Applying Segmented Images by Louvain Method into Content-Based Image Retrieval

Tuyet-Ngan Vo[1], Mickael Coustaty[2], Jean-Loup Guillaume[2], Thanh-Khoa Nguyen[1(✉)], and De Cao Tran[3]

[1] Ca Mau Community College, Ca Mau, Vietnam
{vtngan,khoant}@cmcc.edu.vn
[2] L3i Laboratory, University of La Rochelle, La Rochelle, France
{mickael.coustaty,jean-loup.guillaume}@univ-lr.fr
[3] Can Tho University, Can Tho, Vietnam
tcde@cit.ctu.edu.vn

Abstract. The amount of multimedia data has increased on personal computers and the Internet requires the essential to finding a particular image or a collection of images have enhanced of demands. It urges researchers to propose new sophisticated methods to retrieve the information one desires. In the case of, the legacy approach cannot grow up with the rapid rate of available data anymore. Therefore, content-based image retrieval (CBIR) has attracted many researchers to various fields. Content-based image retrieval models attempt to effort to automate data analysis and indexing. In this paper, we propose a content-based image retrieval system for real images. This method is using segmented images by the Louvain method [26] to create features in order to apply to the CBIR system based on the Bag-of-Visual-Words (BoVW) model. In order to evaluate the proposed method, we selected the Corel dataset which is composed of 10 classes [14] total of 1000 images in the dataset for the experiment. The experimental results are shown using qualitative and quantitative evaluations.

Keywords: Louvain method · Image content-based retrieval · Bag of Visual Words · Shape representation · Image content-based searching · Image retrieval

1 Introduction

Recently, image retrieval (IR) has become a fascinated topic in the computer vision field. Especially, with the dramatic increase in multimedia data, the classical information retrieval techniques do not meet the demand for users. Content-based image retrieval (CBIR) has become an alternate technique for information retrieval.

Image retrieval is the discipline of research that focused on finding, searching, and retrieving digital images that are stored. For traditional image retrieval systems, embedded metadata such as captioning, keywords, or descriptions of the

P. Cong Vinh and A. Rakib (Eds.): ICCASA 2021, LNICST 409, pp. 77–90, 2021.
https://doi.org/10.1007/978-3-030-93179-7_7

images is used in order to perform image retrieval through the annotation words. Recently, the amount of digital multimedia data has increased rapidly. Therefore, information retrieval techniques have shifted from text-based to become semantic-based or content-based.

Text-based image retrieval (TBIR) is a traditional image retrieval method that uses artificial notes to retrieve image results. This image retrieval method not only requires huge labors to manually annotate the images but also takes time-consuming and complicated work.

The content-based image retrieval (CBIR) means that the search analyzes the content of the image based on features related to color, texture, shape, or any other informative that can be derived from image properties themselves. This image retrieval method is desirable since most image search engines reply purely on metadata, and this process produces a lot of unrelated images.

The semantic-based image retrieval (SBIR) is an automatic method of extracting semantically meaningful representations from low-level features of images to support queries by high-level semantic concepts in the user's mind. Reducing the semantic gap between the low-level features and high-level semantic concept is a main concern of semantic image retrieval efforts.

In fact, the way of image features can distinguish and what people perceive from the image are not common. Semantic-based image retrieval is often built by using extracted low-level features methods to determine meaningful while interesting objects related to the similarity properties of the visual features. In a sense, the semantic process extracts object features to produce a semantic description of images in order to be stored in database. To grab Image, we can make a query based on high-level concepts. The query could work thanks to a set of textual words which translate into semantic features in the query. The local features is used for the semantic mapping process that will be performed through supervised or unsupervised learning tools to cooperate the low-level features. [35,36]. The Semantic content can be retrieved by textual annotation or by high sophisticated inference procedures based on visual content. Recently, deep learning has enabled detection objects belong to many classes exceed traditional computer vision techniques [28,29,34].

2 Bag of Visual Words Model

The Bag-of-Visual-Words (BoVW) model is used popularly in CBIR models. The image is represented as a collected local feature. The local features are used to represent groups of local descriptors. The local descriptors obtained by the extraction process for each image could be extremely huge. Moreover, it takes a lot of time to search in the image query to find the nearest neighbors for the local descriptors. For this reason, bag-of-visual-words were proposed as an approach to solving this problem by quantizing descriptors into "visual words" to reduce the number of descriptors significantly. Bag-of-Visual-Words can create the descriptor that becomes more robust to change. The BoVW model is more likely to the traditional description of texts in information retrieval, however, we consider for images retrieval in this case [6,16]. Commonly, Bag-of-Visual-Words reply on the Keypoint detectors and Keypoint descriptors. Besides, there are many researches using methodology with graphs [5,30,31], Strokes [17,27], Bag-of-ARSRG (attributed relational SIFT (scale-invariant feature transform) regions graph) words (BoAW) [18], and *etc.*

In our work, we attempt to use our image segmentation method in the BoVW model by extracting features in that one segmented image to build a CBIR. The CBIR based on the BoVW method comprises three main stages: (1) finding the region of interests (interesting part) of an image using Keypoint detectors in general; (2) computing a summary of the Region through the use of a feature vector (like the Keypoint descriptors); and (3) building the vocabulary to define a common subspace (the vocabulary) for all images. This subspace will then make images comparable. We present a detailed description in the following sequential subsections.

2.1 Keypoint Detection

Keypoint detection is a process that uses a feature detector to determine the unique content regions in an image, for instance, corners. Feature detection is used to search points of interest (keypoints) in the image. Keypoints are guaranteed to be no change. Therefore, the feature detector can detect them in case of the presence of scaling or rotation.

Detecting Keypoints is used for the first step in the bag-of-visual-words method. In order to extract features from interest points, these features are counted at predefined locations and scales [15]. The extraction of features is an individually separated process from represented features in the bag-of-visual-words method [20]. In the literature, there are several keypoint detectors that have been used in researches, such as Harris-Laplace, Difference of Gaussian (DoG), Hessian Laplace, and Maximally Stable Extremal Regions (MSER) [8,21,33].

2.2 Keypoint Descriptors

Based on the detected interest points, it is computed a local descriptor. Local descriptors are implemented by image processing of the transformation of a local pixel neighborhood into a compacity of the represented vector.

According to the content of Keypoints, local descriptors are described as multi-dimensional vectors. In fact, features descriptors could use to construct the representation of the neighbor pixels around a localized keypoint [20]. The literature witnesses many kinds of feature descriptors, such as the scale-invariant feature transform (SIFT) [16], speeded up robust features (SURF) [1], gradient location and orientation histogram (GLOH) [19] and histogram of oriented gradients (HOG) [7].

2.3 Building Vocabulary

The amount of extracted feature descriptors are enormous. In order to solve this issue, we need to build a visual vocabulary by using a clustering algorithm to cluster the feature descriptors, for instance, K-Means technique [10]. In the vocabulary, each cluster is represented by its respective cluster center and treated as a distinct visual word. The clustering algorithm determines the size of the vocabulary and is affected by the size and the types of the dataset [1].

The Bag-of-Visual-Words model could be formulated as below. First of all, the Bag-of-Visual-Words defines the training dataset as S including images represented by $S = s_1, s_2, \ldots, s_n$, where s is the extracted visual features. Secondly, Apply the clustering algorithm like K-Means, which is based on a fixed number to visual words W represented by $W = w_1, w_2, \ldots, w_v$, where v is the cluster number. After that, the data is summarized in a $V \times N$ occurrence table of counts $N_{ij} = n(w_i, s_j)$, where $n(w_i, s_j)$ denotes how often the word w_i is presented in an image s_j [16].

3 Our CBIR Architecture

3.1 Converting Images to Complex Networks

We can represent an image as an undirected graph $G = (V, E)$, where V is a set of vertices ($V = \{v_1, v_2, ..., v_n\}$) and E is a set of edges ($E = \{e_1, e_2, ..., e_k\}$). Every vertex $v_i \in V$ corresponds to a pixel and an edge $e_{ij} \in E$ links vertices v_i and v_j. In this research, we define paths towards other pixels that are considered when the distance of two pixels is lower or equal L neighbors pixels. We have also tested for many cases of distances of L for rows and columns directions (see Fig. 2). Empirically, the $L = 15$ value reaches a relatively good performance. The obtained networks by this pattern have few edges ($2.L$ per pixel) which offer a good condition for the Louvain algorithm to compute communities faster. It is also useful to create edges for distant pixels which can offer good criteria for the Louvain method operates [26]. Moreover, edges are weighted. Its weight w_{ij} is a value of the similarities between v_i and v_j. They are defined as:

$$w_{ij} = \begin{cases} 1 & if \quad d^c_{ij} \le t \text{ for all color channels } c \\ nil & otherwise \end{cases} \tag{1}$$

where t is a threshold, d^c_{ij} is a measure of the similarity of pixels i and j intensity for color channel c (among R, G and B). It is defined by $d^c_{ij} = \left|I^c_i - I^c_j\right|$ where I^c_i and I^c_j are the intensity value of pixel i and j.

3.2 Detection Communities in Complex Networks

Community detection is finding the best partition of a network. The community is dense of the internal link of nodes while sparse external connected with others. In the literature, it is witnessed the existence of several algorithms have been

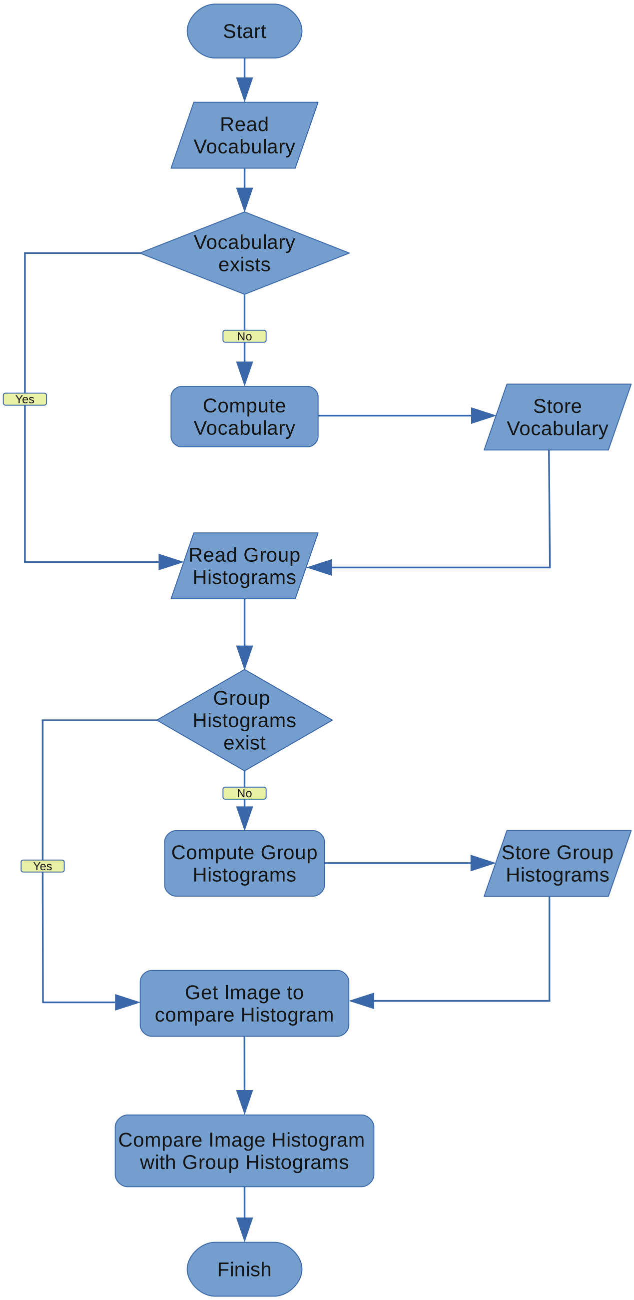

Fig. 1. A general Bag-of-Visual-Word model

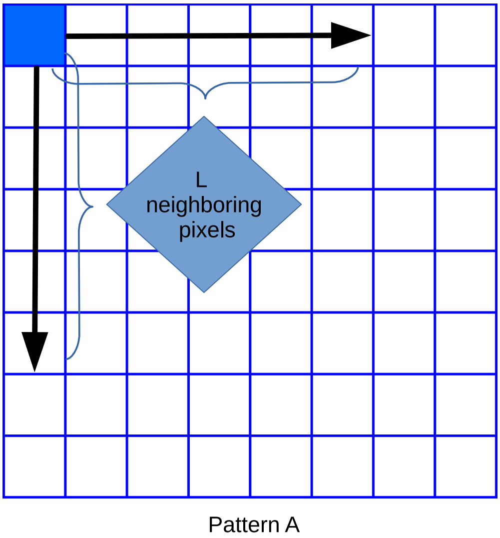

Fig. 2. A pattern for generating complex networks

proposed to find good partitions in a fast way [9]. It is known that the Newman-Girvan modularity [22] is used to measure the qualities of partition results. It is formulated as below:

$$Q = \sum_i (e_{ii} - a_i^2) \tag{2}$$

where e_{ii} denotes the fraction of edges in community i and $a_i (= \sum_j e_{ij})$ if the fraction of ends of edges in community i.

The stronger community structure of the network is indicated by the value of the modularity Q. The community detection algorithms are usually heuristics algorithms. In our research, the Louvain method [2] is used, a hierarchical greedy algorithm, to detect the communities on weighted graphs. The performance of the Louvain method is very quick and very efficient. Its properties are vital to working on graphs built from images. It is inferred that images involve many pixels so graphs are huge and to obtain well-segmented images, communities must be well-identified.

3.3 Extraction of Feature

To extract image features, we consider two aspects of image properties are color and texture. Building a 15-dimensional vector feature for each sub-segmented, namely HOGMeanSD feature [24–26], including 6 elements from Mean and Standard deviation and 9 elements come from HOG feature, detailed in [25].

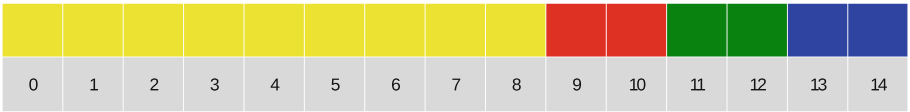

Fig. 3. A 15-dimensional vector for each segment

Based on the results we got in the image segmentation task, we build a content-based image retrieval using the bag-of-visual-words of local features on the segmented regions. The system computes the feature descriptor for those segmented regions. For the validation of our system, we decided to use the K-Means algorithm to cluster the visual vocabulary. As shown in Fig. 4, the proposed system includes two stages: a training stage and a testing stage.

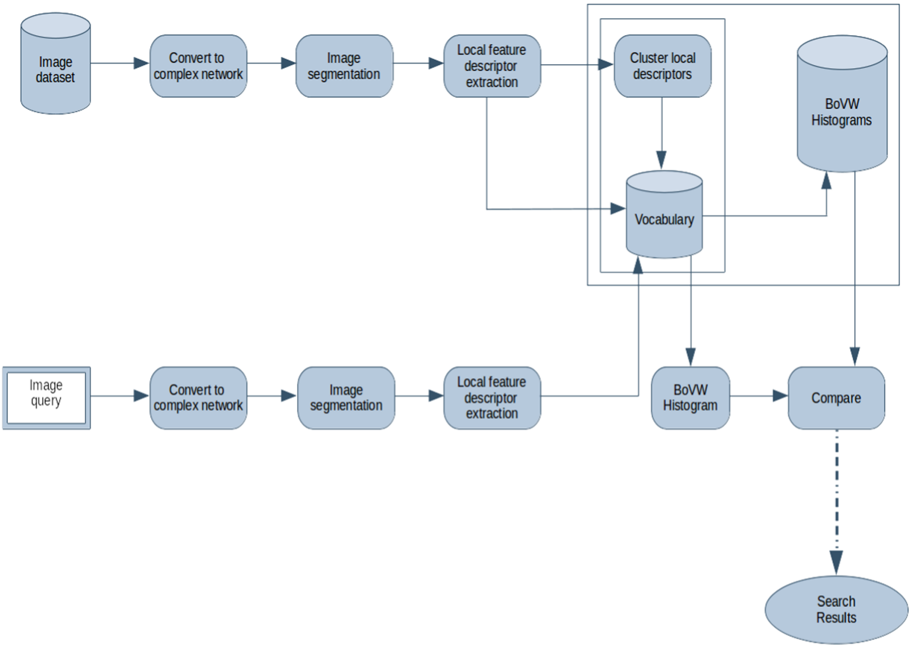

Fig. 4. An illustration of our proposed CBIR model

3.4 Training Stage

In this stage, the proposed system will implement repeatedly for all images in the dataset.

1) For every image
- Read each image individually
- Convert image to a complex network (Graph)
- Segment the image based on community detection: Louvain method and ARA algorithm [23]
- Extract features based on the sub-segments and associates these characteristics to feature descriptors
- Using a K-Means algorithm to cluster the set of local descriptors into K clusters.

2) For every feature descriptor in the image:
- Find the nearest visual word from the vocabulary for each feature vector using distance matching, for instance, BFMatcher, knnMatch, radiusMatch or FlannBasedMatcher
- Compute the Bag-of-Words image descriptor as a normalized histogram of vocabulary words encountered in the image
- Save the Bag-of-Words descriptors for all images in the dataset.

3.5 Testing Stage

In this step, the proposed system will perform for each input image query.

- The query image is converted to a graph
- Segment the image based on community detection: Louvain method and ARA algorithm [23]
- Extract features based on the sub-segments and associates these characteristics to feature descriptors
- Compute the Bag-of-Words vector with the method defined above
- Compare the histogram of query image with groups histograms and retrieve the best results.

4 Experimental Results

We have implemented our CBIR systems to search in the database for images of similarities to a query image. The database is the collected images from the Corel dataset.

4.1 Dataset

In order to evaluate the proposed method, we selected the Corel dataset which is included 10 classes [14]. Each class contains 100 images on the same topic namely, Africans, Beaches, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. There is a total of 1000 images in the dataset for the experiment. We tested 1 image to all images in the dataset, so total test cases are 1000×1000 images. The experimental results are shown using qualitative and quantitative evaluations.

4.2 Results

For qualitative evaluations, we present some retrieved images from image queries implemented by proposed CBIR system. Detail of this qualitative evaluations are displayed in Figs. 5, 6, 7 and 8.

For quantitative evaluations, we present a comparison of our method to others of which used the same Corel dataset. We implement of each image is taken as a query image if an obtained image is within the exactly right class of the query image, it is a recorded success; otherwise, it recorded fail. The precision is given as Eq. 3. In Table 1, we present Top 1 retrieved image including the query image itself on every test case. We also display the statistics of Top 5 and Top 10 first candidates images on retrieved images results. The results infer that our method will reduce the ability to retrieve the similarities images in a large dataset. Table 2 presents the top 5 results of our study compared with those of other studies in the literature. Our experimental results are pretty fine however we need to do more experiments and enhance this method in many ways but this is left for future investigations.

$$Precision = \frac{Number\ of\ Relevant\ Images}{Number\ of\ Relevant\ Images + Number\ of\ Irrelevant\ Images} \tag{3}$$

Table 1. Precision proportion of Top 1, Top 5 and Top 10 implemented on dataset including all classes.

Class name	Top 1	Top 5	Top 10
10 classes	92.25%	60.15 %	52.81%

Fig. 5. Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

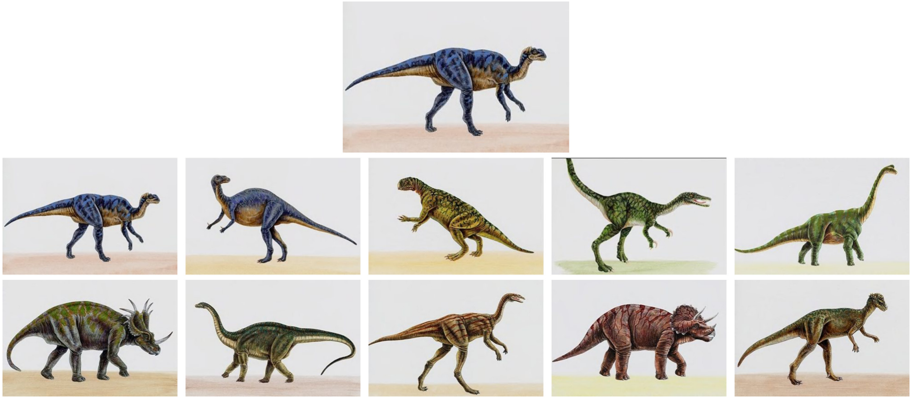

Fig. 6. Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

Fig. 7. Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

Table 2. Comparison of the top 5 results of proposed method with others.

Methods	10 classes	9 classes	6 classes	5 classes
Huang's method [11]	74.41%	78.71%	86.51%	94.23%
Khosla's method [13]	55.00%	–	–	–
Srivastava's method [32]	67.16%	–	–	–
Choudhary's method [4]	–	75.00%	–	–
Choudhari's method [3]	–	–	–	72.00%
Janani's method [12]	–	–	79.00%	–
Our method	60.15%	63.84%	77.20%	80.49%

Fig. 8. Qualitative evaluation of results using image query. *First line*: image query, *Second line*: images retrieved (Top 5), *Third line*: images retrieved (5 last images from Top 10)

5 Conclusion

The proposed content-based image retrieval system has been implemented to look for the most similar images to a given image. Our CBIR system of which the Bag-of-Visual-Words model using the local feature that we have proposed and extracted offers an efficient image retrieval method. Although our CBIR system has not been implemented on a large dataset or in deep and extensive experimental results, it has demonstrated its usefulness. We strongly believe that our research will be more advanced and deeper in several future studies [24].

References

1. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32
2. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theor. Exp. **10**, 10008 (2008). https://doi.org/10.1088/1742-5468/2008/10/P10008
3. Chaudhari, R., Patil, A.M.: Content based image retrieval using color and shape features. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **1**, 386–392 (2012)
4. Choudhary, R., Raina, N., Chaudhary, N., Chauhan, R., Goudar, R.H.: An integrated approach to content based image retrieval. In: 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2404–2410 (2014). https://doi.org/10.1109/ICACCI.2014.6968394

5. Cortés, X., Conte, D., Cardot, H.: Bags of graphs for human action recognition. In: Bai, X., Hancock, E.R., Ho, T.K., Wilson, R.C., Biggio, B., Robles-Kelly, A. (eds.) S+SSPR 2018. LNCS, vol. 11004, pp. 429–438. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97785-0_41
6. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: In Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1–22 (2004)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893, June 2005. https://doi.org/10.1109/CVPR.2005.177
8. Dang, Q.B., Coustaty, M., Luqman, M.M., Ogier, J.M.: A comparison of local features for camera-based document image retrieval and spotting. IJDAR **22**(3), 247–263 (2019)
9. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
10. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-Means clustering algorithm. Appl. Stat. **28**(1), 100–108 (1979)
11. Huang, Y.F., Chen, B.R.: Content-based image retrieval system for real images. In: Benzmüller, C., Sutcliffe, G., Rojas, R. (eds.) GCAI 2016. 2nd Global Conference on Artificial Intelligence. EPiC Series in Computing, vol. 41, pp. 95–108. EasyChair (2016). https://doi.org/10.29007/w4sr, https://easychair.org/publications/paper/Z3T
12. Janani, R.G.: An improved CBIR method using color and texture properties with relevance feedback. Int. J. Innovative Res. Comput. Commun. Eng. **2**, 47–54 (2014)
13. Khosla, G., Rajpal, N., Singh, J.: Evaluation of Euclidean and Manhanttan metrics in content based image retrieval system. In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 12–18, March 2015
14. Liu, G.H., Yang, J.Y., Li, Z.: Content-based image retrieval using computational visual attention model. Pattern Recogn. **48**(8), 2554–2566 (2015)
15. Liu, J.: Image retrieval based on bag-of-words model. CoRR abs/1304.5168 (2013). http://arxiv.org/abs/1304.5168
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
17. Mandal, R., Roy, P.P., Pal, U., Blumenstein, M.: Bag-of-visual-words for signature-based multi-script document retrieval. Neural Comput. Appl. **31**(10), 6223–6247 (2018). https://doi.org/10.1007/s00521-018-3444-y
18. Manzo, M., Pellino, S.: Bag of ARSRG words (BoAw). Mach. Learn. Knowl. Extr. **1**, 871–882 (2019). https://doi.org/10.3390/make1030050
19. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005). https://doi.org/10.1109/TPAMI.2005.188
20. Mikolajczyk, K., et al.: A comparison of affine region detectors. Int. J. Comput. Vis. **65**(1), 43–72 (2005). https://doi.org/10.1007/s11263-005-3848-x
21. Mukherjee, J., Mukhopadhyay, J., Mitra, P.: A survey on image retrieval performance of different bag of visual words indexing techniques. In: Proceedings of the 2014 IEEE Students' Technology Symposium, pp. 99–104 (2014). https://doi.org/10.1109/TechSym.2014.6807922
22. Newman, M.E., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)

23. Nguyen, T., Coustaty, M., Guillaume, J.: A new image segmentation approach based on the Louvain algorithm. In: 2018 International Conference on Content-Based Multimedia Indexing (CBMI), pp. 1–6 (2018). https://doi.org/10.1109/CBMI.2018.8516531
24. Nguyen, T.: Image segmentation and extraction based on pixel communities. Segmentation et extraction d'images basées sur des communautés de pixels. Ph.D. Thesis, University of La Rochelle, France (2019). https://tel.archives-ouvertes.fr/tel-03223157
25. Nguyen, T., Coustaty, M., Guillaume, J.: A combination of histogram of oriented gradients and color features to cooperate with Louvain method based image segmentation. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, pp. 280–291. INSTICC, SciTePress (2019). https://doi.org/10.5220/0007389302800291
26. Nguyen, T.K., Guillaume, J.L., Coustaty, M.: An enhanced Louvain based image segmentation approach using color properties and histogram of oriented gradients. In: Cláudio, A.P., et al. (eds.) Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 543–565. Springer International Publishing, Cham (2020)
27. Okawa, M.: Offline signature verification based on bag-of-visualwords model using KAZE features and weighting schemes. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 252–258 (2016). https://doi.org/10.1109/CVPRW.2016.38
28. Potapov, A., Zhdanov, I., Scherbakov, O., Skorobogatko, N., Latapie, H., Fenoglio, E.: Semantic image retrieval by uniting deep neural networks and cognitive architectures. CoRR abs/1806.06946 (2018). http://arxiv.org/abs/1806.06946
29. Sadeghi-Tehran, P., Angelov, P., Virlet, N., Hawkesford, M.J.: Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology. J. Imaging **5**(3), 33 (2019). https://doi.org/10.3390/jimaging5030033
30. Silva, F., Goldenstein, S., Tabbone, S., Torres, R.: Image classification based on bag of visual graphs, pp. 4312–4316 (2013). https://doi.org/10.1109/ICIP.2013.6738888
31. Silva, F.B., de O. Werneck, R., Goldenstein, S., Tabbone, S., da S. Torres, R.: Graph-based bag-of-words for classification. Pattern Recogn. **74**, 266–285 (2018). https://doi.org/10.1016/j.patcog.2017.09.018
32. Srivastava, P., Prakash, O., Khare, A.: Content-based image retrieval using moments of wavelet transform. In: The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014), pp. 159–164 (2014). https://doi.org/10.1109/ICCAIS.2014.7020550
33. Tsai, C.F.: Bag-of-words representation in image annotation: a review. ISRN Artif. Intell. **2012** (2012). https://doi.org/10.5402/2012/376804
34. Wan, J., et al.: Deep learning for content-based image retrieval: a comprehensive study. In: Proceedings of the 22Nd ACM International Conference on Multimedia, pp. 157–166. MM 2014, ACM (2014). https://doi.org/10.1145/2647868.2654948
35. Wang, H., Mohamad, D., Ismail, N.A.: Image retrieval: techniques, challenge, and trend. World Acad. Sci. Eng. Technol. **60**, 716–718 (2009)
36. Wang, H.H., Mohamad, D., Ismail, N.A.: Approaches, challenges and future direction of image retrieval. CoRR abs/1006.4568 (2010). http://arxiv.org/abs/1006.4568