Enhancing Computer Vision: A Comprehensive Analysis and Optimization of Key Point Descriptors

Frank Sossi

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Computer Science & Software Engineering

University of Washington

2023

Reading Committee:

Professor Clark Olson, Chair

Professor Min Chen

Professor Dong Si

Program Authorized to Offer Degree:

Computer Science & Software Engineering

University of Washington

Abstract

Enhancing Computer Vision:
A Comprehensive Analysis and Optimization of Key Point Descriptors

Frank Sossi

Chair of the Supervisory Committee:
Committee Chair Professor Clark Olson
Computing & Software Systems

In the area of computer vision, the pooling of keypoint descriptors plays an important role in achieving improved performance for various applications. However, a comprehensive understanding of the performance and efficiency of pooling of different descriptors has not been well explored. This study intends to bridge this knowledge gap by implementing, optimizing, and bench-marking a set of both experimental and established descriptors under different pooling configurations. The objective is to examine how the aggregation of individual keypoint attributes into pooled descriptors can lead to a more robust and effective descriptor representation, which would potentially enhance the accuracy and reliability of computer vision tasks. The research is structured around three main objectives: exploring the best pooling of descriptors as opposed to or in combination with stacking, implementing a range of descriptor pooling strategies, and augmenting the efficacy of pooled descriptors through careful optimization processes. Furthermore, the resulting descriptors will be bench-marked to evaluate the performance of these pooled descriptors, with a special focus on precision and recall metrics, which are important for examining the accuracy and completeness of the descriptor matching process. Through an analysis of descriptor pooling performance this study aims to provide substantial evidence contributing to the broader field of computer science, and enhancing practical methodologies within the realm of computer vision. The findings

from this research are anticipated to facilitate a deeper understanding of descriptor pooling strategies but also contribute to advancements in computer vision applications, making a notable contribution to both academic and practical realms.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

Chapter 1
INTRODUCTION

1.1 Goals/Vision

The overall goal of this project is to bridge the existing knowledge gap in the realm of computer vision, particularly focusing on the pooling of keypoint descriptors to enhance the performance of descriptor representations in image recognition tasks. The following are the primary objectives and vision of this research:

- **Understanding Descriptor Pooling:** Conduct research into the implementation of descriptor pooling as opposed to or in conjunction with descriptor stacking, to generate a robust and space efficient descriptor representation. This method aims to reduce dimensionality and achieve invariance to transformations by aggregating features or descriptors over regions of an image [1].

- **Implementation and Optimization:** Implement, optimize, and benchmark a variety of established and experimental descriptors under different pooling configurations. This includes exploring innovative optimization processes to augment the efficacy of pooled descriptors, with a potential stretch goal of leveraging advancements in GPU and tensor cores for enhanced computational efficiency.

- **Performance Evaluation:** Conduct an extensive bench-marking exercise to evaluate the performance of pooled descriptors, particularly focusing on precision and recall metrics to assess the accuracy and completeness of the descriptor matching process.

- **Contribution to Computer Vision:** By investigating the hierarchical performance outcomes of descriptor pooling, this research aims to enhance practical methodologies

and applications in the field of computer vision.

- **Practical Implications:** The anticipated findings from this research could serve as a foundation for developing more advanced descriptor pooling strategies, thereby contributing to the broader field of computer vision, and potentially leading to enhanced accuracy and reliability in computer vision tasks.

The problem at hand is the lack of a thorough understanding concerning the performance and efficiency outcomes when various descriptors are pooled. The beneficiaries of this research span academia, where the findings could inform future research in computer vision, and the industry, particularly sectors reliant on image recognition technologies for various applications such as autonomous vehicles, robotics, and security systems.

1.1.1 Stakeholders and Beneficiaries

The primary stakeholders of this research include the academic community focusing on computer vision and related fields, and industries that heavily rely on image recognition technologies. The beneficiaries extend to sectors like automotive for autonomous driving, robotics for enhanced object recognition, and security systems for better surveillance and monitoring.

1.2 Success Criteria

Success for this project is defined across three tiers: minimum, expected, and aspirational criteria, enabling a clear framework for evaluating progress and outcomes.

Minimum Criteria:

- Successful implementation and documentation of at least two descriptor pooling strategies.

- Demonstration of basic functionality and potential advantages over traditional methods.

Expected Criteria:

- Implementation of a wider range of pooling strategies and their integration with existing descriptors.
- Quantifiable improvement in key performance metrics over baseline descriptors.

Aspirational Criteria:

- Development of novel pooling strategies that set new standards for descriptor performance.
- Significant contribution to the field of computer vision, recognized through publication or widespread adoption.

Quality is measured by the accuracy, efficiency, and robustness of the pooled descriptors in various computer vision tasks.

1.3 Methodology

The project methodology is designed to systematically address the research objectives through the following steps:

1. **Literature Review and Descriptor Selection:** Identifying promising descriptors and pooling strategies from current literature.
2. **Implementation:** Developing software implementations for selected descriptors and pooling strategies.
3. **Optimization:** Applying optimization techniques to enhance performance and efficiency of the implemented methods.
4. **Benchmarking:** Evaluating the performance of pooled descriptors against established benchmarks using precision, recall, and computational efficiency as key metrics.

This structured approach ensures thorough exploration and evaluation of descriptor pooling strategies, facilitating meaningful contributions to the field of computer vision.

1. Average pooling
2. Max pooling
3. Domain Size pooling
4. L1 normalization and Rooting (by itself and in combination with other strategies)
5. L1 normalization after pooling
6. L1 normalization before pooling
7. Rooting (L2 normalization after pooling)
8. Rooting (L2 normalization before pooling)

Sift Descriptor Evaluation

This document presents an evaluation of different SIFT descriptors based on their performance in three key computer vision tasks: Verification, Matching, and Retrieval. The evaluation criteria include the Area Under Curve (AUC) for the Verification task, the mean Average Precision (mAP) for the Matching task, and the mAP for the Retrieval task at various query sizes.

Descriptor Naming Convention

The naming of the descriptors follows a specific pattern based on their configuration:

- **Pooling Strategy:** Indicates the pooling method used. "Avg" for Average Pooling, "Max" for Max Pooling, and "Dom" for Domain Size Pooling (summed as per [1]).

- **Normalization Stage:** The stage at which normalization is applied. "Bef" for Before Pooling, "Aft" for After Pooling. For example "Bef" means the each descriptor is normalized then pooled.

- **Rooting Stage:** Indicates when rooting is applied. "RBef" for Rooting Before Pooling, "RAft" for Rooting After Pooling. For example "RBef" means the square root of each descriptor is done then the results are pooled.

- **Norm Type:** The type of norm used. "L1" for L1 norm, "L2" for L2 norm.

SIFT Descriptor Processing

Given a set of keypoints extracted from an image, the process of generating and processing SIFT descriptors involves several steps, parameterized by the options chosen for pooling, normalization, rooting, and the norm type. Let $\mathbf{D}$ denote the matrix of descriptors extracted for all keypoints.

Pooling Strategies

- **Domain Size Pooling (Dom)**: The descriptors are scaled and summed across different scales. For scale $s \in S$, where S is the set of scales,

$$\mathbf{D}_{\text{Dom}} = \sum_{s \in S} \mathbf{D}_s. \tag{1.1}$$

- **Average Pooling (Avg)**: The descriptors are averaged across scales,

$$\mathbf{D}_{\text{Avg}} = \frac{1}{|S|} \sum_{s \in S} \mathbf{D}_s. \tag{1.2}$$

- **Max Pooling (Max)**: The maximum value is taken from descriptors across scales,

$$\mathbf{D}_{\text{Max}} = \max_{s \in S} \mathbf{D}_s. \tag{1.3}$$

Normalization and Rooting

- **Normalization**: Applied before (Bef) or after (Aft) pooling,

$$\mathbf{D}_{\text{norm}} = \frac{\mathbf{D} - \min(\mathbf{D})}{\max(\mathbf{D}) - \min(\mathbf{D})}, \tag{1.4}$$

  where $\mathbf{D}$ is the descriptor matrix post-pooling, and the norm type dictates the space in which normalization occurs.

- **Rooting**: Applied before (RBef) or after (RAft) pooling,

$$\mathbf{D}_{\text{root}} = \sqrt{\mathbf{D}}, \tag{1.5}$$

  where $\mathbf{D}$ is the descriptor matrix subject to the selected pooling strategy.

Norm Types

- **L1 Norm (L1)**: Normalizes using the L1 norm.

- **L2 Norm (L2)**: Normalizes using the L2 norm.

Each of these steps modifies the set of descriptors $\mathbf{D}$ based on the selected options, affecting the final descriptor matrix used for image matching and recognition tasks.

Verification Task Results

Descriptor	**Noise Level**	**AUC (Balanced)**	**AP (Imbalanced)**
SIFTMaxBefRBefL1	Easy, Hard, Tough	0.917, 0.862, 0.817	0.836, 0.707, 0.601
SIFTAvgAftRBefL1	Easy, Hard, Tough	0.875, 0.830, 0.797	0.771, 0.662, 0.576
SIFTAvgBefRBefL2	Easy, Hard, Tough	0.946, 0.890, 0.834	0.885, 0.750, 0.625
SIFTAvgAftRBefL2	Easy, Hard, Tough	0.919, 0.862, 0.815	0.836, 0.703, 0.595
SIFTDomAftRBefL1	Easy, Hard, Tough	0.875, 0.830, 0.797	0.771, 0.662, 0.576

Matching Task Results

Descriptor	Easy	Hard	Tough	Mean
SIFTMaxBefRBefL1	0.666	0.338	0.139	0.381
SIFTAvgAftRBefL1	0.670	0.340	0.140	0.383
SIFTAvgBefRBefL2	0.666	0.337	0.137	0.380
SIFTAvgAftRBefL2	0.668	0.339	0.139	0.382
SIFTDomAftRBefL1	0.670	0.340	0.140	0.383

Retrieval Task Results

Descriptor	100	500	1000	5000	10000	15000	20000
SIFTMaxBefRBefL1	0.921	0.865	0.842	0.782	0.757	0.742	0.732
SIFTAvgAftRBefL1	0.938	0.886	0.864	0.807	0.783	0.768	0.759
SIFTAvgBefRBefL2	0.905	0.849	0.825	0.762	0.738	0.723	0.713
SIFTAvgAftRBefL2	0.924	0.868	0.845	0.785	0.760	0.745	0.735
SIFTDomAftRBefL1	0.938	0.886	0.864	0.807	0.783	0.768	0.759

Conclusion

The analysis demonstrates that SIFTAvgAftRBefL1 and SIFTMaxBefRBefL1 consistently show high performance across different tasks, making them among the best descriptors based on the provided results.

Performance Improvement Over Baseline SIFT

The baseline SIFT descriptor, referred to as SIFTEXP2, served as the benchmark for evaluating the improvements offered by the modified descriptors. The performance of SIFTEXP2 across the Verification, Matching, and Retrieval tasks is summarized as follows:

Baseline SIFT Performance

- **Verification Task (Balanced AUC):** Easy - 0.956, Hard - 0.887, Tough - 0.812
- **Verification Task (Imbalanced AP):** Easy - 0.911, Hard - 0.762, Tough - 0.620
- **Matching Task (mAP):** Mean - 0.353
- **Retrieval Task (mAP for 10K queries):** Mean across noise levels - 0.645, 0.521, 0.479 (for 100, 500, 1000 queries respectively)

Improvement Analysis

Comparing the enhanced descriptors with the baseline SIFT, we observe notable improvements in certain aspects. For instance, SIFTAvgAftRBefL1 and SIFTMaxBefRBefL1, among others, demonstrated superior performance across various tasks. Specifically:

- In the **Verification Task**, the enhanced descriptors showed robustness across different noise levels, with AUC values close to or surpassing the baseline in certain configurations.
- The **Matching Task** saw modest improvements in the mean Average Precision, indicating enhanced matching capabilities over the baseline.
- For the **Retrieval Task**, the enhanced descriptors outperformed the baseline, especially in scenarios with a larger number of queries, showcasing their effectiveness in retrieval applications.

The improvements highlight the effectiveness of modifying pooling strategies, normalization stages, rooting stages, and norm types in enhancing the overall performance of SIFT descriptors for specific tasks.

Conclusion

The analysis illustrates that tailored modifications to the SIFT descriptor can lead to significant improvements in performance across verification, matching, and retrieval tasks. The descriptors SIFTAvgAftRBefL1 and SIFTMaxBefRBefL1, in particular, stand out as top performers, offering robust and efficient solutions for various computer vision challenges.

Chapter 2

LITERATURE REVIEW

2.1 Initial Literature Review

The investigation of the area of keypoint descriptor methodologies in computer vision has led to innovative strides, particularly in the realm of keypoint recognition and image matching. This literature review underscores these advancements, contextualizing them in light of this study's objective: to enhance the precision and efficiency of descriptor pooling strategies for optimized performance in computer vision applications.

2.1.1 Incorporation of Color Attributes in Descriptors

The integration of color attributes in descriptors has shown promise in enhancing recognition capabilities. C. F. Olson and S. Zhang [3] introduced an innovative descriptor, Histograms of Normalized Colors (HoNC), which computes normalized color histograms within each grid cell of the array used in the SIFT descriptor. Their work demonstrates the potential benefits of color attributes, which is pertinent to this study's exploration of diverse descriptor pooling strategies.

2.1.2 Exploration of Multi-Scale Descriptors

In a significant contribution, J. Lee, E. Park, and S. Yoo [2] explored the potential of multi-scale descriptors. Their advocacy for a more comprehensive approach to descriptor methodology, especially at varying scales, aligns with this research's endeavor to examine a range of pooling strategies across different descriptor configurations.

2.1.3 Advancements in Descriptor Robustness and Efficiency

X. Zhao, X. Wu, W. Chen, P. C. Chen, Q. Xu, and Z. Li [4] proposed the ALIKED method, incorporating a deformable transformation into the descriptors to enhance their robustness while reducing computational overhead. This efficiency in descriptor processing resonates with this study's objective to optimize descriptor pooling strategies without compromising on performance.

2.1.4 Synthesis and Research Trajectory

The existing literature advocates for continuous exploration in descriptor methodologies, suggesting that innovative pooling strategies could unlock higher levels of precision and efficiency in computer vision applications. This research, through its focus on enhancing descriptor pooling strategies, seeks to contribute to this evolving narrative, potentially leading to advancements in both theoretical and practical realms of computer vision.

BIBLIOGRAPHY

[1] Jingming Dong and Stefano Soatto. Domain-size pooling in local descriptors: Dsp-sift. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5097–5106, 2015.

[2] JongMin Lee, Eunhyeok Park, and Sungjoo Yoo. Multi-scale local implicit keypoint descriptor for keypoint matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6144–6153, 2023.

[3] Clark F Olson and Siqi Zhang. Keypoint recognition with histograms of normalized colors. In *2016 13th Conference on Computer and Robot Vision (CRV)*, pages 311–318. IEEE, 2016.

[4] Xiaoming Zhao, Xingming Wu, Weihai Chen, Peter CY Chen, Qingsong Xu, and Zhengguo Li. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation and Measurement*, 2023.