information

Editorial

Large Scale Multimedia Management: A Comprehensive Review

Saïd Mahmoudi * and Mohammed Amin Belarbi

Computer Science Department, Faculty of Engineering, University of Mons, 7000 Mons, Belgium; belarbi.mohammed.amin@gmail.com

* Correspondence: said.mahmoudi@umons.ac.be; Tel.: +32-65-374-056

Abstract: Multimedia applications deal, in most cases, with an extremely high volume of multimedia data (2D and 3D images, sounds, videos). That is why efficient algorithms should be developed to analyze and process these large datasets. On the other hand, multimedia management is based on efficient representation of knowledge which allows efficient data processing and retrieval. The main challenge in this era is to achieve clever and quick access to these huge datasets to allow easy access to the data and in a reasonable time. In this context, large-scale image retrieval is a fundamental task. Many methods have been developed in the literature to achieve fast and efficient navigating in large databases by using the famous content-based image retrieval (CBIR) methods associated with these methods allowing a decrease in the computing time, such as dimensional reduction and hashing methods. More recently, these methods based on convolutional neural networks (CNNs) for feature extraction and image classification are widely used. In this paper, we present a comprehensive review of recent multimedia retrieval methods and algorithms applied to large datasets of 2D/3D images and videos. This editorial paper discusses the mains challenges of multimedia retrieval in a context of large databases.

Keywords: multimedia management; image retrieval; CNN; dimensional reduction; CBIR

check for updates

Citation: Mahmoudi, S.; Belarbi, M.A. Large Scale Multimedia Management: A Comprehensive Review. *Information* **2022**, *13*, 28. https://doi.org/10.3390/info13010028

Received: 21 December 2021
Accepted: 27 December 2021
Published: 10 January 2022

1. Introduction

In recent years, we have witnessed an incredible and rapid growth of multimedia content in its different forms (2D and 3D images, text, video, etc.). Therefore, there is an increasing demand for multimedia-based services. These services often need some research components allowing to retrieve similar contents. In this case, content-based retrieval approaches are applied. However, managing retrieval tasks in large-scale multimedia databases requires considerable resources in terms of processing and storage. These methods target the goal of developing highly efficient and effective ways to index, retrieve and analyze visual content from multimedia databases.

An effective content-based retrieval system allows to retrieve relevant similar images to a user's query. The query could be formulated as an image (2D or 3D) or a sketch in the case of content-based image retrieval [1,2]. The research should also be very fast. This retrieval task should provide a set of results as close as possible to those given by a human visual perception.

In general, there are two different techniques for similarity-based retrieval: text-based and visual-content-based methods. Classical text-based methods use metadata and textual information to describe the multimedia content. On the other hand, content-based methods for similarity retrieval use visual features extracted directly from raw data to describe the content. Traditional text-based methods face some well-known inconveniences such as the annotation task burden, which is very time consuming, and the subjective aspect of annotation with specific content keywords which can be different according to the annotation operator.

If we investigate the methods used for content-based images retrieval, we can notice that the main technologies used for feature extractions are those based on color, shape, texture, spatial positions, etc.

Nowadays, a lot of real applications in industrial, medical, spatial and cultural domains deal with huge databases and need real time and accurate retrieval results. Hence, the principal requirement of these retrieval systems is to be accurate, fast and scalable to large database. However, classical existing systems are not real time and are generally inaccurate when they are applied to large databases. This phenomenon is named "Curse of Dimensionality". That is why the major challenge of actual retrieval systems is to keep good accuracy and fast computing time in the context of big database management. In conclusion, classical retrieval methods need to be adapted to increase their accuracy and reduce their computing times.

One the other hand, the recent development of deep convolutional networks shows that deep CNN show impressive performance improvement in various tasks, such as classification, object detection and segmentation, etc. This is due to the ability of convolutional neural networks to generate accurate representative image features. These features could be extracted from last or intermediate layers of the network [1]. However, feature vectors generated by deep CNN present a high dimensionality (4096 elements in the most common architectures), which means that CNN features as well as classical ones needs to be compressed or adopted to be scalable for large multimedia content.

We can find in the literature different methods allowing to adapt content-based image retrieval approaches to large-scale databases. These solutions can be based on approximation or dimensionality reduction applied to the generated descriptors. In the next section, we will present the main existing approaches in this domain.

2. Large Scale Multimedia Management Method: Overview

In this section, we will briefly present and discuss some recent works in image retrieval for large scale databases. We can split the existing approaches into two main classes: (1) dimensionality reduction methods based on binary/hashing or mathematical classical dimensional reduction methods such as PCA or t-Sne and (2) approximation methods such as VA-FILE, VLAD or tree based approximation approaches.

Binary coding aims to present feature vectors by a compact binary code. For example, in [3], the authors present dual binary embedding applied lo large scale images retrieval. The system uses multiples binary features extracted from SIFT (Scale Invariant Feature Transform) feature and a multi-IDF (Inverse Document Frequency) scheme allowing the association of binary features to the inverted file. Hashing-based approaches aim to construct a set of hash functions allowing to map visual features to binary codes. In some cases, these binary codes are learned from annotated databases. The key idea of hashing is to build or learn compact binary codes that represents a compact description of visual features.

In [4], Liu et al. propose an unsupervised hashing method allowing to learn short binary code associated to images visual features. This method generates one bit for each dimension and ranks the discriminative separability of each bit according to a cost function. Only top-ranked bits are selected and grouped to compose the final binary code.

The main idea of the work proposed by Wu Song et al. in [5] is to compress high-dimensional vectors in a latent space in order to generate compact binary codes. They extract deep binary codes by using as input deep CNN features generated for the retrieval task. These binary codes are calculated by comparing each feature map to the average feature map. They also apply a late fusion process to enhance retrieval precision. In [5], Haiyan et al. propose a retrieval system based on iterative quantization hashing associated with query-adaptive re-ranking.

In [6], the authors propose a method based on labeled data. This method learns binary codes from the hidden layer representing the dominant concepts in the latent space.

One of the well-known approximation methods applied in the context of large-scale retrieval is the VA-FILE (Vector-Approximation File) approach. This method builds an index file by compressing features by using an approximation process instead of only partitioning the search space like with tree-based methods. On top of that, the VA-FILE

method uses a first stage of query processing to filter the database, so that only a subset is visited in further retrieval stages [7].

In [8], the authors propose a retrieval strategy that combines a multi-resolution BOF (Bag Of Features) approach with a non-linear, non-uniform quantization and VA-file. In another work, Sprenger et al. present an interesting comparison of the performance of some multidimensional index structures (MDIS), namely the R*-tree, the kd-tree and the VA-file. In particular, they have proven that all these approaches gain a large benefit from using main memory and parallelization on modern Multi-Core CPU architectures [9].

Principal Component Analysis (PCA) was also used in several situations. For example, in [10], the authors achieved a high compression rate of 70% when they applied PCA to SIFT and SURF features. This high rate of compression features allowed faster retrieval while maintaining the accuracy of the search.

In the last decade, and since the work of Krizhevsky et al. [11], CNN-based methods are achieving incredible success in computer vision and also in retrieval tasks. One of the main methods used in the context of large database management is Regional Maximum Activation of Convolutions (RMAC) [12]. This method aggregates several image regions by using integral pooling to generate compact features with fixed length. Normalization and PCA-whitening are applied afterwards to these features as post-processing step. Other aggregation methods are proposed in the literature: [13–16].

3. Conclusions

In this editorial, we have presented the principal challenges facing multimedia retrieval in large-scale multimedia, especially the well-known Curse of Dimensionality phenomenon, which occurs when classical retrieval approaches are not efficient if applied to large databases. For the classical descriptors, extracted from shape, color and textures features, dimensionality reduction, hashing and/or binary coding methods as well as approximation approaches were proposed in the literature. On the other hand, a lot of work has been conducted in the last decade to deal with dimensionality reduction of CNN-based features. Aggregation methods such as RMAC or MS-RMAC constitute some examples of the existing solutions in this context.

Author Contributions: Conceptualization, S.M.; methodology, S.M.; writing—original draft preparation, S.M.; writing—review and editing, M.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to express their gratitude to Olivier DEBAUCHE for accepting to edit the writing of this paper.

Conflicts of Interest: The guest editor is the main author. This editorial paper is one the papers planned by the guest editor.

References

1. Seddati, O.; Dupont, S.; Mahmoudi, S. Quadruplet networks for sketch-based image retrieval. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Bucharest, Romania, 6–9 June 2017; pp. 184–191. [CrossRef]
2. Ansary, T.F.; Vandeborre, J.P.; Mahmoudi, S.; Daoudi, M. A bayesian framework for 3D models retrieval based on characteristic views. In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2004, Thessaloniki, Greece, 9 September 2004; pp. 139–146. [CrossRef]
3. Zheng, L.; Wang, S.; Tian, Q. Coupled binary embedding for large-scale image retrieval. *IEEE Trans. Image Process.* **2014**, *23*, 3368–3380. [CrossRef] [PubMed]
4. Liu, L.; Yu, M.; Shao, L. Learning short binary codes for large-scale image retrieval. *IEEE Trans. Image Process.* **2017**, *26*, 1289–1299. [CrossRef] [PubMed]

5. Wu, S.; Oerlemans, A.; Bakker, E.M.; Lew, M.S. Deep binary codes for large scale image retrieval. *Neurocomputing* **2017**, *257*, 5–15. [CrossRef]
6. Lin, K.; Yang, H.F.; Hsiao, J.H.; Chen, C.S. Deep learning of binary hash codes for fast image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 27–35.
7. Weber, R.; Schek, H.J.; Blott, S. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In Proceedings of the VLDB 98: Proceedings of the 24rd International Conference on Very Large Data Bases, New York, NY, USA, 24–27 August 1998; Volume 98, pp. 194–205.
8. Lipin, W.; Juncheng, P. Image retrieval based on VA-file and multi-resolution bow. *Appl. Math. Inf. Sci.* **2015**, *9*, 445. [CrossRef]
9. Sprenger, S.; Schäfer, P.; Leser, U. Multidimensional range queries on modern hardware. In Proceedings of the 30th International Conference on Scientific and Statistical Database Management, Bozen-Bolzano, Italy, 9–11 July 2018; pp. 1–12. [CrossRef]
10. Belarbi, M.A.; Mahmoudi, S.; Belalem, G. PCA as dimensionality reduction for large-scale image retrieval systems. *Int. J. Ambient. Comput. Intell. (IJACI)* **2017**, *8*, 45–58. [CrossRef]
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
12. Tolias, G.; Sicre, R.; Jégou, H. Particular object retrieval with integral max-pooling of CNN activations. *arXiv* **2015**, arXiv:1511.05879.
13. Li, Y.; Xu, Y.; Wang, J.; Miao, Z.; Zhang, Y. Ms-rmac: Multiscale regional maximum activation of convolutions for image retrieval. *IEEE Signal Process. Lett.* **2017**, *24*, 609–613. [CrossRef]
14. Kalantidis, Y.; Mellina, C.; Osindero, S. Cross-dimensional weighting for aggregated deep convolutional features. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 685–701.
15. Babenko, A.; Lempitsky, V. Aggregating deep convolutional features for image retrieval. *arXiv* **2015**, arXiv:1510.07493.
16. Radenović, F.; Tolias, G.; Chum, O. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 1655–1668. [CrossRef] [PubMed]

Short Biography of Author

Saïd Mahmoudi graduated from the Computer Science Department, Faculty of Sciences, University of Oran, Algeria. He received his MS in Computer Science from the LIFL Laboratory, University of Lille1, France in 1999. He obtained his PhD in Computer Science at the University of Lille 1, France in December 2003. Between 2003 and 2005, he was an Associate Lecturer at the University of Lille 3, France. Since September 2005, he is Associate Professor at the Faculty of Engineering of the University of Mons, Belgium. His research interests include inernet of things, images processing, computer aided medical diagnosis, 2D and 3D retrieval and indexing and annotation.