Enhancing Fashion Image Retrieval Accuracy Using B+ Tree Indexing and Color Features

تحسين دقة استرجاع صور الأزياء باستخدام فهرسة شجرة B+ وخصائص الألوان

FURQAN TAHSEEN HUSSEIN OKAB فرقان تحسين حسين عكاب

جامعة البيان، كلية القانون و العلوم السياسية

Furqan.t@albayan.edu.iq الايميل

تاريخ تقديم البحث : 2025/04/29
تاريخ قبول النشر : 2026/01/21

Abstract

This research focuses on improving the accuracy of fashion image retrieval systems by integrating efficient feature extraction and indexing techniques. The proposed method employs two main algorithms: the Dominant Color Descriptor (DCD) from the MPEG-7 standard for extracting the dominant color features of images, and the B+ Tree indexing structure for organizing and retrieving image data efficiently. The Euclidean distance measure is used to calculate the similarity between feature vectors in the retrieval process.
To evaluate the performance of the proposed system, two standard metrics—Precision and Recall—were applied. Experimental results conducted on a dataset of fashion images demonstrated satisfactory performance, achieving a Precision of 65% and a Recall of 69%, indicating that the combination of color-based features and B+ Tree indexing enhances the retrieval accuracy and effectiveness in fashion image databases.

Key words: **Color Features extraction, Indexing color, Euclidean distance**

المستخلص

يركز هذا البحث على تحسين دقة أنظمة استرجاع صور الأزياء من خلال دمج تقنيات استخراج وفهرسة فعّالة. تعتمد الطريقة المقترحة على خوارزميتين رئيسيتين: واصف اللون السائد (DCD) من معيار MPEG-7 لاستخراج خصائص الألوان السائدة في الصور، وهيكل فهرسة شجرة B+ لتنظيم بيانات الصور واسترجاعها بكفاءة. يُستخدم مقياس المسافة الإقليدية لحساب التشابه بين متجهات السمات في عملية الاسترجاع.
لتقييم أداء النظام المقترح، تم تطبيق مقياسين قياسيين – الدقة والاسترجاع. أظهرت النتائج التجريبية التي أُجريت على مجموعة بيانات من صور الأزياء أداءً مُرضيًا، حيث حققت دقةً بلغت 65% ونسبة استرجاع بلغت 69%، مما يُشير إلى أن الجمع بين السمات اللونية وفهرسة شجرة B+ يُعزز دقة الاسترجاع وفعاليته في قواعد بيانات صور الأزياء.

الكلمات المفتاحية: استخراج السمات اللونية, فهرسة الألوان,المسافة الإقليدية

1. Introduction

Content-Based Fashion Image Retrieval (CBFIR) has emerged as a prominent application in e-commerce, fashion, and digital marketing. The goal of such systems is to enable users to search for visually similar images based on a query image without relying on textual keywords or annotations [1].
Fashion image retrieval relies on analyzing visual features such as color, texture, shape, and overall clothing patterns. Retrieval methods can be broadly categorized into two main approaches:

Handcrafted features in fashion image retrieval rely on manually extracting specific visual characteristics from images using well-known algorithms. Prominent examples include [2]:

- **Histogram of Oriented Gradients (HOG):** for capturing edge shapes and geometric patterns of garments.
- **Local Binary Patterns (LBP):** for representing local texture and surface patterns.
- **Color Histogram:** for analyzing the distribution of colors across clothing items.

On the other hand, **deep learning–based features** utilize **deep neural networks** to automatically learn discriminative representations from data. Popular networks in this domain include[2]:

- **VGGNet and AlexNet:** classical sequential networks providing strong feature representations for clothing differentiation.
- **ResNet (Residual Network):** deep networks with skip connections that allow high-level feature extraction even with very deep architectures.

Content-based fashion image retrieval allows users to find similar designs, specific color combinations, or particular products efficiently, and serves as a powerful tool for fashion companies and designers to analyze trends, enhance e-commerce experiences, and provide accurate recommendations to customers.

2. Related work

The field of Fashion Image Retrieval (FIR) has witnessed significant development in recent years, with approaches ranging from traditional handcrafted features to deep learning and multimodal techniques [1].

Initially, some studies focused on handcrafted feature-based methods, where visual characteristics such as color, texture, and topological structure of garments are manually extracted. For instance, Zhang et al. proposed a method combining topological features with color and texture to improve retrieval accuracy in fashion image datasets [3].

With the emergence of deep learning, researchers shifted to automatically extracted deep features, which greatly improved the system's ability to distinguish fine-grained clothing patterns. Zhu et al. introduced the Multi-Granular Alignment (MGA) method that integrates global and fine-grained features of garments, achieving superior performance in image retrieval compared to previous methods [4]. Similarly, Kuang et al. employed Graph Reasoning Networks with a similarity pyramid to connect different clothing components, enhancing precise image retrieval [5].

Additionally, multimodal retrieval approaches have been proposed, combining image and text information to enhance search. Gao et al. developed FashionBERT, which links textual descriptions and images using adaptive loss, improving image-text matching in the fashion domain [6]. Moro et al. proposed a Transformer-based multimodal model that integrates text and image representations while optimizing query efficiency [7].

Some studies addressed cross-domain retrieval, aiming to retrieve fashion images across different domains, such as user-uploaded photos versus store catalog images. Deep metric learning techniques have been applied to improve retrieval performance across domains [8][9].

To improve search efficiency and reduce retrieval time in large-scale databases, some researchers applied indexing and feature dimensionality reduction. For example, Güzel utilized clustering and Principal Component Analysis (PCA) to reduce the search space in fashion image retrieval systems [10].

Finally, recent surveys such as the work by Islam et al. provide a comprehensive overview of state-of-the-art FIR methods, highlighting current trends and future challenges in the field [11]. These studies help identify the research gap addressed by the current work, such as combining fine-grained feature extraction with efficient indexing to achieve fast and accurate retrieval in large-scale fashion image databases.

3. Method Proposed

The proposed method for fashion image retrieval consists of three stages:

Color feature extraction
Search space reduction
Similarity measurement

3.1. Dataset

The experiments were conducted using a dataset of 6,236 dress images obtained from the Kaggle platform [12]. The dataset includes a wide variety of colors and dress styles. The Dominant Color Extraction algorithm was applied to all images to extract their main color features, which were then indexed using a B+ Tree structure to improve retrieval speed. The Euclidean Distance measure was employed to compute the similarity between the query image and other images in the database.

3.2.Color feature extraction

Color feature is considered important in fashion image retrieval because color plays a crucial role in representing the visual appearance of clothing items. It is one of the most dominant and easily perceivable features for humans, making it highly effective for distinguishing between different fashion images. Moreover, color information helps in grouping visually similar clothes and significantly enhances the accuracy of retrieval results.

The dominant color extraction algorithm from MPEG-7 is considered effective because it provides a compact and perceptually meaningful representation of color information. It reduces the complexity of color features by identifying the most representative colors in an image, rather than processing all pixels. This makes it robust against background variations and illumination changes. Moreover, it aligns well with human visual perception, which focuses on dominant colors rather than minor details, making it suitable for fashion image retrieval applications.

To extract the color feature from the object (which is obtained from the image using the SRD method), the **Dominant Color Descriptor (DCD)** from MPEG-7 was used. This descriptor provides an **efficient and compact color representation** of the color distribution in the image or object. The DCD is based on a clustering algorithm called **GLA** [13].

The general form of the color descriptor for an image is defined as [14]:

$$\mathbf{DCD}(\mathbf{I}) = \{(\mathbf{c_i}، \mathbf{p_i})\}، \left(\mathbf{i} = \mathbf{1}، \mathbf{2}، \dots ، \mathbf{N}\right)$$

- (I): represents the image or the object.
- $(\mathbf{c_i})$: represents the dominant color value, extracted as a three-dimensional color vector.
- $(\mathbf{p_i})$: represents the percentage of the dominant color, with values ranging from $(0<\mathbf{p_i}<=1)$, such that the sum of all percentages equals one.

The maximum number of dominant colors that can be extracted from an image is **eight colors**.

3.3.Search space reduction

In **content-based image retrieval (CBIR) systems**, particularly in the domain of fashion, **retrieval time is a critical factor** that directly affects system efficiency and user satisfaction. Reducing the time required to retrieve relevant clothing images is important for several reasons:

1. **Improved user experience:** Fast retrieval allows users to quickly find the desired clothing items, enhancing usability and engagement, especially in online shopping and recommendation applications.

2. **Handling large datasets:** Fashion datasets often contain tens of thousands of images. Efficient retrieval reduces computational load and prevents system delays, making large-scale applications feasible.
3. **Real-time applications:** In applications such as virtual fitting rooms or mobile fashion search, low retrieval time is essential to provide instant feedback and maintain user interaction.
4. **Optimized resource usage:** Faster retrieval minimizes processing power and memory requirements, leading to more cost-effective and scalable CBIR systems.
5. **Better integration with advanced features:** Efficient retrieval allows the combination of multiple features (color, shape, texture) without significantly increasing response time, improving the accuracy of results without sacrificing speed.

In summary, **reducing retrieval time is vital** for the practical deployment of fashion CBIR systems, ensuring high performance, scalability, and an enhanced user experience.

Use of B+ Tree in Fashion Image Retrieval Systems

At this stage of the fashion image retrieval system, a **B+ Tree** is employed to enhance the speed of accessing images stored in the database. A B+ Tree is a **balanced and efficient data structure** for search, insertion, and deletion operations, making it highly suitable for managing large datasets [15].

The B+ Tree is used to organize the **feature indices extracted from images** (such as color and shape) in a way that allows the system to **quickly locate matching images and reduce the search space**. With this structure, the system can perform searches **systematically and efficiently** instead of scanning each image individually, significantly reducing the retrieval time.

In summary, the use of a B+ Tree **improves search efficiency and ensures faster image retrieval**, making it a crucial component for enhancing the performance of **content-based image retrieval (CBIR) systems** [15].

How B+ Tree Works

Basic Structure

A **B+ Tree** consists of **internal nodes** and **leaf nodes**.

- **Internal nodes** contain **keys** that guide the search but do not store actual data.
- **Leaf nodes** hold the **actual data or pointers to the data** and are linked together via a **linked list**, enabling sequential search.

Searching for a Value

- The search begins at the **root node**.
- At each internal node, the target value is compared with the stored keys to determine which **child node** to follow.
- The search continues through internal nodes until reaching the correct **leaf node** containing the value or its pointer.
- Once the leaf node is reached, the desired data can be retrieved directly.

Inserting a New Value

- Insertion starts by locating the appropriate **leaf node**.
- If the leaf node has space, the value is inserted directly.

- If the leaf node is full, it is **split into two nodes**, and the middle key is promoted to the parent node.
- If the parent node becomes full due to the split, splitting may propagate **up to the root**, increasing the tree's height if necessary.

Deleting a Value

- The target value is first located in the leaf node.
- If deleting the value causes the node to fall below the minimum number of keys, the node can **borrow a key from a neighboring node or merge with it** to maintain balance.
- The tree remains **balanced**, with all leaf nodes approximately at the same level.

Practical Features

- Search, insertion, and deletion operations in a B+ Tree have **near-constant time complexity O(log n)**, even with very large databases.
- The linked and sorted leaf nodes allow for **efficient range queries**.
- B+ Trees are highly suitable for storing **image features such as color and shape**, enabling fast retrieval of matching images.

3.4.Similarity measurement

Euclidean Distance Formula in RGB Color Space

Given two color vectors (image query and image database) [14]:

ci=(Ri,Gi,Bi)

and

cj=(Rj,Gj,Bj)

The Euclidean distance between them is:

$$d_{i,j} = \sqrt{\left(c_i^R - c_j^R\right)^2 + \left(c_i^G - c_j^G\right)^2 + \left(c_i^B - c_j^B\right)^2}$$

$d_{i,j}$**:** Euclidean distance between two images.

4. Results and Experimental

Settings:

The experimental results were computed based on the following settings:

Number of trials: The experimental results were calculated over ten trials.

Number of retrieved images: Ten images were retrieved in each trial; thus, the evaluation was conducted based on ten retrieved images.

Query images: The query images were selected from the same dataset and represented different colors.

The results of the Dominant Color Extraction algorithm showed good performance in content-based image retrieval. The extracted color features were indexed using the B+ Tree structure, which significantly improved retrieval speed. By applying the Euclidean Distance measure, images with similar dominant colors to the query image were ranked at the top of the retrieved results. Overall, the proposed approach achieved high accuracy in retrieving visually similar images while reducing search time compared to non-indexed methods.

In order to evaluate the performance of the proposed image retrieval system, two widely used metrics, Precision and Recall, were employed. These metrics measure the accuracy and completeness of the retrieved results.

Precision (P) indicates the proportion of correctly retrieved relevant images among all retrieved images, while Recall (R) represents the proportion of correctly retrieved relevant images among all relevant images in the database [14].

The formulas for Precision and Recall are defined as follows:

$$\mathbf{Precision} = \frac{\boldsymbol{Number\ of\ correctly\ retrieved\ similar\ images}}{\mathbf{Total\ number\ of\ retrieval\ images}}$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{Number\ of\ correctly\ retrieved\ similar\ images}}{\boldsymbol{Total\ number\ of\ similar\ images\ in\ the\ database}}$$

The image retrieval results showed that the system achieved a Precision of 65% and a Recall of 69%.
This indicates that approximately 65% of the retrieved images were truly relevant to the query image, while about 69% of all relevant images in the database were successfully retrieved.
These results demonstrate the system's reasonable balance between precision and recall in retrieving similar images.

References

1- Peng, T., Yin, W., Liu, J., Li, L., & Hu, X. (2025). Deconfounded Fashion Image Captioning With Transformer And Multimodal Retrieval. *Virtual Reality & Intelligent Hardware*, *7*(2), 127-138.

2- Singh, M., & Singh, M. K. (2024). Content-Based Medical Image Retrieval Using Deep Learning And Handcrafted Features In Dimensionality Reduction Framework.

3- Zhang X., Sun H., Ma J., Research On Clothing Image Retrieval Combining Topology Features With Color Texture Features, 2024.

4- Zhu J., Huang H., Deng Q., Li X., Fashion Image Retrieval With Multi-Granular Alignment, 2023.

5- Kuang Y., Et Al., Fashion Retrieval Via Graph Reasoning Networks On A Similarity Pyramid, 2019.

6- Gao D., Jin L., Chen B., Qiu M., Li P., Wei Y., Hu Y., Wang H., Fashionbert: Text And Image Matching With Adaptive Loss For Cross-Modal Retrieval, 2020.

7- Moro G., Salvatori S., Frisoni G., Efficient Text-Image Semantic Search: A Multi-Modal Vision-Language Approach For Fashion Retrieval, 2023.

8- Ning C., Di Y., Menglu L., Survey On Clothing Image Retrieval With Cross-Domain, Springer, 2022.

9- Ibrahimi S., Van Noord N., Geradts Z., Worring M., Deep Metric Learning For Cross-Domain Fashion Instance Retrieval, 2019.

10- Güzel, Efficient Image Retrieval In Fashion: Leveraging Clustering And Principal Component Analysis For Search Space Reduction, 2024.

11- Islam S. Maidul, Joardar S., Ahmed Sekh* A., A Survey On Fashion Image Retrieval, 2024.

12- Xiaochun Xu. Kaggle. [Online]. Https://Www.Kaggle.Com/Datasets/Xxc025/Fashiondataset

13- A. Yamada, M. Pickering, S. Jeannin, And L.C. Jens, "MPEG-7 Visual Part Of Experimentation Model Version 9.0-Part 3 Dominant Color," ISO/IEC JTC1/SC29/WG11/N3914, 2001.

14- Ahmed, T. A., Colour-Based Image Retrieval Algorithms Based On Compact Colour Descriptors And Dominant Colour Based Indexing Methods, Universiti Utara Malaysia, 2014.

15- Yildizer, E., Balci, A. M., Jarada, T. N., & Alhajj, R. (2012). Integrating Wavelets With Clustering And Indexing For Effective Content-Based Image Retrieval. Knowledge-Based Systems, 31, 55-66.