Computational Imaging Lab. (CIL)
Department of Computer Science

MULTI-SCALE KEYPOINT MATCHING

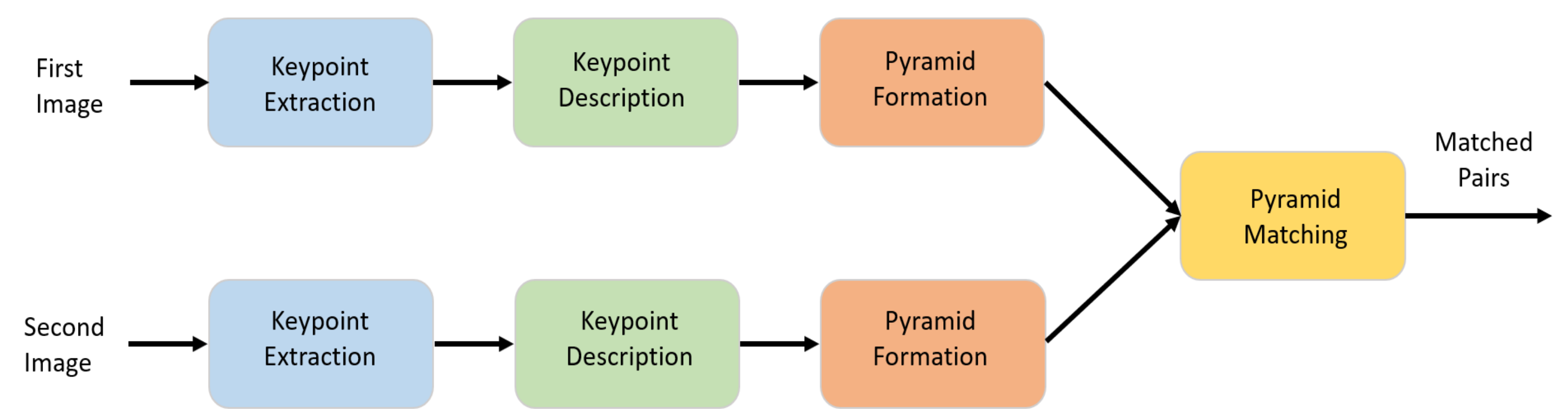
Sina Lotfian and Hassan Foroosh
University of Central Florida, Orlando, Florida, USA

ABSTRACT

We propose a new hierarchical method to match keypoints by exploiting information across multiple scales. Traditionally, for each keypoint a single scale is detected, and the matching process is done in the specific scale. We replace this approach with matching across scale-space. The holistic information from higher scales are used for early rejection of candidates that are far away in the feature space. The more localized and finer details of lower scale are then used to decide between remaining possible points. The proposed multi-scale solution is more consistent with the multi-scale processing that is present in the human visual system.
We evaluate our method on several datasets and achieve state of the art accuracy, while significantly outperforming others in extraction time.

INTRODUCTION

The problem of keypoint localization and matching is one of the most fundamental problems studied in computer vision, with broad applications. Finding point correspondence between two images can be broken into several steps. In the first step sparse set of keypoints are located on the image and scale and orientations are assigned. In the second step the neighborhood around the keypoints are mapped into the feature space based on the detected scale. However, keypoints are rarely distinguishable in their natural scales and the background information is vital for unambiguous matching. We introduce an iterative pruning process that can utilize data from both high and low scale descriptors.

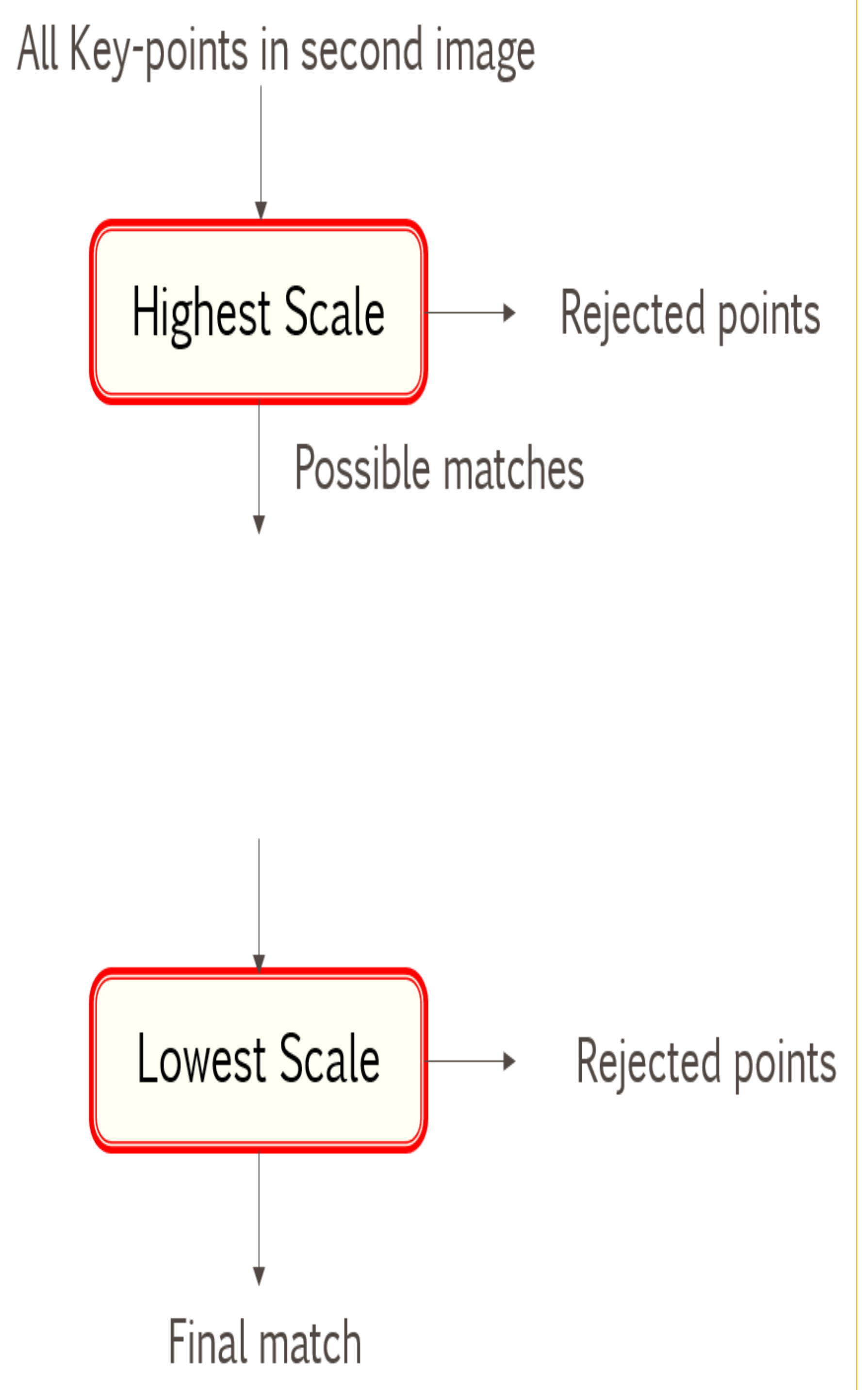


MOTIVATION

The literature on multi-scale descriptor and how to utilize information present in scale-space is relatively thin. Motivated by how humans use clues from multiple scales to find point correspondence, we introduce an iterative matching that uses the data from higher scales to early reject the mismatches.

METHOD

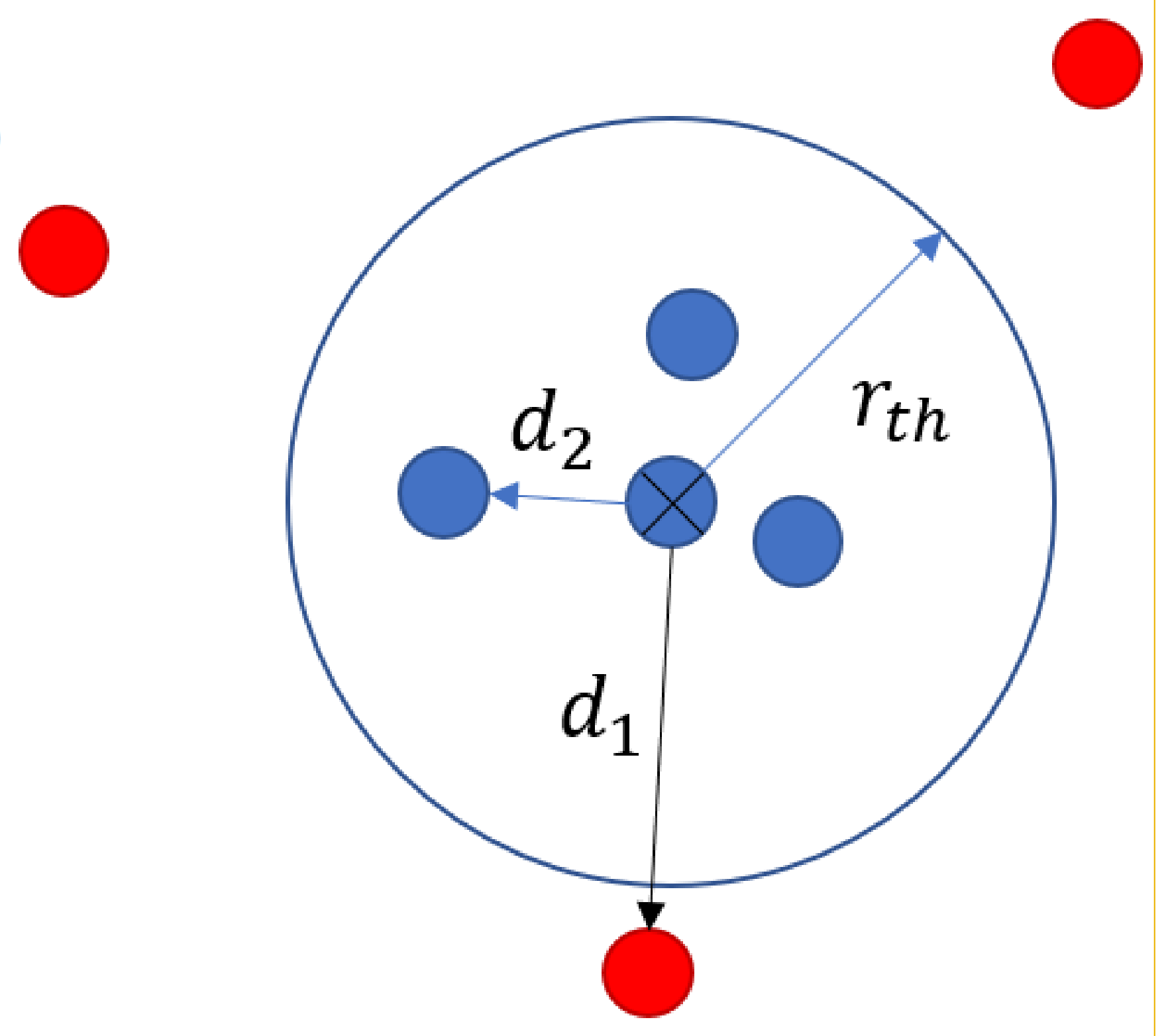
The Process starts at the highest scale. In the highest scale, keypoints are divided into two groups: rejected candidates and possible matches. This process is repeated until the final step is reached or there are only two candidates remaining.



At each scale, the candidates are classified based on their distance from the query keypoint. Let $r_{th}$and a_i be the cutoff distance and ith local feature point on the first image. Let $\{u_1, .., u_n\}$ be the set of accepted points with condition $d_{(a_i, u_j)} < r_{th}$. Let $\{v_1, .., v_n\}$ be the set of rejected points with condition $d_{(a_i, v_j)} > r_{th}$.
The cutoff threshold is then calculated by maximizing the function :

$$r_{th}(a_i) = \arg\max_{r_{th}}(min(\|\mathbf{a_i} - \mathbf{u_j})\| - max(\|\mathbf{a_i} - \mathbf{v_j}\|))$$

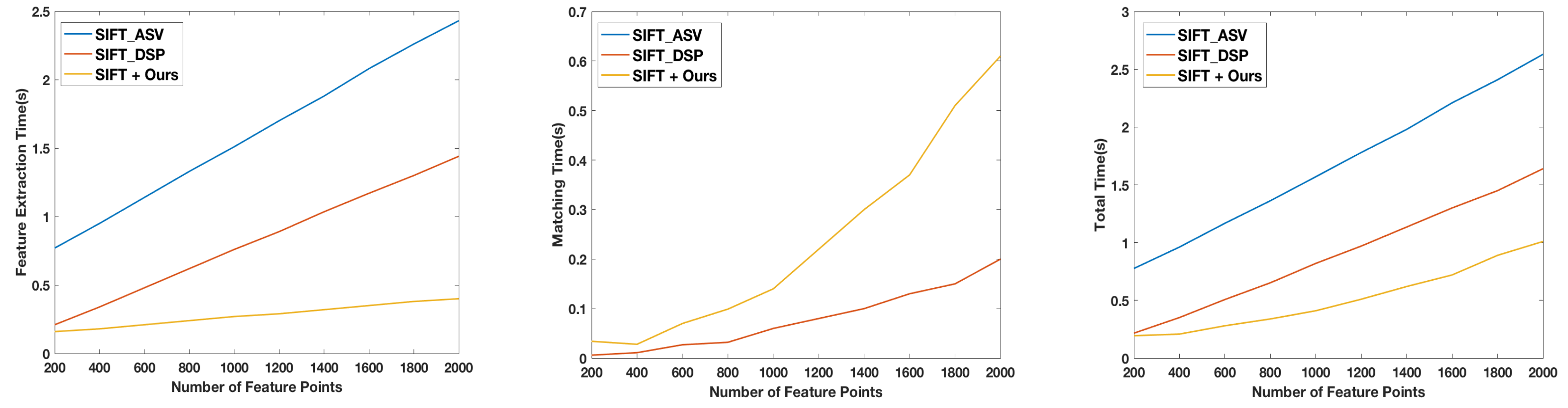
d_1 is the distance to the closest rejected point and d_2 is the distance to the furthest accepted point. The goal is to maximize the difference between the two values so that the rejected points are as far as possible while accepted points cluster closely to the query point. The margin of confidence in this case is $d_1 - d_2$.



Maximum Margin Nearest Neighbor

RESULTS

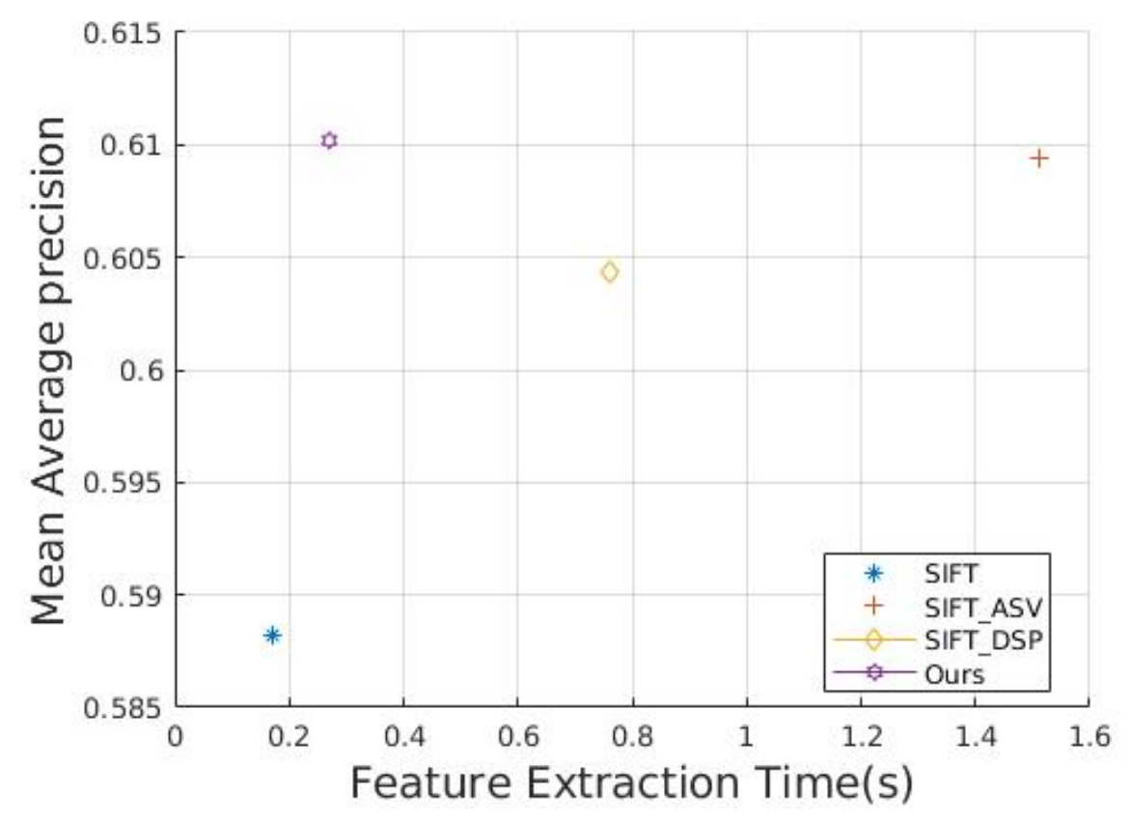
We compare our method with state of the art over multiple datasets. We demonstrate that adding the multi-scale iterative matching can increase the performance of both hand-engineered (Such as SIFT) and learned features (Such as SOS-Net).
Since the proposed method can reach maximum performance with as few as 2 scale, the feature extraction phase takes less time than the competitive methods. The proposed iterative pruning takes more time than the nearest neighbor used in competitors, however since matching is small portion of the overall time, our method is computationally more efficient for up to 2000 keypoints.



Time Analysis

Method	MAP		
	Oxford	Webcam	HPatches
SIFT [1]	58.82	16.29	42.28
SOSNet [2]	63.54	18.94	46.39
Root SIFT[3]	60.11	18.29	44.05
Raw Patch	30.67	3.97	21.13
LIOP [4]	40.54	1.79	33.87
DSP-SIFT [5]	60.43	22.28	45.17
ASV-SIFT [6]	60.94	23.16	45.53
SIFT + ours(top-bottom)	61.02	25.64	46.87
SIFT + ours(bottom-top)	60.41	25.10	47.04
SOSNet [2] + ours(top-bottom)	68.03	27.05	**51.31**
SOSNet [2] + ours(bottom-top)	**68.82**	**27.37**	50.63

Mean Average Precision for Multiple datasets



Time-mAP trade off for multi-scale approaches

DISCUSSION

In this paper we propose a computationally efficient multi-scale matching process. The multi-scale matching can increase the accuracy of both hand-crafted and learned features.

REFERENCES

[1] Lowe, David G. ”Distinctive image features from scale-invariant keypoints.” International journal of computer vision 60.2 (2004): 91-110.
[2]Tian, Yurun, et al. ”SOSNet: Second order similarity regularization for local descriptor learning.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
[3] Arandjelovi´c, Relja, and Andrew Zisserman. ”Three things everyone should know to improve object retrieval.” 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.
[4] Miksik, Ondrej, and Krystian Mikolajczyk. ”Evaluation of local detectors and descriptors for fast feature matching.” Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). IEEE, 2012.
[5] Dong, Jingming, and Stefano Soatto. ”Domain-size pooling in local descriptors: DSP-SIFT.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[6] Yang, Tsun-Yi, Yen-Yu Lin, and Yung-Yu Chuang. ”Accumulated stability voting: A robust descriptor from descriptors of multiple scales.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016