MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

NGUYEN MINH HAI

DEVELOPING A SEMANTIC-BASED IMAGE RETRIEVAL ALGORITHM

SUMMARY OF DISSERTATION ON COMPUTER

Major: Computer Science

Code: 948 01 01

Ho Chi Minh City - 2024

The dissertation is completed at: Graduate University of Science and Technology, Vietnam Academy Science and Technology

Supervisors:

1. Supervisor 1: Assoc. Prof. Dr. Tran Van Lang, Institute of Applied Mechanics and Informatics, Vietnam Academy of Science and Technology
2. Supervisor 2: Dr. Van The Thanh, Ho Chi Minh City University of Education

Referee 1: Assoc. Prof. Dr. Tran Cong Hung

Referee 2: Assoc. Prof. Dr. Huynh Tuong Nguyen

Referee 3: Dr. Le Quang Minh

The dissertation is examined by Examination Board of Graduate University of Science and Technology, Vietnam Academy of Science and Technology at………………………..

The dissertation can be found at:

1. Graduate University of Science and Technology Library
2. National Library of Vietnam

INTRODUCTION

1. Overview

Machine vision mainly focuses on the application of electronic images and computer support, particularly in analyzing images collected from systems such as pedestrian tracking, traffic analysis, security monitoring, and deep space exploration [4]. However, to achieve better results in machine vision applications, there has been an increasing effort to combine machine learning with image processing methods during analysis [5]. This combination is considered a way to create more intelligent systems and has become popular in image retrieval applications [6].

2. Urgency of the thesis

Content-Based Image Retrieval (CBIR) does not analyze semantics and fails to describe image content with high-level semantics, leading to search performance not meeting user requirements. The approach of Semantic-Based Image Retrieval (SBIR) [37, 38], which aims to describe semantic concepts, has been proposed in conjunction with CBIR to enhance search efficiency. However, interpreting an image's semantics by a computer is a highly complex task. The main challenge in SBIR is transforming a large image into low-level features and linking these features to the image’s high-level semantic content. From there, a semantic image retrieval model is created, allowing the computer to understand the actual semantics desired by the user. Thus, semantic-based image retrieval is a highly researched topic worldwide and performs well when addressed effectively. With the desire to contribute an efficient semantic image retrieval method, this thesis undertakes the topic: "Development of a Semantic-Based Image Retrieval Algorithm."

3. Research Objectives of the Thesis

The thesis focuses on the following objectives:

(1) Analyze and extract semantic features from images.

(2) Understand and process the components in an image to generate semantics for the image.

(3) Develop an image search and classification system capable of understanding and effectively responding to user needs.

4. Research Methodology

Theoretical Method: Synthesize related works on semantic-based querying using machine learning methods and tree-structured storage. Analyze the pros and cons of these works; study ontology enrichment methods and develop a semantic image retrieval model based on ontology. Propose semantic-based image retrieval models; conduct experimental evaluations based on the proposed models, comparing the search accuracy with existing works to make suitable adjustments and improvements.

Experimental Method: Based on the methods and models proposed in the thesis, the experimental program is implemented on machines with similar configurations. The experimental data chosen are reliable image datasets, widely published and used in prior research, for comparison with the experimental results of the proposed models to demonstrate the validity and effectiveness of the theoretical basis.

5. Research Subjects and Scope

Search Algorithms: Methods and techniques for developing algorithms to search and classify (clustering and classification) images based on semantic information rather than just image features like color and shape. Simultaneously, study some machine learning methods to improve image search efficiency. Image Semantics: Semantic elements in images need to be understood and identified to support accurate search. Image Data: Image datasets containing the necessary semantic information to train and evaluate the algorithm.

Research Scope: (1) **Semantic Scope**: Recognize specific types of semantics such as objects (e.g., cars, people, animals), actions (e.g., running, jumping), or contexts (e.g., outdoor, indoor) using ontology and the SPARQL query language; (2) **Spatial Scope**: Focus on one or a few image datasets that contain the necessary semantic information to train and evaluate the algorithm, such as standard datasets like Wang, MS-COCO and ImageCLEF; (3) **Methodological Limitation**: Focus on ontology-based methods to link semantics and images.

6. Thesis Contributions

The thesis develops a semantic-based image retrieval algorithm using the visual features of images, based on the GP-Tree data structure to enhance search accuracy, including:

(1) Building a hierarchical clustering GP-Tree data structure to store the feature vectors of images.

(2) Developing the GP-Tree structure using supervised and semi-supervised learning algorithms to improve image search efficiency.

(3) Constructing a semantic image retrieval system based on the GP-Tree structure and ontology to demonstrate the effectiveness of the proposed search methods.

7. Thesis Content and Structure

The thesis is structured as follows:

- **Chapter 1:** Provides an overview of the image retrieval problem, with two main directions: content-based image retrieval and semantic-based image retrieval. Related research works are surveyed and analyzed to identify challenges and limitations in existing methods, from which the specific research direction of the thesis is proposed to address these limitations. This section also details the experimental setup, including the environment, dataset selection, and performance evaluation criteria.

- **Chapter 2:** Presents related studies on using tree structures for storing and indexing image datasets. The GP-Tree clustering structure is described in detail, including the operations of adding, modifying, splitting, and deleting elements. A semantic image retrieval model based on the GP-Tree using ontology is proposed, aiming to improve search performance and accuracy. This image retrieval system has been tested on popular datasets such as Wang, MS-COCO, and ImageCLEF to evaluate the results and effectiveness of the proposed model.
- **Chapter 3:** Discusses methods to improve the GP-Tree clustering structure to enhance image search efficiency. Specifically, methods like Graph-GPTree clustering and the SgGP-Tree network are introduced to improve storage and retrieval of similar elements. Moreover, the semantic-based image retrieval method using ontology is discussed, with the SgGP-Tree structure being utilized to classify objects in images more accurately. A semantic-based image retrieval model combining ontology and the SgGP-Tree structure has been proposed and tested on popular datasets like Wang, MS-COCO, and ImageCLEF to assess the model's effectiveness.
- Conclusion and Future Directions: Summarizes the achievements and presents future development directions of the thesis.
- List of Author's Publications: Lists the works published by the author during the thesis.
- References: Lists the references cited in the thesis.

CHAPTER 1. OVERVIEW OF IMAGE RETRIEVAL

1.1. Overview of Image Retrieval

The image retrieval problem addressed in this thesis is defined as finding a set of images most similar to the input image based on a similarity measure between images [1].

1.2. Common Features in Image Retrieval

A feature is defined as the identification of a visual attribute of an image [2]. In general, image features can be global or local [3].

1.3. Related research works on image retrieval

1.3.1. Text-based image retrieval

Text-based method is a simple traditional keyword-based search method. Images are indexed by content, such as image captions; file names, web page titles and alt tags.... and stored in a database

1.3.2. Content-Based Image Retrieval

In a content-based image retrieval (CBIR) system, the effectiveness of the computational process depends on the ability to extract low-level features and measure similarity. In CBIR methods, low-level visual features such as color, shape, texture, and spatial layout are extracted from the images to perform the retrieval process.

1.3.3. Semantic-Based Image Retrieval

In the field of image retrieval, the main challenge lies in converting images into low-level features that can be computed by the machine and linking them to high-level concepts to reduce the semantic gap. Several semantic-based image retrieval (SBIR) methods have been proposed to bridge the semantic gap [123], such as semantic-based image retrieval using machine learning techniques and ontology-based semantic image retrieval.

1.4. Experimental Setup and Evaluation Methods

The proposed models were built and experimented upon using the dotNet Framework 4.5 and Python 3.7 platforms. The datasets used in the experiments of this thesis are widely used and popular datasets in image retrieval research, including Wang, ImageCLEF, and MS-COCO. The results need to be evaluated for performance, which includes image classification outcomes and the accuracy of image retrieval. The performance evaluation metrics for classification and image retrieval used in this thesis include precision (P), recall (R), and the F-measure (F_m).

1.5. Chapter conclusion

This chapter provides a detailed overview of modern image retrieval methods, including two main approaches: content-based image retrieval and semantic image retrieval.

This chapter also presents specific experimental organization methods, including experimental environment setup, experimental data set selection and preparation, as well as performance evaluation metrics of search methods.

The next chapter presents a new data structure, the hierarchical clustering tree, which is proposed to optimize the storage and indexing of image data sets. This structure not only supports fast retrieval but also enhances the efficiency of semantic image retrieval, allowing the system to process large image data sets quickly and efficiently.

CHAPTER 2. GP-TREE STRUCTURE FOR SEMANTIC IMAGE RETRIEVAL

This chapter presents studies related to tree structures for storing and indexing image datasets. It describes the GP-Tree clustering structure and the operations of adding, modifying, splitting, and deleting elements in the tree. A semantic-based image retrieval model using GP-Tree based on ontology is proposed, and the image retrieval system is tested on popular image datasets such as Wang, MS-COCO, and ImageCLEF to evaluate the performance of the proposed model. The content of this chapter is directly related to two published works [CT4] and [CT6], and indirectly to [CT1], [CT2], and [CT3].

2.1. Introduction

This chapter presents several clustering tree structures and related studies; the technique for designing the balanced GP-Tree clustering structure to store large image datasets; and the operations of adding, modifying, splitting, and deleting elements on the tree. It proposes a model, algorithm, and experiments for semantic-based image retrieval on GP-Tree with the proposed image datasets.

2.2. GP-Tree Data Structure

Based on the multi-branch tree structure and the K-Means clustering method, the GP-Tree structure is built by splitting a leaf node into two nodes if the number of elements in that leaf node exceeds a given threshold M. Each child node can be created based on a threshold θ, which evaluates the similarity of the data. If a data element has a similarity deviation greater than the threshold θ, that data element must belong to a different branch. Thus, the GP-Tree grows in the direction of the leaves and develops into a multi-branch tree where each leaf node is a cluster of similar elements. The data elements, which are feature vectors for each image, are sequentially stored on the GP-

Tree, allowing various operations to be performed. The hierarchical clustering structure of the GP-Tree is illustrated in **Figure 2.1**.

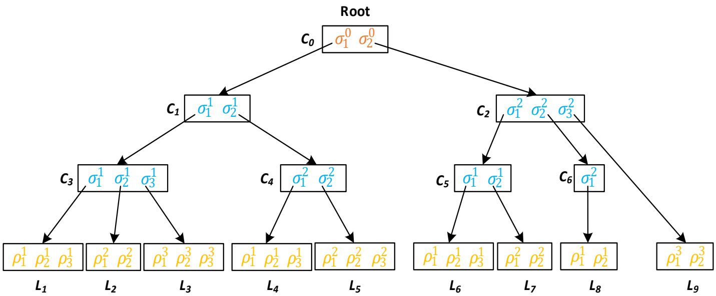


Figure 2.1. Hierarchical clustering GP-Tree with 3 levels

2.3. Image Retrieval Using GP-Tree Structure

2.3.1. GP-Tree based image retrieval system

The image retrieval system using GP-Tree consists of two phases: (1) The preprocessing phase performs segmentation, feature extraction from the image dataset, extracts concept classes, and organizes storage on the GP-Tree; (2) The query phase searches for similar images by content on the GP-Tree and extracts the image semantics. **Figure 2.4** presents the model for semantic-based image retrieval using the GP-Tree structure, with the two phases detailed as follows:

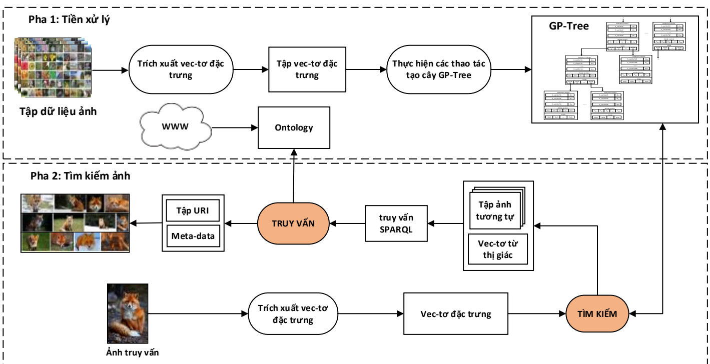


Figure 2.4. Model of the GP-Tree-based Image Retrieval System (GP-SBIR)

2.3.2. Experiment and Evaluation of the GP-SBIR System

Table 2.1: GP-Tree Experiment Results

Image Dataset	**Number of Images**	**Experimental Parameters**		**Experimental Time (seconds)**	**Number of Leaf Clusters**	**Number of Internal Clusters**	**Number of P@ Samples**	**P@ Rate**
		M	**N**					
WANG	10,800	100	40	17,839.47	218	32	2,240	20%
ImageCLEF	20,000	150	50	32,173.95	432	67	6,000	20%
MS-COCO	163,957	300	70	158,765.84	782	314	44,188	20%

Table 2.2: Query Performance of the GP-SBIR System on Test Datasets

Image Dataset	**Evaluation Metrics**			
	Accuracy	**Recall**	**F-measure**	**Average Search Time (ms)**
WANG	0.6078	0.4896	0.5424	98.75
ImageCLEF	0.6062	0.4094	0.4887	132.09
MS-COCO	0.647	0.564	0.603	217.65

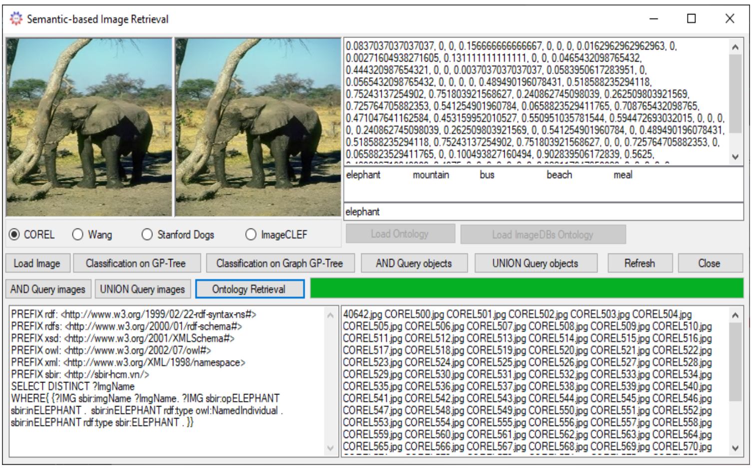


Figure 2.18. GP-SBIR Image Retrieval System

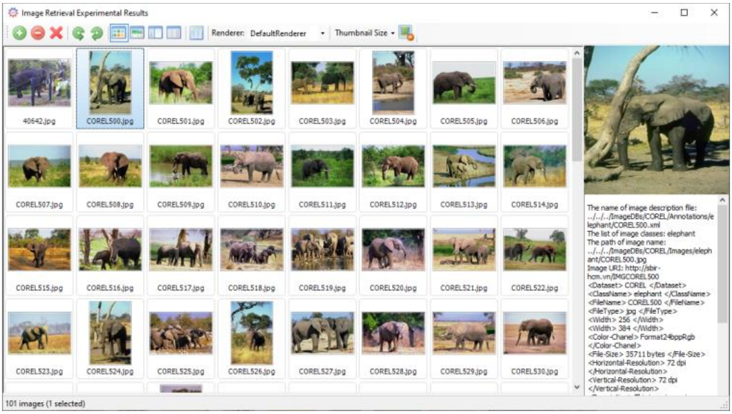


Figure 2.19. Similar Image Set for the Query Image on GP-SBIR

The GP-Tree-based image retrieval system (GP-SBIR) is illustrated in **Figure 2.18**. For each input image I_q (Load Image), selected from the image sets (Wang, ImageCLEF, MS-COCO), feature vectors are extracted and searched on the GP-Tree to retrieve similar images by content. The result set SI for the input image I_q, based on GP-Tree, is a visual term extraction process. The visual term set is built from images in the most frequently occurring class. From there, a SPARQL query is created (Create SPARQL) to search for semantically similar images based on the ontology (Load Ontology). The result set of semantically similar images for the input image is illustrated in **Figure 2.19**.

2.4. Chapter conclusion

The GP-Tree clustering tree can store large datasets and is effective for image retrieval. The GP-Tree performs partitioning and hierarchical clustering, allowing fast searching within branches with the most similar measurements to find similar elements at the leaf nodes. Thus, the GP-Tree offers fast search times and relatively high accuracy. However, its main drawback is that during node splitting, similar elements can be separated into different nodes, which in the worst-case scenario may reside on different branches. This can lead to missed similar elements during the retrieval process, affecting search performance. Therefore, improved tree methods are proposed in the next chapter to enhance retrieval accuracy on the GP-Tree. Additionally, a semantic-based image retrieval model using ontology is proposed to improve search performance and extract semantics for query images. The issue of semantic image retrieval using the GP-Tree structure is discussed in the next chapter.

CHAPTER 3. THE SGGP-TREE STRUCTURE FOR SEMANTIC IMAGE RETRIEVAL

This chapter introduces methods to improve the GP-Tree clustering structure to enhance image retrieval efficiency. Specifically, methods such as Graph-GPTree and SgGP-Tree are discussed to improve the storage and retrieval of similar elements. Additionally, semantic image retrieval methods based on ontology are examined, with the SgGP-Tree structure being used to classify objects in images more accurately. A semantic image retrieval model that integrates ontology with the SgGP-Tree structure is proposed and tested on popular datasets such as WANG, MS-COCO, and ImageCLEF to assess its effectiveness. The key content of this chapter has been published in [CT1], [CT2], [CT3].

3.1. Introduction

This chapter presents improved methods for the GP-Tree clustering structure to enhance image retrieval: Graph-GPTree and the integrated SgGP-Tree network. The chapter discusses semantic image retrieval methods based on ontology, where the SgGP-Tree structure is used to classify objects in images. It also details the development of ontology for multi-object image datasets. A semantic image retrieval model based on ontology combined with the SgGP-Tree structure is proposed and evaluated through experimentation.

3.2. Graph-GPTree

3.2.1. Graph-GPTree structure

Graph-GPTree is built based on sparse graph operations involving leaf nodes obtained from the GP-Tree. Here, vertices represent the leaf nodes, and the edges, weighted, represent their similarity. The sparse graph is created during the GP-Tree generation when, upon splitting each leaf node, the system marks the neighboring levels of the newly split leaf nodes (**Figure 3.2**).

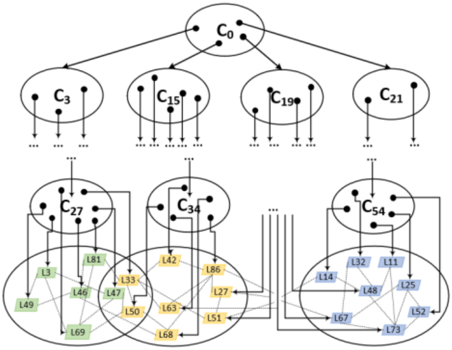


Figure 3.2. The sparse graph created must have a set of GP-Tree leaf nodes.

3.2.2. Graph-GPTree construction process

The graph clustering algorithm GraphGP-Tree, used to find clusters in a dataset, is described in **Figure 3.3**. The algorithm works on a sparse graph where the nodes represent data elements, and the weighted edges signify their similarity. This sparse graph representation allows the clustering algorithm to scale to large datasets.

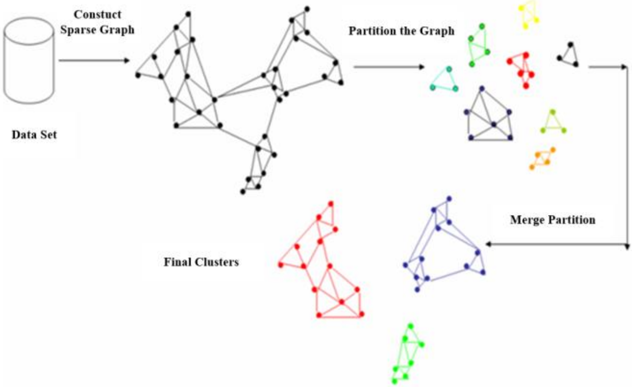


Figure 3.3. Overview of the overall approach of the GP-Tree generation algorithm

3.2.3. Image retrieval model on Graph-GPTree

The image retrieval model based on the neighboring clusters of Graph-GPTree is described in **Figure 3.4**, with the following steps: extract the feature vector for the input image; compare the feature vector with the feature database on the GP-Tree to select the most similar branch and identify the appropriate leaf node; from the selected leaf node on the GP-Tree, use the Graph-GPTree neighborhood graph to find the neighboring leaf nodes, leading to a set of similar images arranged by increasing similarity; use the k-NN algorithm to classify the similar image set and perform visual vocabulary search to explain the query image's semantics.

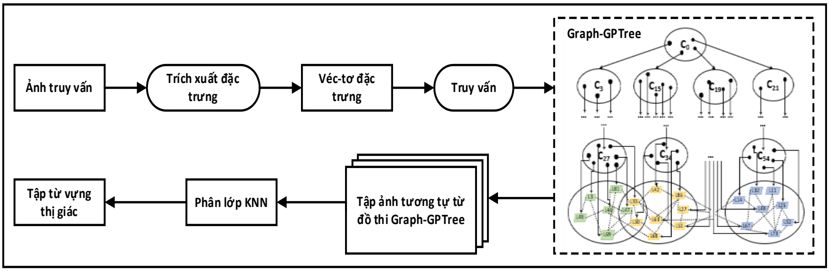


Figure 3.4. Image search model on Graph-GPTree neighborhood cluster graph

3.3. SgGP-Tree Integrated Network

3.3.1. SgGP-Tree Structure

Cấu trúc SgGP-Tree là sự kết hợp giữa cây GP-Tree, Graph-GPTree và mạng SgGP-Tree is a combination of the GP-Tree, Graph-GPTree, and SOM network. In the SOM network, weight adjustments during training achieve optimal clustering. However, the weight adjustment process is costly for large image datasets, and random weight initialization may lead to entirely different maps. Additionally, introducing new data into the map after SOM training will result in misclassification since SOM is static; hence, SOM must be retrained from scratch. The combined model, consisting of the GP-Tree, Graph-GPTree, and grSOM, called SgGP-Tree, is depicted in Figure 3.5.

3.3.2. Image retrieval model on SgGP-Tree integrated network

The image retrieval model on the SgGP-Tree integrated network, known as SBIR-SgGP, uses the SgGP-Tree structure created by combining GP-Tree, Graph-GPTree, and SOM (**Figure 3.5**). During preprocessing, features are extracted from the image dataset and stored in the SgGP-Tree structure. During image retrieval, SgGP-Tree is used to retrieve similar images and the visual vocabulary. The feature extraction block from previous models is reused in this search model. The innovation lies in improving the GP-Tree clustering structure by integrating the Graph-GPTree neighboring cluster graph and the SOM network to create SgGP-Tree (**Figure 3.6**).

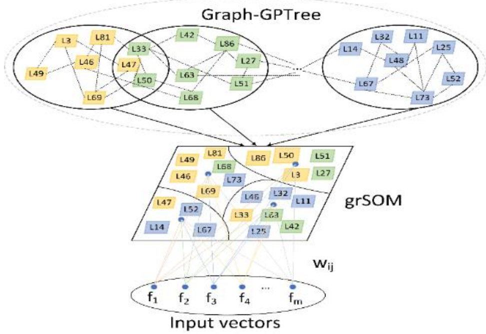
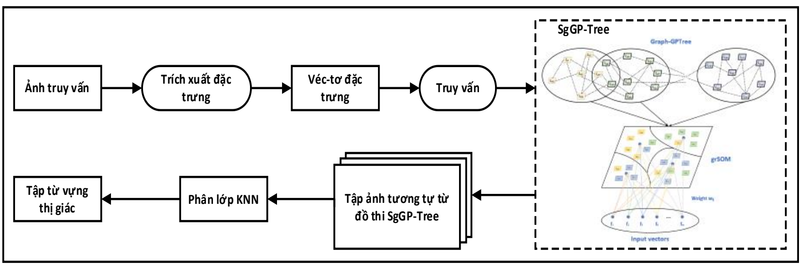


Figure 3.5. SgGP-Tree combination model



Hình 3.6. Mô trình tìm kiếm ảnh trên SgGP-Tree (SBIR-SgGP)

3.4. Semantic Image Retrieval System Based on Ontology

3.4.1. Ontology-based Image Retrieval Model

An ontology-based semantic image retrieval system, OntoSBIR, is built based on the proposed ontology methods. It consists of two phases, as illustrated in **Figure 3.7**. The feature extraction and content-based image retrieval methods are inherited from the previous chapters of the dissertation. The main contribution of this model is the ontology construction and ontology-based image retrieval.

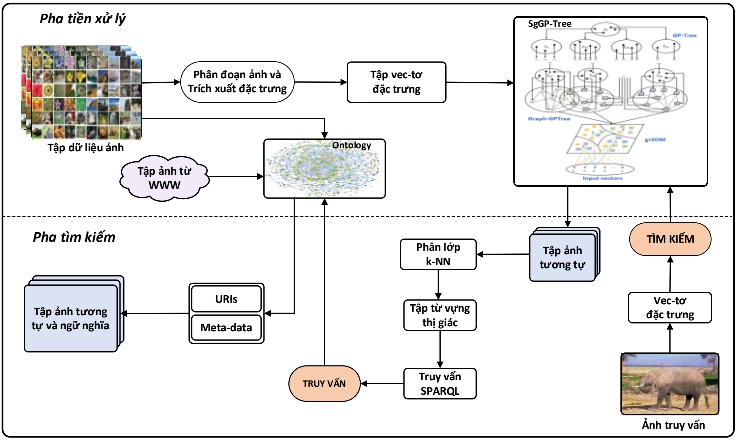


Figure 3.7. SBIR-GP search system model

3.4.2. Experiments and evaluation of the SBIR-GP image retrieval system

The SBIR-GP image retrieval system is designed to perform semantic-based image queries using SgGP-Tree and ontology. When given an input image, the SBIR-GP system first extracts feature vectors and retrieves similar images based on their content sequentially using SgGP-Tree. This process results in a collection of similar images**. Figure 3.8** is the interface of the OnSBIR system with images as input. The result is a list of images with the same semantics as the input image and metadata descriptions of the image set. At the same time, the semantic concept of the classification vocabulary is extracted from the ontology's thesaurus. **Figure 3.9**. is an example of the classification concept of the ontology dictionary.

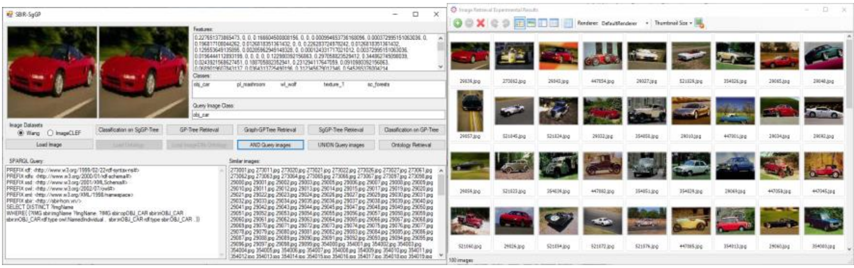


Figure 3.8: A result from the SBIR-GP search system based on the input image.

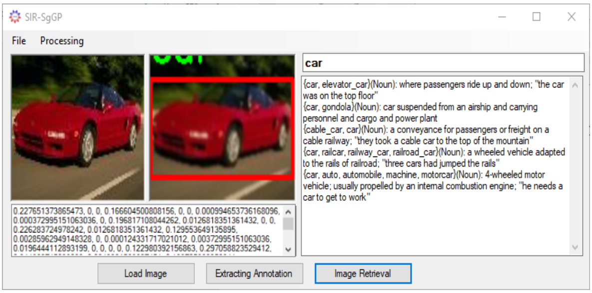


Figure 3.9: Semantic concept for the class.

The image datasets used for the experiments include WANG, MS-COCO, and ImageCLEF datasets. The average performance values and search times for the test datasets are shown in Table 3.1, Table 3.2, and Table 3.3.

Table 3.1. Image search performance on WANG dataset

Method	Precision	Recall	F-Measure	Average Search Time (ms)
GP-Tree	0.6780	0.6840	0.6810	39.75
Graph-GPTree	0.7665	0.6677	0.7137	202.79
SBIR-GP	0.8004	0.7040	0.7491	696.19

Table 3.2. Image search performance on ImageCLEF dataset

Method	Precision	Recall	F-Measure	Average Search Time (ms)
GP-Tree	0.6802	0.7750	0.7245	44.09
Graph-GPTree	0.8168	0.7637	0.7894	239.29
SBIR-GP	0.8926	0.8764	0.8844	868.51

Table 3.3. Image search performance on MS-COCO dataset

Method	Precision	Recall	F-Measure	Average Search Time (ms)
GP-Tree	0.717	0.724	0.7205	102.32
Graph-GPTree	0.873	0.764	0.815	198.47
SBIR-GP	0.875	0.724	0.783	265.45

From the above tables, we can observe that improving the GP-Tree structure leads to better precision for the WANG, ImageCLEF, and MS-COCO datasets. The Graph-GPTree clustering method outperforms GP-Tree but is still slightly behind SBIR-GP in performance. However, the search time for GP-Tree is faster compared to both Graph-GPTree and SBIR-GP.

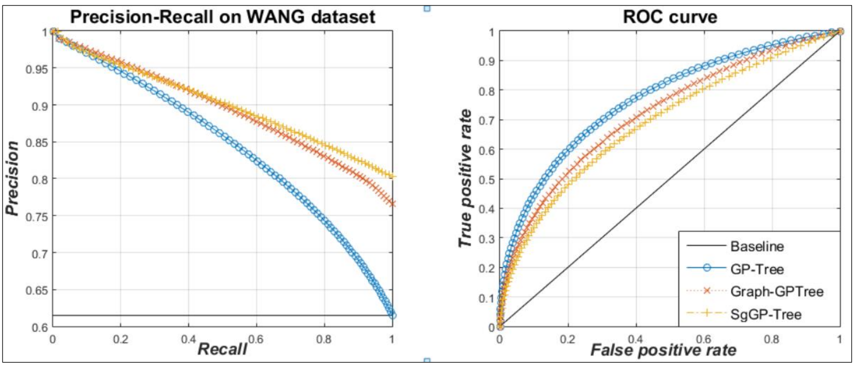


Figure 3.10. Image search performance on GP-Tree, Graph-GPTree, and SBIR-GP (SgGP-Tree) for WANG dataset.

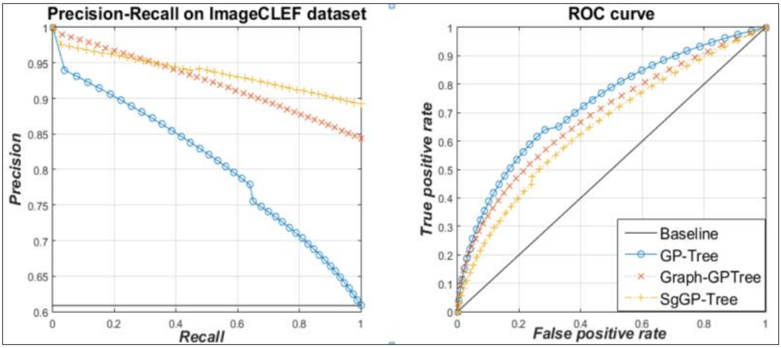


Figure 3.11. Image search performance on GP-Tree, Graph-GPTree, and SBIR-GP (SgGP-Tree) for ImageCLEF dataset.

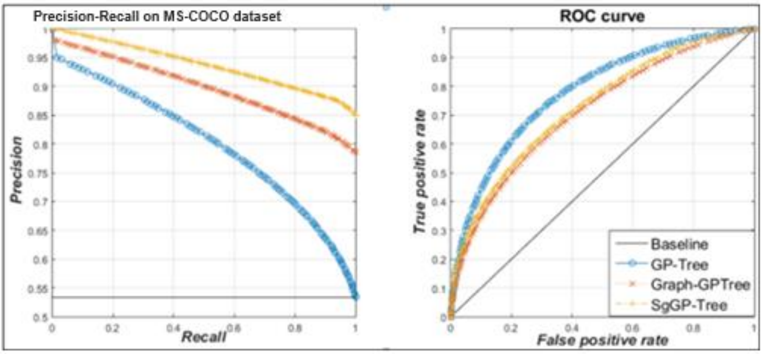


Figure 3.12. Image search performance on GP-Tree, Graph-GPTree, and SBIR-GP (SgGP-Tree) for MS-COCO dataset.

Based on experimental data, the PR (Precision-Recall) curves and ROC (Receiver Operating Characteristic) curves were generated to assess the accuracy of the SBIR-GP system (**Figure 3.10, Figure 3.11,** and **Figure 3.12**). To evaluate the accuracy and efficiency of the SBIR-GP image search system, we compared its performance with other research works on the same image datasets. The results in **Tables 3.4, 3.5,** and **3.6** compare the average precision values of different search methods on the WANG, ImageCLEF, and MS-COCO datasets.

Table 3.4. Comparison of Image Retrieval Methods on the WANG Dataset

Method	Average Precision
K. Kanwal et al. (2020) [4]	0.5067
H. Zeng et al. (2021) [5]	0.6600
O. Sikha and K. Soman (2021) [6]	0.8030
S. Dhingra and P. Bansal (2021) [7]	0.6000
A. Ouni et al. (2022) [8]	0.7800
Graph-GPTree	**0.7665**
SBIR-GP (SgGP-Tree)	**0.8004**

Table 3.5. Comparison of Image Retrieval Methods on the ImageCLEF Dataset

Method	Average Precision
A. Yang et al. (2019) [9]	0.8030
Y. Qiang et al. (2020) [10]	0.6670
X. Yue et al. (2021) [11]	0.7140
N. T. U. Nhi et al. (2022) [12]	0.6510
X. Wang et al. (2023) [13]	0.7270
Graph-GPTree	**0.8168**
SgGP-Tree (SBIR-GP)	**0.8926**

Table 3.6. Comparison of Image Retrieval Methods on the MS-COCO Dataset

Method	Average Precision
Y. Cao, 2018 [14]	0.8576
Y. Xie, 2020 [15]	0.8628
Wen Gu, 2019 [16]	0.8350
Graph-GPTree	**0.8730**
SgGP-Tree (SBIR-GP)	**0.8753**

3.5. Chapter Summary

This chapter proposed methods to enhance image retrieval efficiency on the GP-Tree. First, a model combining the neighborhood graph with the GP-Tree, known as Graph-GPTree, was created to connect similar elements branched during node splitting on the GP-Tree. Next, a model combining grSOM and Graph-GPTree, known as SgGP-Tree, was developed to enhance image search efficiency. SgGP-Tree adds criteria to select the winning leaf nodes, enabling better clustering and more accurate image searches. Experiments were conducted on the WANG, ImageCLEF, and MS-COCO datasets. The SBIR-GP system demonstrated superior accuracy compared to the author's previous proposals. Experimental performance was compared with other methods on the same image datasets to evaluate the proposed model, method, and algorithm. The comparison results showed that the SBIR-GP system achieved higher accuracy than other studies on the same image datasets, proving the effectiveness and suitability of the dissertation's proposals.

CONCLUSION AND DEVELOPMENT DIRECTIONS

1. Contributions of the Thesis

In this thesis, semantic image retrieval methods have been proposed and developed through an in-depth analysis of related studies to build high-performance image retrieval models. The thesis focuses on combining clustering techniques and ontologies, with three main contributions:

(1) Building the GP-Tree Structure for Storing and Indexing Image Data: GP-Tree is based on hierarchical clustering, allowing efficient storage of large image data through low-level feature vectors. The connection between image features and semantic vocabulary helps minimize data size while enhancing the speed and accuracy of the retrieval process.

(2) The Semantic Image Retrieval System GP-SBIR Based on GP-Tree: GP-SBIR combines GP-Tree with ontology, functioning through two phases: preprocessing and retrieval. The preprocessing phase constructs GP-Tree and a semi-automated ontology framework based on RDF,

while the retrieval phase performs similar image searches and semantic queries through SPARQL. The accuracy on the WANG, ImageCLEF, and MS-COCO datasets reaches 0.6078, 0.6062, and 0.717, respectively.

(3) Development of GP-Tree with Two New Models: Graph-GPTree and SgGP-Tree: Graph-GPTree improves retrieval accuracy by linking similar elements located on different branches of the tree. Meanwhile, SgGP-Tree combines with SOM to optimize clustering and select more effective leaf nodes, thereby enhancing retrieval performance. Experiments on the WANG, ImageCLEF, and MS-COCO datasets show that the accuracy of Graph-GPTree reaches 0.7665, 0.8168, and 0.8730, while SgGP-Tree achieves 0.8004, 0.8926, and 0.8753. These results confirm that the improved GP-Tree models provide high effectiveness in image retrieval.

The thesis has conducted experiments on well-known datasets such as WANG, ImageCLEF, and MS-COCO, yielding superior accuracy results compared to other modern retrieval methods. These results not only enhance the effectiveness of semantic image retrieval but also open up potential for development in this field.

2. Development Directions

In addition to the achieved results mentioned above, there are still some areas for further development:

(1) Comparison with Modern Methods: Investigating deep neural network (DNN), CNN, R-CNN, GCN-based image retrieval methods to compare with the proposed methods in the thesis.

(2) Applications in Practical Fields: Developing semantic image retrieval applications in specific fields such as identifying tourist locations from images, diagnosing diseases through medical imaging, distinguishing

types of soil rocks, and retrieving images from information on social networks.

(3) Enriching Semantic Ontology: Developing tighter semantic relationships between objects and actions in images to enhance accuracy in understanding and retrieving images.

(4) Development of Vietnamese Ontology: Expanding the construction of ontology for semantic image retrieval in Vietnamese, opening up useful applications for domestic users.

REFERENCES

[1] S. Angra and S. Ahuja, "Machine learning and its applications: A review," in 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), 2017, pp. 57-60: IEEE.

[2] D. Karimi, G. Akbarizadeh, K. Rangzan, and M. Kabolizadeh, "Effective supervised multiple-feature learning for fused radar and optical data classification," *IET Radar, Sonar Navigation,* vol. 11, no. 5, pp. 768-777, 2017.

[3] G. Duan, J. Yang, and Y. Yang, "Content-based image retrieval research," *Physics Procedia,* vol. 22, pp. 471-477, 2011.

[4] A. Alzu'bi, A. Amira, N. Ramzan, and I. Representation, "Semantic content-based image retrieval: A comprehensive study," *Journal of Visual Communication,* vol. 32, pp. 20-54, 2015.

[5] 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, vol. 2, pp. 560-565.

[6] L. M. Thanh and e. al, "Image retrieval system based on EMD similarity measure and S-tree," (in V), *Intelligent Technologies and Engineering Systems, Springer, New York, NY,* pp. 139-146, 2013.

[7] L. Piras and G. J. I. F. Giacinto, "Information fusion in content based image retrieval: A comprehensive overview," vol. 37, pp. 50-60, 2017.

[8] S. Bruch, *Foundations of Vector Retrieval*. Springer, 2024.

[9] M. A. Alzubaidi, "A new strategy for bridging the semantic gap in image retrieval," *International Journal of Computational Science Engineering,* vol. 14, no. 1, pp. 27-43, 2017.

[10] C. S. Wickramasinghe, K. Amarasinghe, and M. Manic, "Parallalizable deep self-organizing maps for image classification," in *2017 IEEE symposium series on computational intelligence (SSCI)*, 2017, pp. 1-7: IEEE.

[11] K. Kanwal, K. T. Ahmad, R. Khan, A. T. Abbasi, and J. Li, "Deep learning using symmetry, FAST scores, shape-based filtering and spatial mapping integrated with CNN for large scale image retrieval," *Symmetry,* vol. 12, no. 4, p. 612, 2020.

[12] G.-H. Liu and J.-Y. Yang, "Deep-seated features histogram: a novel image retrieval method," *Pattern Recognition,* vol. 116, p. 107926, 2021.

[13] O. Sikha and K. Soman, "Dynamic Mode Decomposition based salient edge/region features for content based image retrieval," *Multim. Tools Appl.,* vol. 80, no. 10, pp. 15937-15958, 2021.

[14] S. Dhingra and P. Bansal, "Relative examination of texture feature extraction techniques in image retrieval systems by employing neural network: an experimental review," in *Proceedings of International Conference on Artificial Intelligence and Applications: ICAIA 2020*, 2021, pp. 337-349: Springer.

[15] A. Ouni, E. Royer, M. Chevaldonné, and M. Dhome, "Leveraging semantic segmentation for hybrid image retrieval methods," *Neural Computing Applications,* vol. 34, no. 24, pp. 21519-21537, 2022.

[16] A. Yang, X. Yang, W. Wu, H. Liu, and Y. J. I. a. Zhuansun, "Research on feature extraction of tumor image based on convolutional neural network," vol. 7, pp. 24204-24213, 2019.

[17] Y. Qiang, C. Sheng, and D. Yin, "Method of tire pattern image retrieval based on wavelet transform and Siamese network," in *Proceedings of the 2020 International Conference on Aviation Safety and Information Technology*, 2020, pp. 587-592.

[18] X. Yue *et al.*, "Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 13834-13844.

[19] N. T. U. Nhi and T. M. Le, "A model of semantic-based image retrieval using C-tree and neighbor graph," *International Journal on Semantic Web Information Systems,* vol. 18, no. 1, pp. 1-23, 2022.

[20] X. Wang, D. Peng, M. Yan, and P. Hu, "Correspondence-free domain alignment for unsupervised cross-domain image retrieval," *arXiv preprint arXiv:.06081,* 2023.

[21] Y. Cao, M. Long, B. Liu, and J. Wang, "Deep cauchy hashing for hamming space retrieval," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1229-1237.

[22] Y. Xie, Y. Liu, Y. Wang, L. Gao, P. Wang, and K. Zhou, "Label-Attended Hashing for Multi-Label Image Retrieval," in *IJCAI*, 2020, pp. 955-962.

[23] W. Gu, X. Gu, J. Gu, B. Li, Z. Xiong, and W. Wang, "Adversary guided asymmetric hashing for cross-modal retrieval," in *Proceedings of the 2019 on international conference on multimedia retrieval*, 2019, pp. 159-167.

LIST OF PUBLISHED ARTICLES RELATED TO THE THESIS

1. **N. M. Hai**, T. V. Lang, and V. T. Thanh, "Semantic-Based Image Retrieval Using Hierarchical Clustering and Neighbor Graph," in World Conference on Information Systems and Technologies, 2022, pp. 34-44: Springer, DOI: https://doi.org/10.1007/978-3-031-04829-6_4 (**Scopus, Q4**)
2. **N. M. Hai**, V. T. Thanh, and T. V. Lang, "A method for semantic-based image retrieval using hierarchical clustering tree and graph," Telkomnika, vol. 20, no. 5, pp. 1026-1033, 2022, DOI: http://doi.org/10.12928/telkomnika.v20i5.24086 (**Scopus, Q3**)
3. **N. M. Hai**, T. Van Lang and T. T. Van, "Improving The Efficiency of Semantic Image Retrieval using A Combined Graph and SOM Model," in IEEE Access, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3333678 (**SCIE, Q1**)
4. **N. M. Hai**, V. T. Thanh, and T. V. Lang, "The improvements of semantic-based image retrieval using hierarchical clustering tree," FAIR’2020, 2020, pp. 557-570: Natural Science and Technology Publishing House, DOI: https://doi.org/10.15625/vap.2020.00213
5. **Nguyễn Minh Hải**, Trần Văn Lăng, Văn Thế Thành, “Một tiếp cận tìm kiếm ảnh theo ngữ nghĩa dựa trên mạng nơ-ron tích chập và ontology”, TCKHTrường ĐH Sư phạm TP. HCM, 2022. tr. 48-59. DOI: https://doi.org/10.54607/hcmue.js.19.3.3272(2022)
6. **N. M. Hai**, V. T. Thanh, and T. V. Lang, "A method of semantic-based image retrieval using graph cut," Computer Science and Cybernetics, vol. 38, no. 2, pp. 193-212, 2022, DOI: https://doi.org/10.15625/1813-9663/38/2/16786