

©Copyright 2023

Frank Sossi

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

	Page
List of Figures	ii
Chapter 1: Introduction	1
1.1 Goals/Vision	1
1.2 Success Criteria	2
1.3 Methodology	3
Chapter 2: Literature Review	5
Bibliography	6

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 [3].
- **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.

Chapter 2

LITERATURE REVIEW

BIBLIOGRAPHY

- [1] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5173–5182, 2017.
- [2] Vassileios Balntas, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Learning local feature descriptors with triplets and shallow convolutional neural networks. In *Bmvc*, volume 1, page 3, 2016.
- [3] 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.
- [4] Editor. Hyphenation exception log. *TUGboat*, 7(3):145, 1986.
- [5] Benoit Gallet and Michael Gowanlock. Computing double precision euclidean distances using gpu tensor cores. *arXiv preprint arXiv:2209.11287*, 2022.
- [6] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The L^AT_EX Companion*. Addison-Wesley, 1994.
- [7] Donald E. Knuth. *The T_EX book*. Addison-Wesley, 1984.
- [8] Donald E. Knuth. *The T_EX Book*. Addison-Wesley Professional, 1986.
- [9] Donald E. Knuth. *T_EX: The Program*. Addison-Wesley, 1986.
- [10] Donald E. Knuth. *Computer Modern Typefaces*. Addison-Wesley, 1986.
- [11] Donald E. Knuth. *The Metafont book*. Addison-Wesley, 1986.
- [12] Leslie Lamport. *L^AT_EX: A Document Preparation System*. Addison-Wesley, 2nd edition, 1994.
- [13] 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.

- [14] Qi Li, Vojislav Kecman, and Raied Salman. A chunking method for euclidean distance matrix calculation on large dataset using multi-gpu. In *2010 Ninth International Conference on Machine Learning and Applications*, pages 208–213. Ieee, 2010.
- [15] Clark F Olson, Sam A Hoover, Jordan L Soltman, and Siqi Zhang. Complementary keypoint descriptors. In *Advances in Visual Computing: 12th International Symposium, ISVC 2016, Las Vegas, NV, USA, December 12-14, 2016, Proceedings, Part I 12*, pages 341–352. Springer, 2016.
- [16] 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.
- [17] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007.
- [18] William Shakespeare. *Hamlet*. F.S. Crofts & Co., Inc., NY, 1946. Act I, Scene 3, Lines 70-72, are apropos.
- [19] Hao Shao, Tomáš Svoboda, and Luc Van Gool. Zubud-zurich buildings database for image based recognition. *Computer Vision Lab, Swiss Federal Institute of Technology, Switzerland, Tech. Rep*, 260(20):6, 2003.
- [20] Spivak, M.D., Ph.D. *PCT_EX Manual*. Personal T_EX, Inc., CA, 1985.
- [21] Spivak, M.D., Ph.D. *The Joy of T_EX*. American Mathematical Society, RI, 1986.
- [22] Le You, Han Jiang, Jinyong Hu, C Hwa Chang, Lingxi Chen, Xintong Cui, and Mengyang Zhao. Gpu-accelerated faster mean shift with euclidean distance metrics. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 211–216. IEEE, 2022.
- [23] 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.