Pacific Graphics (2023)
R. Chaine, Z. Deng, and M. H. Kim (Editors)

Poster

Progressive Graph Matching Network for Correspondences

Huihang Feng, Lupeng Liu and Jun Xiao [†]

School of Artificial Intelligence, University of Chinese Academy of Sciences

Abstract

This paper presents a ***progressive graph matching network*** *shorted as* ***PGMNet****. The method is more explainable and can match features from easy to hard. PGMNet contains two major blocks: sinkformers module and guided attention module. First, we use sinkformers to get the similar matrix which can be seen as an assignment matrix between two sets of feature keypoints. Matches with highest scores in both rows and columns are selected as pre-matched correspondences. These pre-matched matches can be leveraged to guide the update and matching of ambiguous features. The matching quality can be progressively improved as the the transformer blocks go deeper as visualized in Figure 1. Experiments show that our method achieves better results with typical attention-based methods.*

CCS Concepts

• ***Computing methodologies*** → ***Matching;*** *Mixed / augmented reality;*

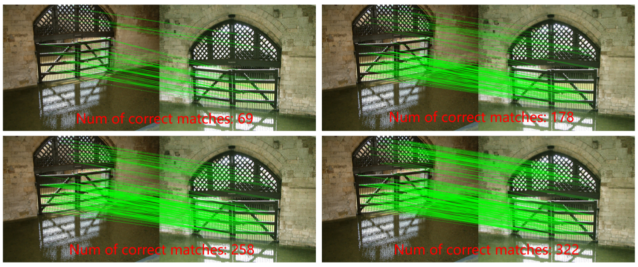
Num of correct matches: 69
Num of correct matches: 178
Num of correct matches: 258
Num of correct matches: 322

Figure 1: *The visualization of the intermediate matching results after the* 1*-th,* 3*-th,* 5*-th and* 7*-th block of PGMNet. Note that the number of matching points are progressively improved.*

1. Introduction

Feature matching is a fundamental task for 3D computer vision and augmented reality tasks such as Simultaneous Localization and Mapping, structure-from-motion, and visual localization. In classic pipelines, correspondences are obtained by nearest neighbour searching of local features and further filtered by mutual nearest neighbour and ratio test. Such methods focus only on local similarities between two sets of feature descriptors and are challenged by large viewpoints, occlusions, and texture-less regions.

The success of deep learning methods in computer vision tasks has motivated its applications in feature matching tasks and superior results have been attained. SuperGlue [SDMR19] is the seminal work and can achieve the SOTA results. It adapts the transformer architecture and optimal transport algorithm to jointly match features and filter outliers by leveraging both spatial relationships of keypoints and visual features.

However, these methods work in an end-to-end way and can only get the matching results after the last network block. When asked to match ambiguous keypoints, humans look back-and-forth at two images. We first spot deterministic and easy-to-find keypoints, and use them as structural information and guidance to match ambiguous keypoints. This basic perception scheme inspires us to develop a progressive matching framework based on sinkformers as shown in Figure 2. Sinkformers can make attention matrices doubly stochastic on both rows and columns based on sinkhorn's algorithm. The attention matrix between keypoints can be seen as matching results. By leveraging sinkformers, easy-to-spot keypoints can assist the matching of ambiguous keypoints and we can match features progressively.

2. Method and Results

In this paper, we present a progressive matching network for learning correspondences between two sets of keypoints (P_A,P_B) from image A and image B, their associated visual descriptors are denoted with (D_A,D_B). The network architecture is shown in Figure 2 and consists of an initialization layer, a seeded attentional GNN block and a matching layer.

The initialization layer is employed to integrate positional in-

[†] Corresponding Author.

DOI: 10.2312/pg.20231285

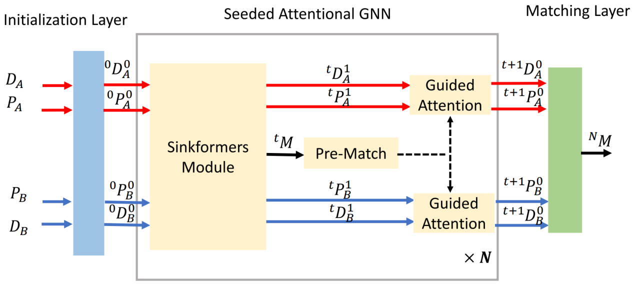

Figure 2: *The network architecture of PGMNet, which takes the descriptors and keypoint positions as input.*

Feature	*Matcher*	**AUC**			**M.S.**	**Prec.**
		@5°	*@10°*	*@20°*		
RootSIFT	NN + RT [Low04]	9.08	19.75	32.66	2.28	28.83
	AdaLAM [CLO*20]	8.24	18.57	31.01	3.10	47.59
	OANet [ZSL*19]	10.71	23.10	37.42	3.20	36.93
	SuperGlue [SDMR19]	**13.12**	**27.99**	**43.92**	8.50	42.53
	PGMNet (ours)	12.61	27.71	43.79	**10.23**	**43.57**
SuperPoint	NN+RT [Low04]	9.44	21.57	36.41	13.27	30.17
	AdaLAM [CLO*20]	6.72	15.82	27.37	13.19	44.22
	OANet [ZSL*19]	10.04	25.09	38.01	10.56	44.61
	SuperGlue [SDMR19]	13.95	29.48	46.07	15.82	44.18
	PGMNet (ours)	**14.82**	**30.11**	**46.30**	**16.27**	**45.49**

Table 1: *Comparison results tested on the Scannet dataset. The best results are highlight in bold.*

formation to descriptors. Then keypoint features are fed to several seeded attentional GNN blocks to update keypoint inter-image and intra-image to increase their distinctiveness. Afterward, the updated features are sent to the matching layer to obtain the keypoint-wise assignment matrix M between the i-th keypoint in image A and j-th keypoint in image B. The design of the initialization layer and the matching layer follows SuperGlue.

Recent works employ fully connected graph attentional networks for feature matching, where exists noise from unmatched keypoints. The intermediate matching results can be obtained by supervision to avoid the noise, we propose the Sinkformer module to predict the optimal assignment between two keypoint sets $M = \{M_{i,j} \in \{0,1\}\}$, where $M_{i,j} = 1$ means keypoint i in an image is assigned to keypoint j in the other image. To make the assignment doubly stochastic on both sets of keypoints, we have two constraints: $M^T \cdot \mathbf{1}_{|D_B|} = \mathbf{1}_{|D_A|}$ and $M \cdot \mathbf{1}_{|D_A|} = \mathbf{1}_{|D_B|}$, where $|\cdot|$ counts the number of keypoints in an image and $\mathbf{1}_{|D_A|}$ represents a vector with a dimension of $|D_A|$. This assignment problem can be solved by sinkhorn algorithm [Cut13] as follows:

$$^tM = sinkhorn\left({}^tD_I^0\left({}^tD_J^0\right)^T\right), I,J \in A,B, I \neq J. \tag{1}$$

Then the sinkformer module can be defined as:

$$^tD_I^1 = {}^tD_I^0 + MLP({}^tD_I^0 || \Delta_1), I,J \in A,B, I \neq J, \tag{2}$$

where $\Delta_1 = {}^tM(MLP({}^tD_I^0 + {}^tP_I^0))$ and $||$ means concatenation along row dimension. By the sinkhorn attention module, features in **A** retrieve and aggregate information from element in **B**. Like SuperGlue, we also use dustbin for unmatched keypoints, although we do not show it in the formula.

The guided attention model is designed to guide the feature updates of unreliable matches with those inliers. To this end, we first get the pre-matched correspondences as inlier predictions from the similar matrix ^tM based on thresholding. Then we use the pre-matched correspondences as seeded matches to guide other features. In the guided attention module, we only pass messages from pre-matched features to the other features intra-image.

The optimization objective of our method follows SuperGlue and is enforced on the assignment matrix after every processing block. Our network can be trained end-to-end with supervision from indices of ground truth matches $G_m = \{(i,j)\}$ and unmatchable points G_{uA}, G_{uB}. The total loss is:

$$L = -\sum_{k=0}^{N}\Big[\sum_{(i,j)\in G_m} log({}^kM_{i,j}) + \sum_{i\in G_{uA}} log({}^kM_{i,m+1}) + \sum_{i\in G_{uB}} log({}^kM_{n+1,j})\Big]. \tag{3}$$

We trained our network on the GL3D dataset and evaluated the results on the scannet dataset. We provide experimental results of our methods under the image matching task. We report 1) the area under curve (AUC) of pose errors; 2) the mean matching score (M.S.); 3) the mean precision (Prec.). Experimental results in Tab. 1 show that our method achieves better performance on matching score and precision compared with baseline methods. Combining with learned features like superpoint, our method achieves better performance in all metrics.

3. Conclusions

In this paper, we propose PGMNet for progressively keypoint matching. PGMNet can firstly find keypoints without any ambiguity and then further use them to guide the matching of ambigous points. Experiments show that our method achieves better performance compared with baseline methods.

4. Acknowledgement

This work is supported by the State Key Laboratory of Robotics and Systems (HIT) (SKLRS-2022-KF-11), and the Fundamental Research Funds for the Central Universities.

References

[CLO*20] Cavalli L., Larsson V., Oswald M. R., Sattler T., Pollefeys M.: Handcrafted outlier detection revisited. In Computer Vision–ECCV 2020 (2020), Springer, pp. 770–787. 2

[Cut13] Cuturi M.: Sinkhorn distances: Lightspeed computation of optimal transport. NeurIPS 26 (2013). 2

[Low04] Lowe D. G.: Distinctive image features from scale-invariant keypoints. IJCV 60 (2004), 91–110. 2

[SDMR19] Sarlin P.-E., DeTone D., Malisiewicz T., Rabinovich A.: Superglue: Learning feature matching with graph neural networks. arXiv preprint arXiv:1911.11763 (2019). 1, 2

[ZSL*19] Zhang J., Sun D., Luo Z., Yao A., Zhou L., Shen T., Chen Y., Quan L., Liao H.: Learning two-view correspondences and geometry using order-aware network. In Proceedings of CVPR (2019), pp. 5845–5854. 2