FashionFinder: Efficient Clothing Retrieval with Deep Features and Vocabulary Trees

Om Mali, Naga Kushal Ageeru
Khoury College of Computer Sciences
Northeastern University
Boston, MA 02120

1 Overview

This project aims to develop an efficient clothing image retrieval system that can quickly identify and return visually similar clothing items from a large database. The input to the system is a clothing image, and the core focus is on experimenting with different feature extraction methods to find the most effective approach for capturing the unique characteristics of fashion items. Traditional techniques like Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT) will be tested alongside other deep CNN-based feature descriptors. These feature extraction techniques will capture various aspects of clothing images, including shape, texture, and color patterns, which are crucial for distinguishing different styles and designs.

Once the features are extracted, the project employs a vocabulary tree model to organize and index the clothing images for fast retrieval. The vocabulary tree uses hierarchical clustering to speed up the process of finding the top-k similar images by efficiently searching through the feature space. By combining the chosen feature descriptors with the vocabulary tree, the system aims to balance retrieval accuracy with computational efficiency.

The project will evaluate the effectiveness of different feature extraction methods in conjunction with vocabulary tree-based retrieval, comparing their performance in terms of speed, precision, and scalability. This exploration is particularly relevant for fashion applications, where quick and accurate search capabilities are essential for recommendation systems, online shopping platforms, and trend analysis.

2 Literature Survey

The project draws upon insights from several pivotal papers in the field of computer vision.

Nister and Stewenius [2006] introduces an efficient method for large-scale object recognition using hierarchical vocabulary trees. Their approach employs hierarchical k-means clustering on image features, enabling rapid and accurate image retrieval—a fundamental component of this project.

Ojala et al. [1994] presents Local Binary Patterns (LBP) as a robust texture descriptor. LBP enhances texture classification by comparing pixel neighborhoods and encoding local texture information, demonstrating significant performance improvements over traditional methods.

Additionally, Vasileva et al. [2018] proposes a method for creating type-aware embeddings that capture the compatibility among fashion items. This approach employs a multi-branch neural network architecture to effectively handle various fashion categories, showcasing one of the many feature descriptors to be explored in this project.

Finally, the Liu et al. [2016] paper presents a comprehensive dataset enriched with annotations, including landmarks and attributes, which significantly enhance clothing recognition and retrieval tasks. It introduces the FashionNet model, which integrates deep learning with clothing attributes and

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

landmarks, improving the accuracy of feature extraction for clothing items and providing a robust foundation for the project.

3 Data sets

The training and testing data will be sourced and preprocessed from the DeepFashion dataset. This dataset offers a comprehensive range of annotated fashion images, making it ideal for robust model development and evaluation. It contains 800K images, which are annotated with 50 categories, 1,000 attributes, clothing landmarks (each image has 4 - 8 landmarks), and over 300K image pairs.

4 Activity Plan

Activities to Complete:

- Data Analysis and Preprocessing: Exploration of the dataset, followed by cleaning, and normalization to ensure quality inputs for the model.
- Experimentation with Feature Extraction Algorithms: Test and compare different algorithms such as SIFT, HOG, LBP, FashionNet, etc to identify the most effective feature descriptors.
- Building Vocabulary Trees: Construct hierarchical vocabulary trees to facilitate fast and efficient retrieval of similar images based on extracted features.
- Integrating Feature Extraction and Vocabulary Tree Models: Combine the selected feature extraction method with the vocabulary tree for end-to-end image retrieval functionality.
- Evaluation Based on Similarity Scores: Evaluate the model's performance by analyzing the similarity scores obtained during the retrieval process.
- Building an Application for the Pipeline: Develop a user-friendly application that allows real-time image input and retrieval using the optimized model.

Both students will collaborate on data analysis and preprocessing. The feature extraction algorithms will be divided, with each student responsible for experimenting with different methods. After extracting features, the team will jointly construct the vocabulary tree, experimenting with various combinations of feature descriptors. We aim to complete the implementation by the last week of November, allowing sufficient time for evaluation, building application, and the preparation of the final report.

References

Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1096–1104, 2016. doi: 10.1109/CVPR.2016.124.

D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2161–2168, 2006. doi: 10.1109/CVPR.2006.264.

T. Ojala, M. Pietikainen, and D. Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 582–585 vol.1, 1994. doi: 10.1109/ICPR.1994.576366.

Mariya I. Vasileva, Bryan A. Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David A. Forsyth. Learning type-aware embeddings for fashion compatibility. *CoRR*, abs/1803.09196, 2018. URL `http://arxiv.org/abs/1803.09196`.