TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 20, No. 5, October 2022, pp. 1026~1033
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v20i5.24086 ❒ 1026

A method for semantic-based image retrieval using hierarchical clustering tree and graph

Nguyen Minh Hai[1], Thanh The Van[2], Tran Van Lang[3]
[1]Faculty of Physics, HCMC University of Education, Ho Chi Minh City, Vietnam
[2]Faculty of Information Technology, HCMC University of Education, Ho Chi Minh City, Vietnam
[3]Journal Editorial Department, HUFLIT University, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Jul 11, 2021
Revised Jul 31, 2022
Accepted Aug 08, 2022

Keywords:

Data mining
Image retrieval
Ontology
SBIR
Similar image

ABSTRACT

Semantic extraction for images is an urgent problem and is applied in many different semantic retrieval systems. In this paper, a semantic-based image retrieval (SBIR) system is proposed based on the combination of growth partitioning tree (GP-Tree), which was built in the authors' previous work, with a self-organizing map (SOM) network and neighbor graph (called SgGP-Tree) to improve accuracy. For each query image, a similar set of images is retrieved on the SgGP-Tree, and a set of visual words is extracted relying on the classes obtained from mask region-based convolutional neural networks (R-CNN), as the basis for querying semantic of input images on ontology by simple protocol and resource description framework query language (SPARQL) query. The experiment was performed on image datasets, such as ImageCLEF and MS-COCO, with precision values of 0.898453 and 0.875467, respectively. These results are compared with related works on the same image dataset, showing the effectiveness of the methods proposed.

Corresponding Author:

Tran Van Lang
Journal Editorial Department, HUFLIT University, Ho Chi Minh City, Vietnam
Email: langtv@huflit.edu.vn

1. INTRODUCTION

In recent years, there have been many research groups to improve the efficiency of semantic-based image retrieval (SBIR) based on the built ontology [1]–[5]; built the image retrieval systems based on natural language analysis to generate a simple protocol and resource description framework query language (SPARQL) query that searches images set relied on image description resource description framework (RDF) [6]–[10]; proposed the image retrieval system based on relevant feedback techniques [11]; the image retrieval relied on ontology applying to text queries, multimedia data or to determine relationships between images by through image annotations and features [9], [12]–[17]. However, the set of similar images obtained has not really responded to user needs because of the difference between computational representations in machines and natural language. With the aim of minimizing the semantic gap to improve the performance of image retrieval.

The published works show that the image retrieval problem has many interests of the authors. Furthermore, applying a hierarchical clustering tree to perform semantic-based similar image retrieval is a viable and challenging approach. On the basis of inheriting from existing works and overcoming the limitations of related published methods [18]–[20], a semantic image retrieval system by combining graph-GPTree and self-organizing map (SIR-SgGP) is built. After classifying the query image based on mask region-based convolutional neural networks (R-CNN), the SPARQL statement is generated to query semantics and extract the uniform resource identifier (URI) of the images on an ontology structure that we proposed.

Our ontology-based model is proposed to support two main functions for semantic retrieval for image datasets: 1) retrieving a similar image set of given images and 2) mapping low-level features into high-level semantics of images based on ontology. A hierarchical clustering tree, called GP-Tree, was published in [20] to automatically store images indexed from low-level features of an image. The advantages of GP-Tree are multi-branch tree and clustering of feature vectors, so it can store large amounts of data and retrieve images quickly. However, the retrieval on the GP-Tree is performed by finding the branch with the highest similarity to the query image. Therefore, the query performance is not really high, so it is necessary to improve the retrieval efficiency on the GP-Tree.

This paper proposes a model to combine GP-Tree with self-organizing map (SOM) network and neighbor graph to limit the omission of similar elements occurring during branching in order to improve the accuracy of image retrieval. The input image is segmented to determine the classes of objects using the mask region-based convolutional neural networks (R-CNN) network. For each image segment, low-level features are extracted to form a combined feature descriptor. These feature sets will be retrieved on GP-Tree to extract a similar set of images. The SPARQL query is generated based on the visual words obtained from the set of classes by the mask R-CNN network and is queried on the ontology to extract the semantics of the query image. Experimental results on ImageCLEF and MS-COCO datasets and compared with published results in related works to evaluate the effectiveness of the proposed method. The contributions of the article include: 1) proposed a model combining GP-Tree with SOM network and neighbor graph (SgGP-Tree) to improve image retrieval efficiency; and 2) proposed a semantic-based image retrieval model that combines machine learning SgGP-Tree and ontology. The rest of the paper presents the necessary steps on the image query method according to the semantic approach as the main contribution of the paper (part 2), the experimental results on the datasets, as well as the assessment are presented in section 3; some conclusions are presented in the final section.

2. RESEARCH METHOD

2.1. Image segmentation and classification of objects in the image

In this paper, the pre-trained mask R-CNN model is used to detect objects in the image; from there, determine the classes for the input image. Figure 1 depicts the results of object recognition and classification on MS-COCO dataset by mask R-CNN based on ResNet-101-FPN [21]. For each extracted image segment, low-level features (color, texture, shape) are used to form an associative feature descriptor [22]. An 81-dimensional low-level feature vector is extracted for the image retrieval system in this paper.

Figure 1. Mask R-CNN results using ResNet-101-FPN on images in the MS-COCO dataset

2.2. Description of GP-Tree

GP-Tree [20] consists of a root, a set of nodes T, and a set of leaves L. Nodes are connected through the path of the parent–child relationship. The leaves L, which are nodes without child nodes, contain element data η so that $L = \{\eta_l \mid l = 1..M\}$, in which M is the maximum number of elements in a leaf. The element data $\eta = (f, \tau, c)$ contain the following elements: the feature vector of an image f, the identifier of an image τ, and classes of the image c. The nodes T have at least two child nodes, which contain the center element μ so that $T = \{\mu_k \mid k = 1..N\}$, where N is the number of elements in a node. Each element of a node is linked to its adjacent child node through τ of that node. The representative element $\mu = (f_c, l, \sigma)$ contains the following components f_c , or the center of feature vectors at a child node that has the path linking l to μ, and σ, which is the value used to check if the next subcluster has must be a leaf or not.

2.3. Neighbor cluster graph

Image retrieval on GP-Tree has not achieved high performance in the case the splitting node are many times. When splitting a leaf, similar elements can be split into separate branches. The graph-GPTree neighbor graph is built based on the set of leaves of the GP-Tree. New leaves created during splitting a leaf will be marked neighbors according to different criteria in order to link related leaves together, avoiding missing data in the retrieval process, thereby increasing performance in retrieving similar images. Based on the above analysis, the neighbor cluster graph is defined as:

a. Definition 1. The graph-GPTree neighbor cluster graph
   Graph-GPTree $G = (V, E)$ is an undirected graph, including:
   – The set of vertexes V The set of vertexes is are the clusters of leaves of the GP-Tree;
   – The set of edges $E \subseteq V \times V$ are the links of a pair of leaves, formed according to the neighbor relationship.

b. Definition 2. Neighbor cluster graph
   – The neighbor level-1st: let f_{cent_i}, f_{cent_j} be the center vector of the leaves L_i, L_j, respectively, where $f_{cent_i} = average\{f_i | f_i \in L_i, i = 1..|L_i|\}, f_{cent_j} = average\{f_j | f_j \in L_j, j = 1..|L_j|\}$;
     if $d_E\left(f_{cent_i}, f_{cent_j}\right) < \theta$, then L_i, L_j are neighbor level-1st, with d_E be the Euclidean distance and θ be a given threshold value.
   – The neighbor level-2nd: let m, n be the number of classes of the image appearing in two leaves L_t, L_k, respectively; c_t, c_k be the class that occurs most of those two leaves, where: $c_t = max\{count(\eta_i.c_j) \big| \eta_i \in L_t, i = 1..|L_t|, j = 1..m\}$,
     $c_k = max\{count(\eta_i.c_j) \big| \eta_i \in L_k, i = 1..|L_k|, j = 1..m\}$;
     if $c_t \equiv c_k$, then L_t, L_k are neighbor level-2nd.

Algorithm 1, the Algorithm of splitting a leaf and creating graph-GPTree is:

Algorithm 1. Algorithm neighbor graph: split the leaf on GP-Tree and create graph-GPTree

```
Input: Threshold θ, Leaf node L, Graph-GPTree;
Output: Graph-GPTree
Function neighborGraph (θ, Leaf, Graph-GPTree)
Begin
   # Find the two furthest elements in a leaf node
   center = average{L.EDi.f, i=1..M};
   left = argmax{Euclidean(center, Leaf.ED_i.f), i=1..M};
   right = argmax{Euclidean(L.ED_left.f, L.ED_i.f), i=1..M};
   EDLeft = L.ED_left; EDRight = L.ED_right;
   # Create new two leaf nodes
   L_l = L_l ∪ EDLeft; L_r = L_r∪ EDRight;
   # Allocates elements to two new leaf nodes
   Foreach ed in L do
      If (Euclidean(ed.f, EDLeft.f < Euclidean(ed.f, EDRight.f)) then L_l = L_l ∪
   ed;
      Else L_r= L_r∪ ed; EndIf
   EndForeach
   # Create center elements for two nodes: L_l & L_r
   ECLeft = average{L_l .ED_i.f, i=1..|L_l |};
   ECRight = average{L_r.ED_i.f, i=1..|L_r|};
   #Update presentation elements to parent
   L.Parent = Leaf.Parent ∪ {ECLeft, ECRight};
   # Determine the 1st-level neighbors of the two newly split leaf nodes
   If (Euclidean(ECLeft, ECRight) < θ) then
      Neighbor[1].L_l = Neighbor[1].L_l ∪ {L_r};
      Neighbor[1].L_r= Neighbor[1].L_r∪ {L_l };
   EndIf
   # Determine the 2nd-level neighbors of the two newly split leaf nodes
   LeftClass = argmax{count(L_l .ED_i.cla), i=1..|L_l |};
   RightClass = argmax{count(L_r.ED_i.cla), i=1..|L_r|};
   If (LeftClass = RightClass) then
      Neighbor[2].L_l = Neighbor[2].L_l ∪ {L_r};
      Neighbor[2].L_r= Neighbor[2].L_r∪ {L_l };
   EndIf
   Graph_GPTree = Graph_GPTree ∪ {Neighbor[1], Neighbor[2]};
   Return Graph_GPTree;
End
```

2.4. GP-Tree-graph-SOM

The graph-GPTree neighbor cluster graph has solved most of the problems of GP-Tree, improving image retrieval performance. However, the cluster selection criterion on the graph is by measure, so it can lead to measurement error when the tree performs node splitting many times; because when splitting a leaf, two new leaves may not create neighbors, but the representative elements of the split leaf (the most elements) have been allocated to the two new leaf nodes. Now more criteria are needed to select the winning leaf based on the weight of the representative element of the leaf. Therefore, the SOM network is built on the GP-Tree and graph-GPTree, called SgGP-Tree; to form a model that combines tree - graph - SOM.

The SgGP-Tree network consists of two layers: input and output. The input layer is the image feature vectors $f = (f_1, f_2, \dots f_m)$, where each f_i vector has n dimensions $f_i = (v_1, v_2, \dots, v_n)$. The output layer is the neuron containing the leaf set of the GP-Tree. The number of neurons is the number of representative classes of the leaf set of the GP-Tree and the neurons are labeled according to the classes; the leaf set of the GP-Tree is allocated to the neurons based on the representative classes of the leaves. The input layer is fully connected to the output layer by the weight vector set $W = (W_1, W_2, \dots W_p)$, where p is the number of neurons of the output layer and $W_i = (w_1, w_2, \dots w_n)$. Figure 2 depicts the structure of SgGP-Tree network based on GP-Tree and graph-GPTree graph. On the basis of SOM network, the SgGP-Tree network structure is defined as:

a. Definition 3. The SgGP-Tree network

The SgGP-Tree network is a SOM network whose input is the feature vectors of an image $f = (f_1, f_2, \dots f_m)$, where each f_i vector has n dimensions $f_i = (v_1, v_2, \dots, v_n)$, $f_i \in \{0,1\}$ and the output layer is the neurons containing the leaf set of the GP-Tree. The input and output layers are fully connected by the weight vectors $W_i = (w_1, w_2, \dots w_n)$, $w_i \in \{0,1\}$.

The purpose of the SgGP-Tree network is to classify the input data. The SgGP-Tree training process is the weight-training process. Adjusting the weights will make the network achieve the best classification requirements. However, the weight adjustment process takes a lot of time with large input datasets and incorrect random weighting. Therefore, instead of taking a random weight, a set of weight vectors trained on the GP-Tree is extracted. This weight vector is defined as:

b. Definition 4. Weight vector

Let w be the weight vector of data elements η at the leaf. The weight vector w is the center of the feature vectors of the most frequently occurring classes in the leaf.

$$w = \frac{\sum_{i=1}^{n} f_i}{n} \tag{1}$$

Where f_i is the feature vector value of the n classes that appear the most. The SgGP-Tree training process includes the following steps:

– Step 1: allocate leaves from GP-Tree to SgGP-Tree network.
– Step 2: initialize the initial weight w_i from the set of weights obtained during GP-Tree training.
– Step 3. randomly select a vector f_i as the training sample.
– Step 4: find the winning neuron using Euclidean distance.
  $i(f_i) = argmin_j \|f_i - w_j\|$
– Step 5: Update the weights of the neurons.
  $w_j = w_j + \gamma(t)(f_i - w_j)$
  Where: $\gamma(t) = \gamma_0 \exp\left(\frac{t}{\varphi_1}\right)$ is learning rate, t is epoch
– Step 6: repeat step 3 until training is over.

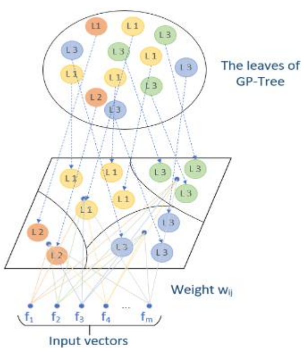


Figure 2. The SgGP-Tree combined network structure

2.5. Query image semantics on ontology

In order to query images according to the semantic approach, an ontology framework for images is proposed, using an ImageCLEF [22] image dataset. SPARQL is a query language on data sources described as RDF or web ontology language (OWL) triples. With the input query image that can contain one object or many objects, mask R-CNN is used to extract the visual words vector; this vector contains one or more semantic classes of the image, and automatically generates a SPARQL statement (and/or), from which to query the ontology to find the annotation of the query image [22]. The query result on the ontology is a set of URIs and the metadata of the query image.

3. RESULTS AND DISCUSSION

3.1. Datasets

To demonstrate the effectiveness of the proposed method, two popular image datasets are used for testing: ImageCLEF and MS-COCO. The ImageCLEF dataset consisting of 20,000 images and 276 class labels. Each image represents a single object and is annotated with a number of semantically relevant text tags. The MS-COCO dataset consists of 123,000 and 80 class labels. Each image consists of many objects and there is a caption for each object in the image.

3.2. The proposed SBIR system

The semantic-based image retrieval system based on SgGP-Tree and ontology is called SIR-SgGP. The query system consists of two phases, preprocessing and image retrieval. Figure 3 depicts the architecture of SIR-SgGP consisting of two specific phases as:

a. Pre-processing phase

1) the input image is segmented to determine the class of the objects in the image, and at the same time, the low-level features of the objects are extracted; from there, create data samples representing the image set of feature vectors and corresponding classes. Then, allocate the dataset on GP-Tree; 2) create a combined model GP-Tree-graph-SOM from the set of GP-Tree leaves; and 3) build a semi-automatic ontology framework from the WWW dataset and image dataset.

b. Image retrieval phase

1) create a representative data sample for the query image consisting of feature vectors and corresponding classes; 2) first, perform retrieval on SgGP-Tree to find the winning cluster. From there, based on the neighbor graph to retrieve the similar image set to the query image; 3) the visual words vector is created based on the classes of the query image obtained; from which the SPARQL query is generated and executes the query on a built ontology framework; the result of the query is the annotation of the query image; and 4) combining the results obtained from 2) and 3), we get a set of similar images and annotation of the query image.

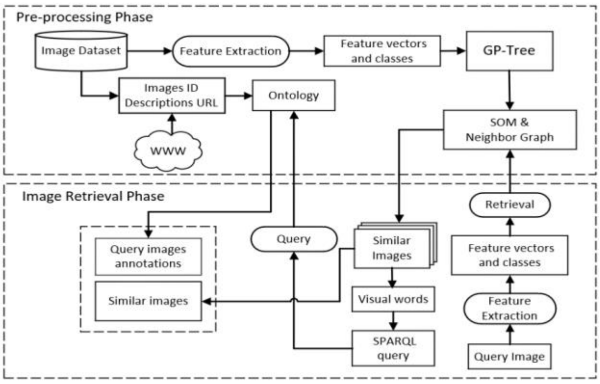
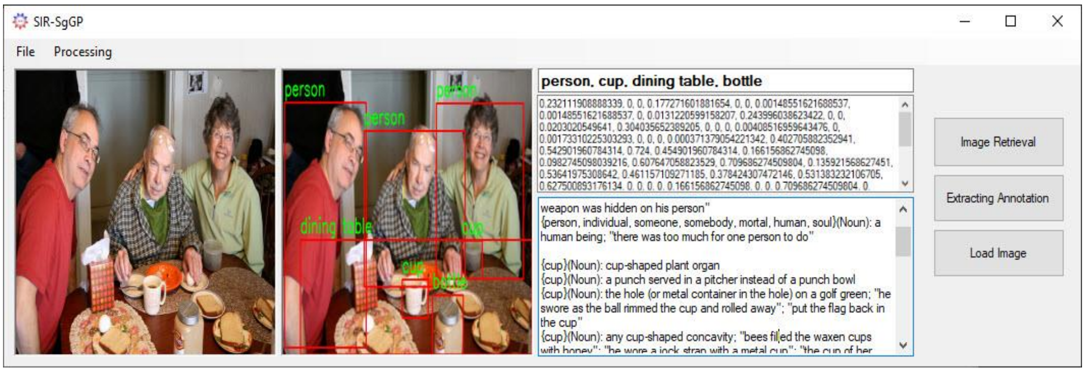


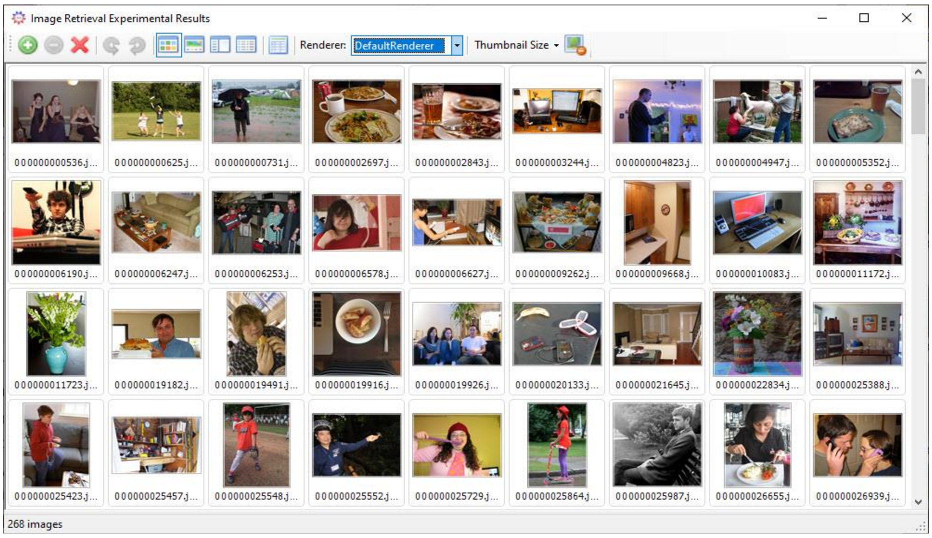
Figure 3. The SIR-SgGP model

3.3. Application

For each input image, the feature vectors and image classes are extracted using mask R-CNN. The set of feature vectors stored on the GP-Tree basing on the Euclidean similarity measure. With a query image, the SIR-SgGP system extracts features and classifies the image by mask R-CNN and searches the image according to the content on the SgGP-Tree association network to find a set of similar images according to the content. From the obtained classes of the query image, visual words are extracted; at the same time, the SPARQL query is also automatically generated to query the semantics of the input image on the ontology. Figure 4(a) describes the classification of objects in the image, annotations query corresponding to the objects, and image features extraction. Figure 4(b) depicts the image set similar to the query image extracted based on the image features in the SIR-SgGP system.



(a)



(b)

Figure 4. A similar image retrieval result of the SIR-SgGP system: (a) the results of classification, feature extraction, and annotations query for the 000000003827.jpg image of the MS-COCO dataset; and (b) the image set is similar to the query image extracted by the SIR-SgGP system

4. EXPERIMENTAL EVALUATION

To evaluate image search efficiency, the article uses evaluation factors including precision, recall and F-measure, and query time (milliseconds). On the basis of the available performance values, the values about performance and time search average for ImageCLEF and MS-COCO datasets of SIR-SgGP is summarized in Table 1. To assess the accuracy and efficiency of the proposed image retrieval system, the experimental results are compared with other studies on the same image datasets. Table 2 and Table 3 show that the retrieval results of the proposed method are relatively accurate compared to semantic-based image retrieval systems. The data of the tables show that our proposed method has higher accuracy when compared with other retrieval methods on the same image dataset. This proves that our proposed method is effective.

Table 1. Performance of the image retrieval system SIR-SgGP

Image dataset	Average precision	Average recall	Average F-measure	Average query time (ms)
ImageCLEF	0.898453	0.823434	0.813241	633.356
MS-COCO	0.875467	0.723454	0.783452	746.345

Table 2. Comparison of mean average precision of methods on ImageCLEF dataset

Method	Mean average precision (MAP)
D. Hu, 2019 [23]	0.643344
D. Wang, 2018 [24]	0.655644
SIR-SgGP	0.898453

Table 3. Comparison of mean average precision of methods on MS-COCO dataset

Method	MAP
Y. Cao, 2018 [25]	0.857645
Y. Xie, 2020 [26]	0.862848
SIR-SgGP	0.875467

5. CONCLUSION

In this paper, a semantic-based image retrieval method is proposed with the combination of GP-Tree-graph-SOM (SgGP-Tree). For each input image, features and image classes are extracted by mask R-CNN to create a visual word vector. From there, the SPARQL query is automatically generated from the visual word vector and executes a query on the ontology to retrieve the similar image set and annotation of the query image. An image retrieval model based on SgGP-Tree and ontology (SIR-SgGP) is proposed and experimented on ImageCLEF and MS-COCO datasets with the accuracy of 0.898453 and 0.875467, respectively. The experimental results are compared with other studies on the same set of image datasets, showing that our proposed method has higher accuracy. In the future research direction, we continue to improve the feature extraction methods of images to further improve the similar image set retrieval performance.

REFERENCES

[1] A. B. Spanier, D. Cohen, and L. Joskowicz, "A new method for the automatic retrieval of medical cases based on the RadLex ontology," *International journal of computer assisted radiology and surgery*, vol. 12, pp. 471-484, 2017, doi: 10.1007/s11548-016-1496-y.

[2] O. Allani, H. B. Zghal, N. Mellouli, and H. Akdag, "Pattern graph-based image retrieval system combining semantic and visual features," *Multimedia Tools and Applications*, vol. 76, pp. 20287-20316, 2017, doi: 10.1007/s11042-017-4716-8.

[3] B. Yu, "Research on information retrieval model based on ontology," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 30, 2019, doi: 10.1186/s13638-019-1354-z.

[4] M. N. Asim, M. Wasim, M. U. Ghani Khan, N. Mahmood, and W. Mahmood, "The Use of Ontology in Retrieval: A Study on Textual, Multilingual, and Multimedia Retrieval," in IEEE Access, vol. 7, pp. 21662-21686, 2019, doi: 10.1109/ACCESS.2019.2897849.

[5] B. Zhong, H. Li, H. Luo, J. Zhou, W. Fang, and X. Xing, "Ontology-based semantic modeling of knowledge in construction: classification and identification of hazards implied in images," *Journal of construction engineering management*, vol. 146, no. 4, 2020, doi: 10.1061/%28ASCE%29CO.1943-7862.0001767.

[6] M. A. Alzubaidi, "A new strategy for bridging the semantic gap in image retrieval," vol. 14, no. 1, 2017, doi: 10.1504/IJCSE.2017.10002209.

[7] V. Vijayarajan, M. Dinakaran, P. Tejaswin, and M. Lohani, "A generic framework for ontology-based information retrieval and image retrieval in web data," *Human-centric Computing Information Sciences,* vol. 6, no. 18, 2016, doi: 10.1186/s13673-016-0074-1.

[8] J. Filali, H. B. Zghal, and J. Martinet, "Towards visual vocabulary and ontology-based image retrieval system," in *International Conference on Agents and Artificial Intelligence*, 2016, pp. 560-565. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01557742/document

[9] O. Bchir, M. M. B. Ismail, and H. Aljam, "Region-based image retrieval using relevance feature weights," *International journal of fuzzy logic and intelligent systems*, vol. 18, no. 1, pp. 65-77, 2018, doi: 10.5391/IJFIS.2018.18.1.65.

[10] U. Manzoor, M. A. Balubaid, B. Zafar, H. Umar, and M. S. Khan, "Semantic image retrieval: An ontology based approach," vol. 4, no. 4, pp. 1-8, 2015, [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.7287&rep=rep1&type=pdf

[11] M. R. Hirwane, "Semantic based Image Retrieval," *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),* vol. 6, no.4, pp. 120-122, 2017. [Online]. Available: https://ijarcce.com/upload/2017/april-17/IJARCCE%2023.pdf

[12] S. Jabeen, Z. Mehmood, T. Mahmood, T. Saba, A. Rehman, and M. T. Mahmood, "An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model," *PloS one*, vol. 13, no. 4, p. e0194526, 2018, doi: 10.1371/journal.pone.0194526.

[13] J. Filali, H. B. Zghal, and J. Martinet, "Ontology-based image classification and annotation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 11, p. 2040002, 2020, doi: 10.1142/S0218001420400029.

[14] N. M. Shati, N. K. Ibrahim, and T. M. Hasan, "A review of image retrieval based on ontology model," *Journal of Al-Qadisiyah for computer science and mathematics*, vol. 12, no. 1, pp. pp. 10-14, 2020. [Online]. Available: https://www.iasj.net/iasj/download/602c59b8c067a036

[15] R. T. Icarte, J. A. Baier, C. Ruz, and A. Soto, "How a General-Purpose Commonsense Ontology can Improve Performance of Learning-Based Image Retrieval," *arXiv e-prints*, 2017, doi: 10.48550/arXiv.1705.08844.

[16] C. Wang, X. Zhuo, P. Li, N. Chen, W. Wang, and Z. Chen, "An Ontology-Based Framework for Integrating Remote Sensing Imagery, Image Products, and In Situ Observations," *Journal of Sensors,* vol. 2020, 2020, doi: 10.1155/2020/6912820.

[17] X. Wang, Z. Huang, and F. V. Harmelen, "Ontology-Based Semantic Similarity Approach for Biomedical Dataset Retrieval," in *International Conference on Health Information Science*, 2020, pp. 49-60, vol. 12435, doi: 10.1007/978-3-030-61951-0_5.

[18] N. M. Hai, V. T. Thanh, and T. V. Lang, "The improvements of semantic-based image retrieval using hierarchical clustering tree," in *Proceedings of the 13th National Conference on Fundamental and Applied Information Technology Research (FAIR'2020)*, 2020, pp. 557-570. [Online]. Available: http://vap.ac.vn/Portals/0/TuyenTap/2021/6/18/db78b606b7cc49b6873265977492fa99/71_FAIR2020_paper_58.pdf
[19] N. M. Hai and T. V. Lang, and T. T. Van, "Semantic-Based Image Retrieval Using Hierarchical Clustering and Neighbor Graph," in *World Conference on Information Systems and Technologies*, 2022, vol. 470, pp. 34-44, doi: 10.1007/978-3-031-04829-6_4.
[20] N. M. Hai, V. T. Thanh, and T. V. Lang, "A method of semantic-based image retrieval using graph cut," *Journal of Computer Science and Cybernetic*, vol. 38, no. 2, pp. 193-212, 2022. [Online]. Available: https://vjs.ac.vn/index.php/jcc/article/view/16786/2543254403
[21] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
[22] N. T. U. Nhi, T. M. Le, and T. T. Van, "A Model of Semantic-Based Image Retrieval Using C-Tree and Neighbor Graph," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1-23, 2022, doi: 10.4018/IJSWIS.295551.
[23] D. Hu, F. Nie and X. Li, "Deep Binary Reconstruction for Cross-Modal Hashing," in *IEEE Transactions on Multimedia*, vol. 21, no. 4, pp. 973-985, 2019, doi: 10.1109/TMM.2018.2866771.
[24] D. Wang, Q. Wang and X. Gao, "Robust and Flexible Discrete Hashing for Cross-Modal Similarity Search," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2703-2715, 2018, doi: 10.1109/TCSVT.2017.2723302.
[25] Y. Cao, M. Long, B. Liu and J. Wang, "Deep Cauchy Hashing for Hamming Space Retrieval," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1229-1237, doi: 10.1109/CVPR.2018.00134.
[26] Y. Xie, Y. Liu, Y. Wang, L. Gao, P. Wang, and K. Zhou, "Label-Attended Hashing for Multi-Label Image Retrieval," in *International Joint Conferences on Artificial Intelligence Organization (IJCAI)*, 2020, pp. 955-962, doi: 10.24963/ijcai.2020/133.

BIOGRAPHIES OF AUTHORS

Nguyen Minh Hai received his BSc degree in Computer Science Teacher Education in 2006 from Education University HCMC, Vietnam. In 2011, he obtained a Master degree in Data Transmission and Computer Network from Posts and Telecommunications Institute of Technology, Vietnam. From 2019 to now, he is a PhD student in Computer Science at Institute of Information Technology, Vietnam Academy of Science and Technology, Vietnam. His research interests include data mining, image processing, and image retrieval. He can be contacted at email: hainm@hcmue.edu.vn.

Thanh The Van received his BSc degree in Mathematics & Computer Science in 2001 from University of Science-Vietnam National University HCMC. In 2008, he obtained a Master degree in Computer Science from Vietnam National University HCMC. In 2016, he received a PhD degree in Computer Science from University of Science, Hue, Vietnam. His research interests include image processing, image mining, and image retrieval. He can be contacted at email: thanhvt@hcmue.edu.vn.

Tran Van Lang is a senior principal research scientist in computer science at the Vietnam Academy of Science and Technology (VAST), who has been interested in the development of bioinformatics and parallel computing in Vietnam. He is currently the editor-in-chief of HUFLIT Journal of Science (HJS); the deputy editor-in-chief of the Vietnam Journal of Computer Science and Cybernetics (JCC); and the chief editor of the Electronics and Telecommunication Section of the Vietnam Journal of Science and Technology (VJST). He was a member of the directorship of the Institute of Applied Mechanics and Informatics, of the VAST; as well as the dean of the Information Technology Faculty, Lac Hong University (LHU) and Nguyen Tat Thanh University (NTTU). He received a B.Sc. degree in mathematics (in 1982) and a Ph.D. degree in mathematics–physics (in 1995) from the HCMC University of Natural Sciences (Vietnam). And he has been an Associate Professor in computer science at the Graduate University of Science and Technology, of the VAST. He also has worked as a researcher in computational mathematics at the Dorodnitsyn Computing Center, Russian Academy of Science. His interests include bioinformatics, parallel and distributed computing, computational intelligence, scientific computation, and computational mathematics. His work has appeared in Vietnamese and international journals and proceedings. He also was head of about 20 ICT projects. He can be contacted at email: langtv@huflit.edu.vn.