Supplemental Material of Learning and Matching Multi-View Descriptors for Registration of Point Clouds

Lei Zhou[1], Siyu Zhu[1], Zixin Luo[1], Tianwei Shen[1],
Runze Zhang[1], Mingmin Zhen[1], Tian Fang[2], Long Quan[1]

[1] Hong Kong University of Science and Technology,
{lzhouai,szhu,zluoag,tshenaa,rzhangaj,mzhen,quan}@cse.ust.hk
[2] Shenzhen Zhuke Innovation Technology (Altizure),
fangtian@altizure.com

In the supplemental material, we first provide a more comprehensive introduction to the SfM training database and the full MVDesc network in Section 1. Second, we present the proof of the convergence's condition of the proposed RMBP and analysis about its parameter settings and potential break-down cases in Section 2. Third, in Section 3, we show extended discussion about the evaluation on EPFL benchmark [8] (in Section 5.3 of the paper body) and a more challenging registration experiment.

1 Supplements to MVDesc

In this section, we present a more comprehensive introduction to the collected SfM database used for training and the full MVDesc network.

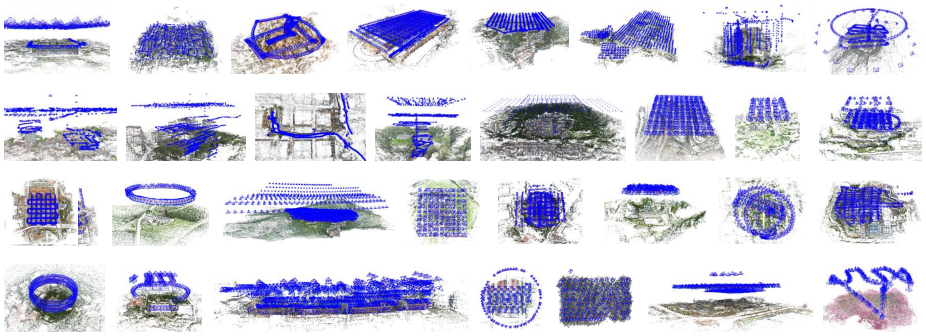


Fig. 1: The collected SfM database of 31 scenes. The blue pyramids denote the cameras and the colored points represent tracks. Up to 10 million positive and negative pairs of triple-view patches are extracted for learning MVDesc based on the SfM database

1.1 The SfM database

In Fig. 1, we visualize the collected SfM database of 31 scenes. In total, it comprises more than 27k high-resolutional images and 9M tracks with at least 6 projections. Up to 10 million positive and negative pairs of triple-view patches are extracted for training the MVDesc network proposed in Section 3.2 of the paper body.

1.2 The MVDesc Network

The MVDesc network proposed in Section 3.2 of the paper body is shown in Fig. 2 in full. It borrows the feature network from MatchNet [2] and produces 128-dimensional MVDesc descriptors. The parameters in brackets of convolutional layers denote kernel sizes, output channel numbers and strides. And the parameters in brackets of pooling layers denote kernel sizes and strides. Each convolutional layer is followed by ReLU except the last one Conv7. Layers with the same name share the same parameters.

2 Supplements to RMBP

In this section, we present the the proof of the convergence's condition, the guideline on the parameter settings and the analysis about the potential breakdown cases of RMBP.

2.1 Proof of the convergence's condition of RMBP

Here, we provide the proof for the derivation of the convergence's condition of the RMBP proposed in Section 4 of the paper body.

Considering the graphical model proposed in Section 4.1 of the paper body with pairwise interactions, the probabilistic distribution of the full set of random variables factors into

$$p(\mathcal{X}) = \prod_{(x_i,x_j)\in\mathcal{E}} \psi_{ij}(x_i, x_j) \prod_{x_i\in\mathcal{X}} \psi_i(x_i), \tag{1}$$

where $\mathcal{X}$ and $\mathcal{E}$ are the node set and edge set of the defined graph respectively. $\psi_i(x_i)$ is the potential function defined independently on node x_i and its observation node, whereas $\psi_{ij}(x_i, x_j)$ is defined on pairwise neighboring nodes x_i and x_j. In our case, the definition of pairwise potential function assumes two forms on compatible or incompatible neighbors respectively. We refer to the two forms as $\psi^+(,)$ and $\psi^-(,)$. They are explicitly expressed in the compatibility matrices given in Eq. 7 of the paper body, where the values of elements reflect the joint probability of neighboring binary variables, *i.e.*,

$$\mathbf{F}^+ = \begin{bmatrix} \psi^+(0,0) & \psi^+(0,1) \\ \psi^+(1,0) & \psi^+(1,1) \end{bmatrix}, \quad \mathbf{F}^- = \begin{bmatrix} \psi^-(0,0) & \psi^-(0,1) \\ \psi^-(1,0) & \psi^-(1,1) \end{bmatrix}. \tag{2}$$

Combining Eq. 2 and Eq. 7 of the paper body, we obtain

$$\psi^{+}(x,y)=\begin{cases}\lambda & x=1,y=1\\ 1 & otherwise\end{cases},\quad \psi^{-}(x,y)=\begin{cases}1 & x=1,y=1\\ \lambda & otherwise\end{cases} \tag{3}$$

In order to compute the marginal distribution $p(x_i)$ on each node of a cyclic network, the Loopy Belief Propagation (LBP) takes the form of a iterative message-passing algorithm between nodes as demonstrated in Eq. 6 of the paper body. As given in [9], *Simon's condition* provides an easily-verifiable and sufficient condition that guarantees convergence of LBP:

Theorem 1. *Loopy belief propagation is guaranteed to converge if*

$$\max_{x_j\in\mathcal{X}}\sum_{i\in\partial j}\log d(\psi_{ij})<1, \tag{4}$$

where $d(\psi)$ characterizes the strength of potential ψ.

In our case, the strength of the pairwise potential is defined as

$$d(\psi_{ij})^2=\max_{a,b,c,d}\frac{\psi_{ij}(a,b)}{\psi_{ij}(c,d)}. \tag{5}$$

By substituting Eq. 3, the strengths of potential functions defined on compatible or incompatible neighboring nodes both take the form:

$$d(\psi^{+})=d(\psi^{-})=\sqrt{\lambda}. \tag{6}$$

After combining Eq. 4 and 6, the Simon's Condition of RMBP's convergence is written into below format:

$$\max_{x_i\in\mathcal{X}}|\partial i|\cdot\log\lambda<2. \tag{7}$$

2.2 Parameter settings

The hyper-parameters k and l in Conditions 2, 3, 4 and 5 of the paper body determine the conditions on defining compatible and incompatible neighbors. In our implementation, we empirically set them as

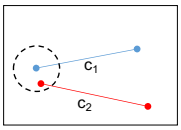
$$k=\min(5,0.01\times|\mathcal{X}|), \tag{8}$$

$$l=\max(100,0.1\times|\mathcal{X}|), \tag{9}$$

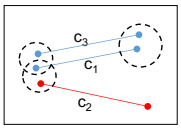
where $|\mathcal{X}|$ is the number of nodes of the graphical model. This setting can be justified from two perspectives. First, the parameter k has the maximum value of 5 and l has the minimum value of 100. It is desirable to ensure strict conditions on defining compatible and incompatible neighbors when building the graphical matching model, because we find that including an outlier neighboring relationship does more harm than excluding an inlier neighboring relationship. Second, the parameters can be adjusted adaptively according to the number of nodes. For example, a larger number of nodes lead to stricter condition on defining incompatible neighbors, while a smaller number of nodes result in stricter condition on defining compatible neighbors.

2.3 Potential break-down case of RMBP

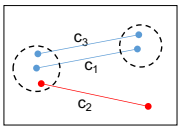
In Fig. 3 below, we demonstrate three cases with outlier match pairs, with Case1 being the break-down case and Case2 & 3 being the remedy to it. Inlier match pairs are marked in blue while outliers in red. Points in the same circle are mutually k-nearest neighbors. Let p_i be the probability of being an outlier for correspondence c_i. The analytical solutions of p_i for the three cases are written in Fig. 3(d). Values smaller than 0.5 are in blue. In Case1, both c_1 and c_2 are seen as outliers (p_1=p_2>0.5). However, in Case2 and Case3 where an inlier c_3 exists as a compatible neighbor of c_1, correct inferences are performed (p_1<0.5, p_2>0.5, p_3<0.5). The common practice of coexisting inlier neighbors like c_1 & c_3 helps to eliminate the negative effect of the break-down configuration like c_1 & c_2.



(a) Case1



(b) Case2



(c) Case3

	c_1	c_2	c_3
Case1	$\frac{2\lambda}{3\lambda+1}$	$\frac{2\lambda}{3\lambda+1}$	-
Case2	$\mathbf{\frac{4\lambda}{\lambda^2+6\lambda+1}}$	$\frac{\lambda^2+3\lambda}{\lambda^2+6\lambda+1}$	$\mathbf{\frac{3\lambda+1}{\lambda^2+6\lambda+1}}$
Case3	$\mathbf{\frac{3\lambda+1}{\lambda^2+4\lambda+3}}$	$\frac{\lambda}{\lambda+1}$	$\mathbf{\frac{3\lambda+1}{\lambda^2+4\lambda+3}}$

(d) Table of outlier probabilities

Fig. 3: (a∼c) Three of the minimal cases of neighboring matching pairs. Inlier match pairs are marked in blue while outliers in red. Points in the same circle are mutually k-nearest neighbors. (d) The analytical solutions of the outlier probability for the correspondences in the three cases. Values smaller than 0.5 are in blue

3 Supplements to Experiments

In this section, we first present further discussion on the comparative experiments on the EPFL benchmark [8] presented in Section 5.3 of the paper body. Finally, we add a challenging experiment of geometric registration of construction sites with great geometric differences.

3.1 Discussion about the Evaluation on EPFL

As a comparison to Fig. 9 of the paper body, we show the correspondences of our MVDesc-32 features in Fig. 4. Although MVDesc-32 suffers from the symmetric ambiguity like CGF [3], it still helps to find substantially sufficient number of true correspondences after putative matching and RMBP filtering for robust registration, as shown in Fig. 4. It suggests the superior description ability of the proposed MVDesc over the geometric descriptors [5, 4, 11, 10, 12, 3] in the challenging scenarios of the EPFL benchmark [8].

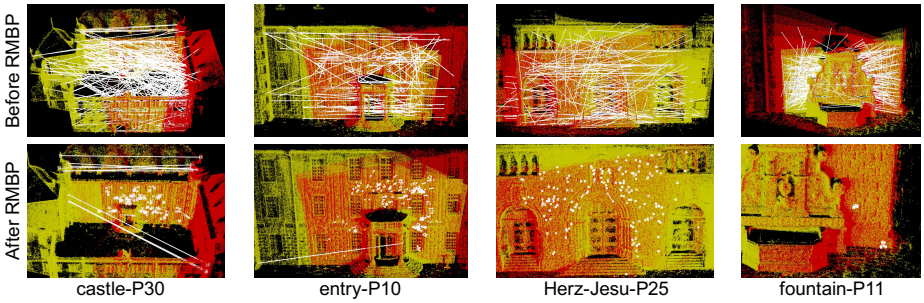


Fig. 4: The correspondences of our MVDesc-32 features as a comparison to Fig. 9 of the paper body. The white lines link between spurious corresponding points, while the white dots mark the true correspondences between the two point clouds. Although MVDesc-32 also suffers from the symmetric ambiguity like CGF [3], it still helps to find substantially sufficient number of true correspondences after RMBP

3.2 Registration of Challenging Construction Sites

Aside from the validation on the standard public datasets like ScanNet [1] and EPFL [7] in Section 5.3 of the paper body, we additionally run registration experiments on the challenging dataset of construction sites.

Setup. The dataset includes three scenes of construction sites. Each scene is comprised of two models captured by more than 300 high-resolutional images at different time and built by the standard 3D reconstruction pipeline [7, 6]. The registration of the two models is very challenging because the geometry has experienced substantial changes, which results in very limited overlap between point clouds. In Fig. 5(a), we visualize the distance from each point in one cloud to the closest point in another cloud. We sample keypoints from the point clouds randomly. And the descriptors like PFH [5], FPFH [4], SHOT [11], USC [10], 3DMatch [12], CGF-32 [3] and our MVDesc-32 are produced in the same way as the EPFL [7] experiments of Section 5.3 of the paper body. We estimate Euclidean transformations by first putative matching, then RMBP filtering and finally RANSAC.

Results. All the geometric descriptors, PFH [5], FPFH [4], SHOT [11], USC [10], 3DMatch [12] and CGF-32 [3], fail to find valid transformations due to the excessive mismatches even the RMBP filtering is applied. However, our MVDesc-32 accomplishes successful registration of all the three cases when the RMBP is equipped, as shown in Fig. 5(b).

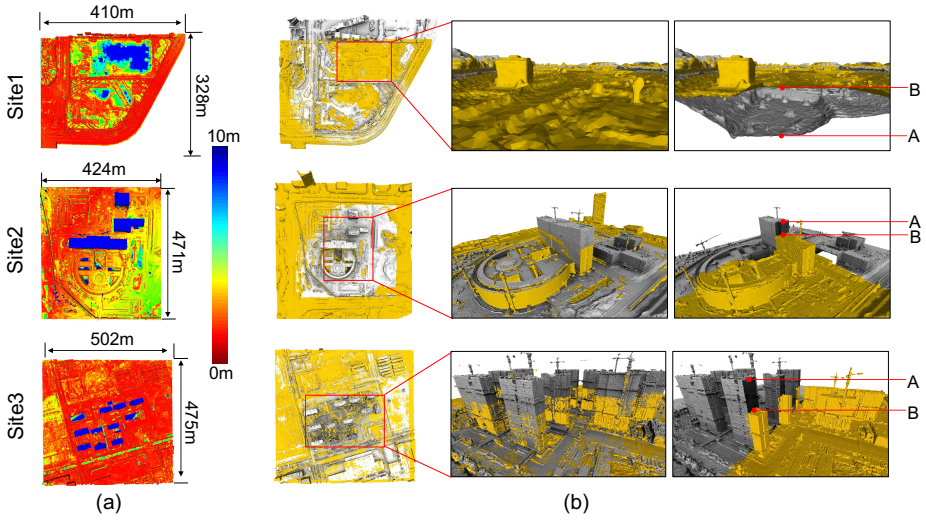


Fig. 5: (a) Visualization of the distance from each point in one point cloud to the closest point in another cloud. The warmer color represents the smaller distance. (b) The registration results of pairs of models achieved based on our MVDesc-32 plus RMBP. The last column of side-by-side visualization reveals the substantial differences in geometry between the two models

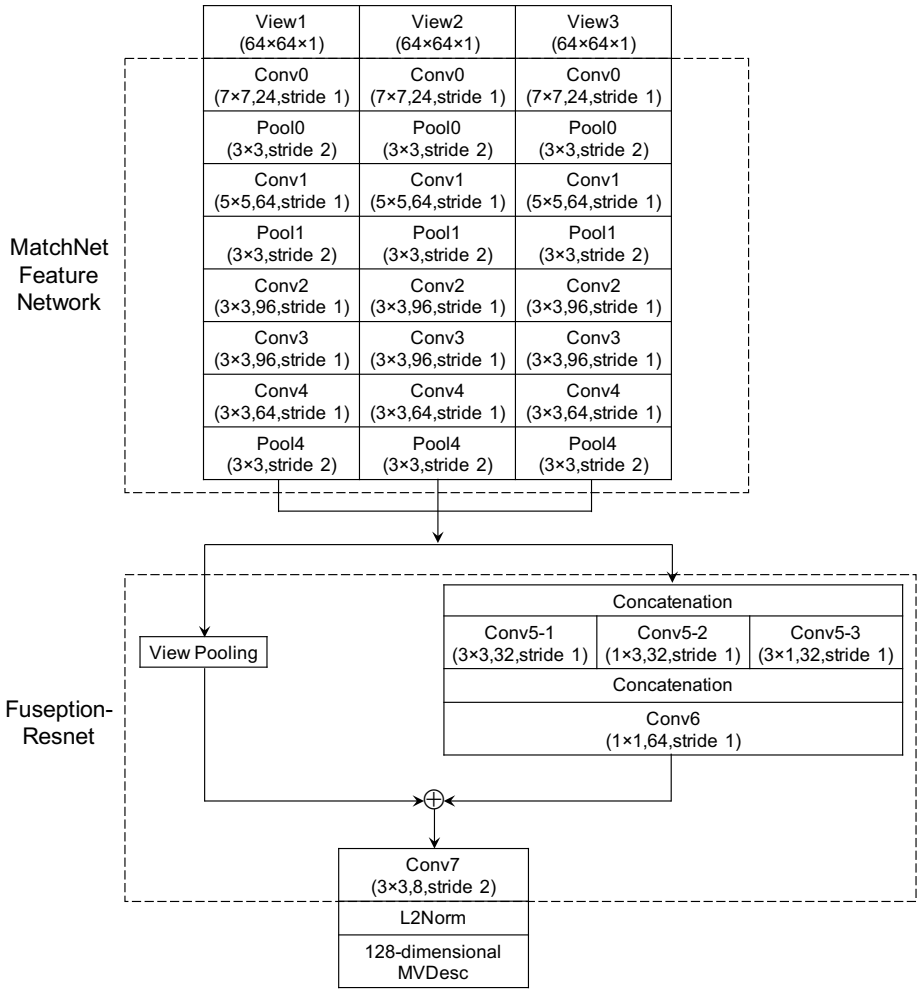


Fig. 2: The full MVDesc network that produces 128-dimensional MVDesc descriptors. The parameters in brackets of convolutional layers denote (kernel size, output channel number, stride). And the parameters in brackets of pooling layers denote (kernel size, stride). Each convolutional layer is followed by ReLU except Conv7. Layers with the same name share the same parameters

References

1. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: CVPR (2017)
2. Han, X., Leung, T., Jia, Y., Sukthankar, R., Berg, A.C.: Matchnet: Unifying feature and metric learning for patch-based matching. In: CVPR (2015)
3. Khoury, M., Zhou, Q.Y., Koltun, V.: Learning compact geometric features. In: ICCV (2017)
4. Rusu, R.B., Blodow, N., Beetz, M.: Fast point feature histograms (fpfh) for 3d registration. In: ICRA (2009)
5. Rusu, R.B., Blodow, N., Marton, Z.C., Beetz, M.: Aligning point cloud views using persistent feature histograms. In: IROS (2008)
6. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: CVPR (2016)
7. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: ECCV (2016)
8. Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR (2008)
9. Tatikonda, S.C., Jordan, M.I.: Loopy belief propagation and gibbs measures. In: UAI (2002)
10. Tombari, F., Salti, S., Di Stefano, L.: Unique shape context for 3d data description. In: Proceedings of the ACM workshop on 3D object retrieval (2010)
11. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: ECCV (2010)
12. Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T.: 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In: CVPR (2017)