Monday May 15, 2023

Enhanced Vocabulary Trees for Real-Time Object Recognition in Image and Video Streams

Sugam Jaiswal1, Arsheya Raj1, and Josiah Zacharias1

1University of Washington Computing and Software Systems, Bothell WA USA

Corresponding author: Clark Olson (e-mail: cfolson@uw.edu).

ABSTRACT This research paper explores the application of k-Means clustering trees in the field of computer vision, focusing on the influential work by Nister and Stewenius (CVPR 2006) and its extensions. The primary objective is to investigate the effectiveness and efficiency of this approach for visual recognition tasks. The study delves into the hierarchical structure of k-means clustering trees and evaluates their impact on recognition accuracy and computational efficiency. The research methodology involves an in-depth analysis of the original work and subsequent literature, along with implementing a program design based on the proposed methodology. The design includes steps such as dataset installation, feature detection using the ORB detector, k-means clustering, max voting, cluster reassignment, and noise calculations.

To evaluate the performance of the approach, experiments are conducted using publicly available datasets, with future plans to explore larger datasets such as COCO. The research also considers the exploration of alternative feature detectors and clustering algorithms like minimal cut and grab cut, aiming to reduce noise and achieve higher accuracy. The findings of this research contribute to greater understanding of k-means clustering trees in computer vision and provide insights into their practical implementation. By comparing the results with the original research paper and exploring different methodologies, the study seeks to highlight the strengths and weaknesses of this approach and inspire further advancements in the field.

INDEX TERMS Computer vision, image processing, feature extraction, image edge detection, clustering algorithms and pattern clustering, hierarchical systems

1. INTRODUCTION

This In recent years, computer vision has emerged as a vital field of research, with applications spanning various domains, including image recognition, object detection, and scene understanding. The ability to extract meaningful information from visual data plays a crucial role in numerous real-world applications, such as autonomous driving, surveillance systems, and augmented reality. One of the fundamental challenges in computer vision is developing efficient and accurate algorithms that can handle the complexity and variability of visual data.

This research paper focuses on the application of k-means clustering trees in computer vision, inspired by the seminal work presented by Nister and Stewenius in their paper "*Scalable Recognition with a Vocabulary Tree*"[[1]](#endnote-2). The primary objective is to explore the effectiveness and efficiency of this approach for visual recognition tasks. By investigating the hierarchical structure of k-means clustering trees, the study aims to enhance recognition accuracy while minimizing computational complexity.

The motivation behind utilizing k-means clustering trees lies in their potential to address the limitations of traditional methods in visual recognition. Conventional approaches often struggle to efficiently handle large-scale datasets and achieve robust recognition in the presence of noise and variations. The hierarchical structure of k-means clustering trees offers a systematic framework for organizing visual data, enabling efficient search and retrieval.

The research methodology involves a comprehensive analysis of the original work by Nister and Stewenius, along with subsequent literature and implementations that have built upon their foundations. Furthermore, a program design is developed, incorporating steps such as dataset installation, feature detection using the ORB detector, k-means clustering, max voting, cluster reassignment, and noise calculation. The performance of the proposed approach is evaluated using publicly available datasets, with plans to explore larger datasets like COCO in the future.

By examining the strengths and weaknesses of the k-means clustering tree approach, this research aims to contribute to the existing body of knowledge in computer vision. Through empirical evaluation and comparison with the results presented in the original research paper, this study seeks to shed light on the practical implications and potential improvements in visual recognition tasks. The findings and insights derived from this research can pave the way for future advancements in the field, leading to more accurate and efficient computer vision systems.

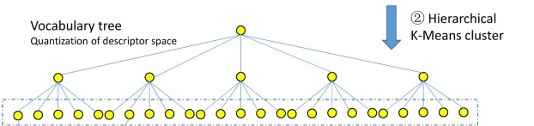
APPROACH & METHODOLOGY

1. Feature Extraction

The authors adopt the bag-of-visual-words representation, where local image features (e.g., SIFT descriptors) are extracted and quantized into visual words using clustering algorithms like k-means. This representation captures the frequency distribution of visual words in an image.

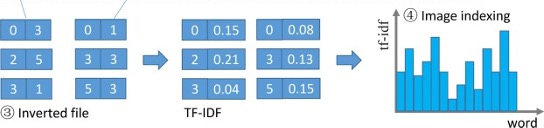
A picture containing screenshot, text, rectangle, diagram

Description automatically generated

1. Feature extraction, which we will test several.
2. k-Means Clustering Tree

The authors propose constructing a hierarchical data structure called a Vocabulary Tree. The Vocabulary Tree organizes the visual words in a tree-like structure, where each node represents a cluster of visual words. The construction process involves recursively partitioning the visual word space using clustering algorithms like k-means.

1. Creation of k-means clustering tree.
2. Image Indexing

To enable efficient retrieval, the authors introduce an inverted file indexing technique. Each leaf node of the Vocabulary Tree corresponds to an inverted file entry that stores the indices of images containing the visual word associated with that node.

1. Inverted image indexing.

PRELIMINARY RESULTS & DISCUSSION

**Image Retrieval:** Given a query image, the Vocabulary Tree is traversed by assigning the query's visual words to the appropriate nodes in the tree. The traversal path forms a histogram, and image retrieval is performed by searching the inverted file entries corresponding to the histogram bins. The retrieved images are ranked based on their similarity to the query.

**Scalability and Efficiency:** The authors highlight the scalability and efficiency of their method. The Vocabulary Tree allows for a compact representation of visual words, reducing memory requirements. Additionally, the inverted file indexing and hierarchical structure enable efficient image retrieval.

The paper provides experimental results demonstrating the effectiveness and efficiency of the Vocabulary Tree approach on large-scale image datasets, showing improved performance compared to traditional bag-of-visual-words methods.

Overall, "Scalable Recognition with a Vocabulary Tree" presents a method that addresses the challenges of large-scale image recognition by introducing the Vocabulary Tree data structure and efficient indexing techniques, offering an effective solution for image retrieval tasks.

CONCLUSION

Conclusion to come with final paper.

APPENDIX A

REFERENCES

More to come

ACKNOWLEDGMENT

More to come

FOOTNOTES

1. 1. Nister, D., & Stewenius, H. (2006). *Scalable Recognition with a Vocabulary Tree*. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) (Vol. 2, pp. 2161-2168).IEEE.[https://inst.eecs.berkeley.edu//~cs294-6/fa06/papers/nister\_stewenius\_cvpr2006.pdf](https://inst.eecs.berkeley.edu/~cs294-6/fa06/papers/nister_stewenius_cvpr2006.pdf)

   [↑](#endnote-ref-2)