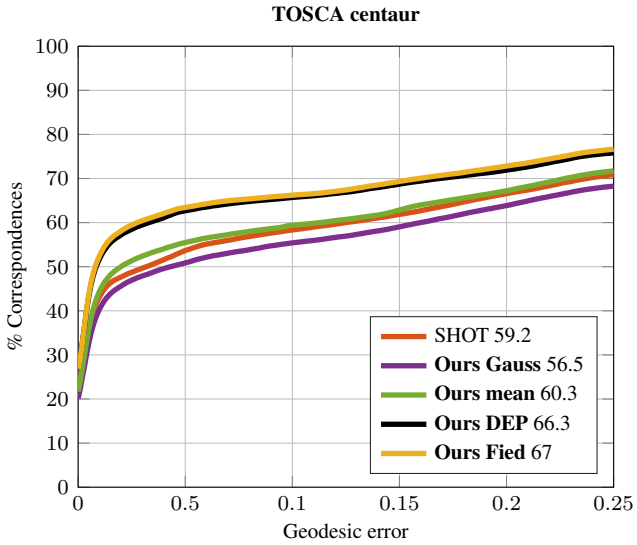

Figure 5. Correspondence results on the deformable centaur class from TOSCA [1].