A Review on Enhancing Retrieval Performance of Content Based Image Retrieval System Based on Early and Late Fusion Techniques

[1] Pranav Lod [2]Prof. V. S. Mahalle
[1]Department of Computer Science and Engineering SSGMCE, Shegaon
[2]Sant Gadge Baba Amravati University, Amravati, Maharashtra

Abstract

The query image is used to get pictures from a database or image repository by Content-Based Image Retrieval (CBIR) systems. CBIR is crucial in a number of industries, including architecture, web-based search, criminal prevention, and medical diagnostics. In CBIR, there are primarily two stages: the first is feature extraction, and the second is similarity matching. The effectiveness and efficacy of CBIR may be increased in a number of ways, including segmentation, relevance feedback, expanding searches, and fusion-based techniques. Several techniques for mixing and fusing several picture descriptors have been proposed in the literature. Early and late fusion techniques make up the two main categories into which fusion strategies are commonly separated. While late fusion refers to either the combining of outputs provided by distinct retrieval methods or the combination of different rankings of similarity, early fusion is the integration of visual characteristics from several descriptors into a single vector prior to the similarity computation.

Key Words: Content-based image retrieval, Feature extraction, Fusion method.

1. Introduction

Large image databases can be searched using one of two methods. The first method, text-based image retrieval (TBIR), uses manually annotated images that have appropriate names attached to them. The end user then uses keywords to search for and find the required images. The manual annotating process takes a large amount of time, which is the first obvious drawback of this strategy. The second drawback is the amount of onus placed on the end user to create their own queries. Content-based image retrieval (CBIR), an alternative and efficient technique, offers several benefits and addresses the key issues raised by the prior technique. Due to the fast growth of multimedia data through contemporary sources including the Internet, smart devices, internet of things (IoT) devices and sources, social networks, and medical imaging sources during the past two decades, CBIR also became the primary approach and a focus of extensive study. However, single feature spaces are not always enough to produce the greatest retrieval results. The majority of the older CBIR research were based on individual or single feature descriptors retrieved from colour, texture, or form contents/descriptors. As a result, the majority of current research has incorporated many descriptors from the aforementioned instances. The results of an early survey of CBIR's vast and varied application areas were presented.

While late fusion refers to either the combining of outputs provided by distinct retrieval methods or the combination of different rankings of similarity, early fusion is the integration of visual characteristics from several descriptors into a single vector prior to the similarity computation. In this study, it is suggested that a set of colour and texture attributes be employed for both fusion procedures.

2. Review of literature

[1] Ali Ahmed et al. suggested a retrieval technique that makes use of colour and linguistic information, implements and contrasts the two popular forms of fusion, and then recommends the optimal strategy to recover related photos. The following are the primary contributions of our work:

• For extracting comparable pictures based on the fusion of visual and linguistic properties, a completely automated CBIR technique is given.When high-quality representative extracted fractures of both colour and text-based pictures are concatenated into a single vector, the suggested technique provides satisfactory retrieval outcomes.

• The system's performance retrieval results improve as the accuracy of the characteristics it uses increases. The technique also supports the concept of combining several results from related measurements. Any combination of similarity measurements that the user chooses may provide an excellent final fused precision value that is superior to the sum of the separate similarity values.

Various methods to merge and fuse different picture descriptors have been suggested in the literature throughout the years. Early and late fusion are the two categories into which fusion methods are often divided. Five categories of fusion strategies have been developed: traditional early fusion, features reweighting for early fusion, representation by multi-feature spaces for late fusion, relevance feedback approaches, and multimodal retrieval. The definition and several recent research that used these concepts are provided in the sections that follow. Multiple picture representations (such as colour, texture, and form) are integrated into a single feature vector in normal or classic early fusion. This method is clear-cut, extremely easy to understand, and often assigns an equivalent weight value scheme to any of the feature spaces. Many research have used this approach. Early fusion with feature weighting is appropriate when the retrieval process is [1].

[1] Ahmed A et al. noted that feature selection is an important pre-processing operation in data mining and machine learning processes; it has a favourable effect in terms of reducing the information repetition and high dimensionality of data. The practise of eliminating elements from the data collection that are irrelevant to the task at hand is referred to as feature selection. The choice of features is important for several reasons, including performance, computational effectiveness, feature interpretability, and simplicity. By minimising unimportant and potentially redundant features, increasing precision, reducing unpredictability and the associated computational cost, and increasing likelihood, feature selection is used to achieve the shared objective and common goal of improving accuracy of the classifier[2].

[3,4] Malebary S et al. discovered that there are now more diverse digital picture databases available. In response to customer feedback that searching for and recovering necessary photos from big collections is a highly challenging undertaking, successful and efficient retrieval techniques have been created. Numerous visual qualities have been regarded as indirect techniques of retrieving pictures from databases in the majority of content-based image retrieval (CBIR) systems[5,6]. A methodology for obtaining medical pictures based on automatically deduced visual attributes, such as colour and texture, is known as content-based medical image retrieval (CBMIR). Retrieval performance remains one of the most difficult issues in current CBMIR studies, despite the many methods and approaches that have been proposed. This is because of the well-known "semantic gap" problem, which occurs between machine-captured low-level image features and human-perceived high-level semantic concepts. There has been a lot of planned research to close this gap. The retrieval model's precision can be increased and improved using two expansion strategies presented in this paper, both of which rely on the top-ranked pictures. Our first expansion technique involves reformulating the new expand query picture based on the average values for the characteristics of the top-ranked photos, whereas our second expansion method just considers the most crucial aspects. Twelve texture characteristics and eighteen colour features were derived from two widely used medical picture datasets, Atrey PK, Hossain MA, and El Saddik[7,8], and used in the expansion procedure.

3. Research Methodology

3.1. Process flow

- The image is preprocessed using image resizing, and the features are retrieved using the image's colour and texture.
- The user can choose an image from the dataset.
- The comparable photos will then be obtained once the similarity is assessed to categorise the image.
- Performance metrics like recall and accuracy will be estimated for the categorization performance.
- Early fusion is predicated on retrieving related photos as soon as the item is discovered. After locating the item, late fusion looks through the most pertinent photos and retrieves them one at a time.

4. Early Fusion Techniques

Unimodal characteristics are initially extracted using indexing strategies that rely on early fusion. Following examination of the several unimodal streams, the retrieved characteristics are integrated into a single representation. Early fusion approaches use supervised learning to categorise semantic ideas after combining unimodal information in a multimodal representation. Due to the fact that all of the characteristics are combined from the beginning, early fusion can provide a genuinely multimodal feature representation. Early fusion also offers the benefit of just requiring one learning step. The difficulty of fusing all the traits into a single representation is this method's drawback.

5. Late Fusion Techniques

The TBIR and CBIR subsystems are fused in the late multimedia fusion method, which refers to the fusion of the TBIR and CBIR subsystems. It offers conclusions that take the shape of probabilities or ratings that have numerical similarities. Therefore, using the aggregation functions Barnard K and Forsyth D[8], the probabilities were combined or fused. Algorithms for late fusion are superior than those for early fusion. Image re-ranking is a method that obtains a group of rated items from a textual subsystem and then reorders them in accordance with the visual score (Pi). The TBIR subsystem's chosen objects are the only ones used by the CBIR subsystem to compute visual scores (Pi).

6. Dataset

We utilised the Corel-1K and GHIM-10K datasets[9,10]. The first of these two picture collections, which was recently utilised, was broken up into 10 separate categories, including Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. One hundred photos with a resolution of 256384 or 384256 pixels make up each class or category. The second dataset utilised in this study is GHIM-10K, which covers a variety of items and has a total of 10K photos (10 times more than Corel-1K). It is thought to be more difficult and diversified than Corel-1K. Each group or class has 500 photographs at a resolution of 300400 pixels or 400300 pixels. There are 10,000 images total, split into 20 groups or categories.

7. Expected Results

It was discovered that in addition to measuring the degree of agreement between several sets of rankings for the same set of retrieved items, it could also be used to assess the efficacy of various retrieval procedures. The current investigation utilised the suggested system, where the classes of images serve as judges, the average precision objects serving as the present context's judges, and the accuracy rates of the three separate approaches serving as the study's objects.

8. Conclusions

A retrieval technique based on early and late fusion methods was suggested in this work. The suggested technique is a straightforward procedure utilised to improve the retrieval performance of image retrieval based on merging feature descriptors from colour and texture spaces; moreover, it incorporated combining similarity measure values for three known distance coefficients. The automated retrieval procedure and the ease of the early and late merging processes are the suggested method's two primary advantages. It also has a respectable amount of improvement. Future research should be conducted with more precise and representative characteristics, and it may also incorporate some additional descriptor features such picture shapes.

It was discovered that this test could be used to assess the efficacy of various retrieval methods and that it could also be used to gauge the degree of agreement between several sets of rankings for the same set of recovered items. In the current investigation, the accuracy rates of the three distinct approaches were regarded objects, and the classes of images served as judges, while the average precision objects were seen as judges in the current context.

References:

[1] Ali Ahmed, Sara Mohamed ," Implementation of early and late fusion methods for content-based image retrieval ", International Journal of Advanced and Applied Sciences, 8(7) 2021, Pages: 97-105

[2] Ahmed A (2020). Implementing relevance feedback for content-based medical image retrieval. IEEE Access, 8: 79969-79976. https://doi.org/10.1109/ACCESS.2020.2990557

[3] Malebary S and Ahmed A (2019). Feature selection and the fusion-based method for enhancing the classification accuracy of SVM for breast cancer detection. International Journal of Computer Science and Network Security, 19(11): 55-60.

[4] Ahmed A and Malebary SJ (2020). Query expansion based on top-ranked images for content-based medical image retrieval. IEEE Access, 8: 194541-194550. https://doi.org/10.1109/ACCESS.2020.3033504

[5] Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, and Khalil T (2019). Content-based image retrieval and feature extraction: A comprehensive review. Mathematical Problems in Engineering, 2019: 9658350. https://doi.org/10.1155/2019/9658350

[6] Alsmadi MK (2020). Content-based image retrieval using color, shape and texture descriptors and features. Arabian Journal for Science and Engineering, 45(4): 3317-3330. https://doi.org/10.1007/s13369-020-04384-y

[7] Atrey PK, Hossain MA, El Saddik A, and Kankanhalli MS (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6): 345-379. https://doi.org/10.1007/s00530-010-0182-0

[8] Barnard K and Forsyth D (2001). Learning the semantics of words and pictures. In the 8th IEEE International Conference on Computer Vision, IEEE, Vancouver, Canada, 2: 408-415. https://doi.org/10.1109/ICCV.2001.937654

[9] Bhowmik N, González R, Gouet-Brunet V, Pedrini H, and Bloch G (2014). Efficient fusion of multidimensional descriptors for image retrieval. In the International Conference on Image Processing, IEEE, Paris, France: 5766-5770. https://doi.org/10.1109/ICIP.2014.7026166

[10] Erkut U, Bostancıoğlu F, Erten M, Özbayoğlu AM, and Solak E (2019). HSV color histogram based image retrieval with background elimination. In the 1st International Informatics and Software Engineering Conference, IEEE, Ankara, Turkey: 1-5. https://doi.org/10.1109/UBMYK48245.2019.8965513