MINISTRY OF EDUCATION AND TRAINING
VIETNAM ACADEMY OF SIENCE AND TECHNOLOGY
GRADUATE UNIVERSITY OF SIENCE AND TECHNOLOGY

……..….***…………

Cu Viet Dung

IMPROVING THE PRECISION OF CONTENT-BASED IMAGE RETRIEVAL THROUGH ON A MANIFOLD LEARNING APPROACH FROM USER FEEDBACK

Major: Computer Science
Major code: 9 48 01 01

SUMMARY OF COMPUTER DOCTORAL THESIS

Ha Noi – 2023

The thesis has been completed at: Graduate University of Science and Technology- Vietnam Academy of Science and Technology

Supervisor 1: **Assoc.Prof. Phd. Nguyen Huu Quynh**
Supervisor 2: **Assoc.Prof. Phd. Ngo Quoc Tao**

Reviewer 1:
Reviewer 2:

The thesis shall be defended in front of the Thesis Committee at Vietnam Academy Of Science And Technology - Graduate University Of Science And Technology, at .... hour......, date...... month.......year 2023

This thesis could be found at:
- The National Library of Vietnam
- The Library of Graduate University of Science and Technology

INTRODUCTION

1. The necessity of the thesis

Content-based image retrieval (CBIR) has garnered significant attention over the past decades. It presents a major challenge due to the gap between low-level features and high-level semantic concepts. To narrow this gap, relevant feedback (RF) has been introduced as a powerful tool to enhance the performance of CBIR. However, image retrieval with relevant feedback encounters several issues: (1) it tends to only explore global Euclidean structures or focus solely on the local structures of samples within the same vicinity; (2) the number of samples obtained from user feedback is often small and imbalanced between positive and negative classes; (3) it lacks consideration for various aspects of image data objects. Consequently, the precision of image retrieval methods use machine learning utilizing feedback is often suboptimal.

Therefore, propose an effective image retrieval method to solve these limitations is a necessary requirement. Hence, the thesis selects the topic "Improving the precision of Content-based image retrieval through on a manifold learning approach from user feedback"

2. Subjectives

Overall objective of the thesis: improving the precision of content-based image retrieval through on a manifold learning approach to dimensionality reduction from user feedback

Specific objectives of the thesis:

Propose a method for finding an optimal projection matrix using a manifold learning approach.

- Propose a method for automatically augmenting positive samples in the training set to solve the issue of imbalanced training data. Simultaneously, leverage different aspects of the objects to create a strong classifier

3. The main contributions of the thesis

(1) Propose an method to find optimal projection matrix finding using a manifold learning approach [CT5]. This method considers the local structures of positive and negative samples in two different neighborhoods to learn a projection that can effectively separate the data in the projected space, leading to improved accuracy for image retrieval

(2) Propose an method to automatic augmenting positive samples to solve the issue of training data imbalance [CT4]. This method can (a) add additional positive samples to the training set and (b) leverage different aspects of objects to create a strong classifier

4. The main contents of the thesis

This thesis is structured into three chapters:

Chapter 1: Introduction to Content-based image retrieval

Chapter 2: Describes the method of finding an optimized projection matrix using a manifold learning approach in image retrieval, called Semantic class discriminant projection for image retrieval (SCDPIR).

Chapter 3: Presents the method of balancing the feedback sample set and combine multi-aspect image retrieval. Finally, the thesis provides some conclusions and directions for future research.

CHAPTER 1. INTRODUCTION TO CONTENT-BASED IMAGE RETRIEVAL

1.1. Introduction to Image Retrieval

The task of a CBIR system using visual content is to automatically extract multidimensional features and find a set of images that are similar to the query image in a large database

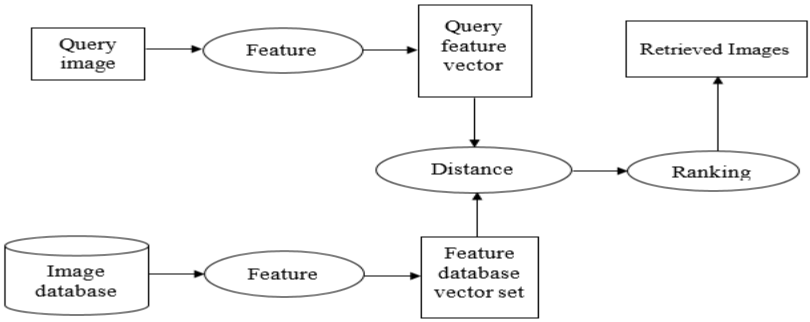


Figure 1.1. Traditional content-based image retrieval diagram

1.2. Introduction to Relevant Feedback

1.2.1. Relevant feedback

In CBIR, users are often involved in each retrieval iteration, and this mechanism is referred to as "relevant feedback" (RF).

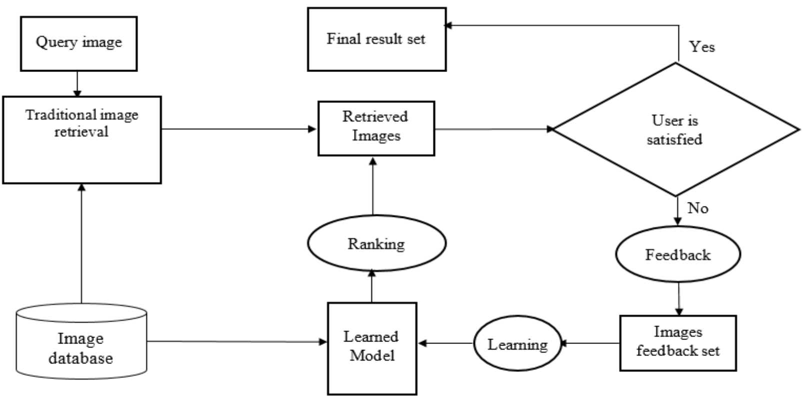


Figure 1.6. Image retrieval with relevant feedback diagram

1.2.2. Manifold learning for content-based image retrieval

The manifold learning aims to create a sub-space where related images are projected closely together, while irrelevant images are projected far apart, by learning the local structure formed by the neighborhood of the query image and the feedback images. This is achieved by embedding the query image and the set of feedback images as data points (nodes) in a k-nearest neighbor graph with weights. The optimal mapping is then determined based on the weight matrix on each edge, ensuring that neighboring points in the graph are mapped to each other by minimizing a cost function. Each image in the database is also mapped to a new embedding space, resulting in a new retrieval that is the nearest neighbor set to the query image. After each iteration, the local structure of the manifold space is relearned.

1.2.3. Review of related research

Initially, the approach to image retrieval with RF assumed the existence of an ideal query point that, if found, would provide the desired result for the user. This approach is known as "Query Point Movement" (QPM). Some image retrieval methods with RF then often rely on support vector machines (SVM) [20,42-45] to separate samples in the entire dataset according to the decision boundary. In image retrieval with relevant feedback, the images provided by users are often very small compared to the dimension of the feature, so we have to solve a problem called “curse of dimensionality”. When the dimension of the feature is too large compared to the number of samples in the training set, machine learning models can overfit. To solve this problem, some authors propose dimensionality reduction techniques such as Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA) [55]. In recent years, many manifold learning algorithms for dimensionality reduction have been proposed to explore the manifold structure. Some of these manifold methods include Locality Preserving Projections, Augmented Relation Embedding, Maximum Margin Projection, Locally Linear Embedding and Laplacian Eigenmaps. However, these methods can only be performed with data points in the training set and do not provide a clear projection that can be performed for new test data points. In addition, these methods only consider the geometric properties within a class while ignoring the relationship of samples from different classes. On the other hand, these methods often do not care about images belonging to different neighborhoods even though they may still be related to the query. Therefore, these image retrieval methods often have limited effectiveness.

1.3. Theory related to the thesis

In this section, concise overview of graph theory, distance metrics, support vector machines, Radial Basis Function (RBF), and use it as the basis for the ranking mechanism for the feedback phase in the proposed system introduced in the following chapters is presented

1.4. Evaluation of CBIR precision

1.4.1. Accuracy and average precision

To evaluate the effectiveness of CBIR systems, precision is commonly used. Precision is the ratio of the number of relevant images to the query image among the top-ranked displayed images, within a specific scope (often denoted as K), referred to as P@K.

The overall retrieval precision of a system is measured by the average of all precision. Average Precision (AP) is calculated as follows:

$$\mathbf{AP} = \frac{\sum_{i=1}^{N} precision(i)}{\mathrm{N}} \tag{1.1}$$

With $precision(i)$ representing the precision of each query and N being the total number of images used as queries.

1.4.2. Several datasets for content-based image retrieval

Dataset names	Number of topics	Number of images
COREL	80	10800
SIMPLIcity	10	1000
Oxford	11	5062
Caltech 101	101	8742

1.4.3. Automatic feedback scheme in experiment

In the real image retrieval system, a query image is often not in the image database, so the thesis uses four cross-validation to evaluate algorithms

The selection of feedback information is automatically simulated based on information from the ground truth files. For each query sent, the system retrieves and ranks images in the database. The initial result set consists of the top K ranked images, which are selected as the feedback images. Users interact with the system by marking the images in the initial retrieval result set that share the same topic (same concept) as the query image as relevant images (positive feedback samples), and the remaining images are marked as irrelevant images (negative feedback samples). Additionally, K/2 next-ranked images are selected as unlabeled samples from the initial retrieval result set

1.5. Conclusion of Chapter 1.

In Chapter 1, the thesis has presented a overview of a content-based image retrieval system and relevant feedback techniques. Additionally, it has analyzed several related feedback methods aimed at reducing semantic gaps. Through this analysis, the strengths and weaknesses of existing CBIR methods have been evaluated to propose new approaches to solve the analyzed limitations.

CHAPTER 2. SEMANTIC CLASS DISCRIMINANT PROJECTION FOR IMAGE RETRIEVAL WITH RELEVANT FEEDBACK METHOD

In this Chapter 2, the thesis propose a semantic class discriminant projection (SCDP) for image retrieval with relevant feedback method for dimensionality reduction in [CT5] to solve the limitation where the the number of dimensions feature is often much higher than the number of images in the feedback set, and the images residing in two different subspaces (two neighborhoods) have not been considered

2.1. Introduction

With image data, the number of dimensions of low-level visual feature space is often very high, from tens to hundreds, leads to the "dimensionality curse" problem. In these situations, people often use dimensional reduction techniques to map high- dimensional spaces to a lower-dimensional subspace. Processing unlabeled data, people use unsupervised methods, including principal component analysis (PCA), locality preserving projections (LPP), locally linear embedding (LLE), neighborhood preserving embedding (NPE), WeightedIso, and Supervised Isomap (S-Isomap). Many supervised dimensional reduction methods have achieved significant success including linear discriminant analysis (LDA), local discriminant embedding (LDP), supervised optimal locality preserving projection (SoLPP), marginal Fisher analysis (MFA), discriminant neighborhood embedding (DNE), linear regression classification steered discriminative projection (LRCDP), and discrimina- tive globality and locality preserving graph embedding (DGLPGE). Semi-supervised dimensional reduction methods including Therefore, several methods of reducing the dimension according to the semi-supervised approach have been proposed. Typical methods of this approach include augmented relation embedding (ARE), maximize margin projection (MMP) and semisupervised discriminant analysis (SDA). However, this method only cares about compressing and distinguishing points that belong to the same neighborhood, but ignores compression and separates other neighboring points, ie, do not guarantee relevant points that in different neighborhoods are near the query image in the lower-dimensional subspace. Besides, these methods only work with data points in the training set, and it does not explicitly give the projection possible for new test points. Therefore, these methods are not effective for image retrieval.

To overcome the above problem, thesis propose a Semantic Class Discriminant Projection (SCDP) [CT5] method. In SCDP, can honestly preserve the local structure of data points in the original multi-dimensional visual feature space and find a good projection matrix for them.

2.2. Related work

In this section, thesis present DNE, ARE, MMP and DAG-DNE, which is the basis in proposed method

2.3. Proposing a semantic class discriminant projection learning method on manifold data

The objective function

Give a set of points $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N$ in space $\mathbb{R}^n$, find a transformation matrix $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ which maps these N points to a set of $\mathbf{y}_1, \mathbf{y}_2, \dots \mathbf{y}_N$ in $\mathbb{R}^d$ $(d \ll n)$ subject to $\mathbf{y}_i$ represents$\mathbf{x}_i$, have $\mathbf{y}_i = \mathbf{U}^T\mathbf{x}_i$

Let $\mathbb{Q} \subset \mathbb{R}^n$ be an image feature space consisting of n dimensions, and $\sigma: \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ is some distance function. Given matrix $\mathbf{X} =$

$[\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N] \in \mathbb{R}^{n\times N}$ represents N mages in the image set, and N data points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_\mathrm{N}\}$ are sampled from the basic sub-manifold M. In these N points, we assume that N_1 points are labeled, and the remaining N_2 points are unlabeled, and $N_1 + N_2 = N$. To take advantage of local geometry information of M, first, we construct a low-level feature relational graph G^F. We find a k nearest neighbor of each data point $\mathbf{x}_\mathrm{i}$, then create an edge between $\mathbf{x}_\mathrm{i}$ and its neighbors get a matrix $\mathbf{W}^{\boldsymbol{F}} \in \mathbb{R}^{N\times N}$ as follows.

$$w_{ij}^F = \begin{cases} e^{-\frac{\rho^2(\mathrm{x}_i, \mathrm{x}_j)}{\tau}}, & \text{if } \mathbf{x}_i \in k - NN(\mathbf{x}_j) \\ & \text{or } \mathbf{x}_j \in k - NN(\mathbf{x}_i) \\ 0, & \text{otherwise;} \end{cases} \tag{2.1}$$

where $\rho^2(\mathbf{x}_i, \mathbf{x}_j)$ is Euclide distance (L_2), τ is a positive scalar number, and $k - NN$ is the symbol for the k nearest neighbor.

With the relevance feedback, thesis denote **IR** as the set of images returned by the system that is not related to the query image, **R** as the set of images returned by the system that is related to the query image and **UL** as the set of images without labels. To explore both the discriminative and geometric information of data manifolds, construct two graphs are the relevant similarity relation G^R and the irrelevant similarity relation G^{IR}

Weight matrices $\mathbf{W}^{\boldsymbol{R}} \in \mathbb{R}^{N\times N}$ and $\mathbf{W}^{\boldsymbol{IR}} \in \mathbb{R}^{N\times N}$ of G^R, G^{IR} respectively as follows:

$$w_{ij}^R = \begin{cases} \alpha, \text{if } (w_{ij}^F > 0) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 1, \text{if } (w_{ij}^F > 0) \wedge (\mathbf{x}_i \in \mathbf{UL} \wedge \mathbf{x}_j \in \mathbf{UL}) \\ 0, \text{otherwise} \end{cases} \tag{2.2}$$

$$w_{ij}^{IR} = \begin{cases} 1, \text{if } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{R} \ \wedge \ \mathbf{x}_j \in \mathbf{IR}) \\ \quad \text{or } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{IR} \ \wedge \ \mathbf{x}_j \in \mathbf{R}) \\ 0, \ \text{otherwise} \end{cases} \tag{2.3}$$

In Eq. (2.2), the value of α is high when the two images i and j belong to the same neighborhood and same label.

We define semantic similar information between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ through storing $\mathbf{x}_i$ and $\mathbf{x}_j$ (the two samples of $\mathbf{x}_i$ and $\mathbf{x}_j$ are not necessarily in the same neighborhood) in the matrix $\mathbf{S\_S} \in \mathbb{R}^{N \times N}$ as the equation:

$$s\_s_{ij} = \begin{cases} 1, \text{if } \mathbf{x}_i \ \in \mathrm{R} \ \wedge \ \mathbf{x}_j \ \in \mathrm{R} \\ 0, \text{otherwise} \end{cases} \tag{2.4}$$

Let **U** is a projection which map a sample $\mathbf{x}_i$ in the original space to $\mathbf{y}_i$ in a lower-dimensional space.

$$\mathbf{y}_i = \mathbf{U}^T \mathbf{x}_i \tag{2.5}$$

It is obvious that in the local neighborhood of sample $\mathbf{x}_i$, the mean of samples of the same neighborhood, and the same label may be calculated as follows:

$$\mathbf{m}_i = \textstyle\sum_j \mathbf{x}_j w_{ij}^{R} \tag{2.6}$$

On the projected space, the mean of samples of the same neighborhood and the same label may be calculated from (2.6) and (2.7)

$$\mathbf{m}_i^{(\mathbf{y})} = \textstyle\sum_j \mathbf{y}_j w_{ij}^{R} \tag{2.7}$$

Optimizing the two objective functions (2.8) and (2.9) under appropriate constraints is a criterion for selecting a good mapping.

$$\min_{\mathbf{U}} \textstyle\sum_{ij} \left(\left\| \mathbf{y}_i - \mathbf{y}_j \right\|^2 w_{ij}^{R} + \left\| \mathbf{m}_i^{(\mathbf{y})} - \mathbf{m}_j^{(\mathbf{y})} \right\|^2 s\_s_{ij} \right) \tag{2.8}$$

$$\max_{\boldsymbol{U}} \textstyle\sum_{ij} \left(\left\| \boldsymbol{y}_i - \boldsymbol{y}_j \right\|^2 w_{ij}^{IR} + \left\| \boldsymbol{m}_i^{(\boldsymbol{y})} - \boldsymbol{m}_j^{(\boldsymbol{y})} \right\|^2 (1 - s\_s_{ij}) \right) \tag{2.9}$$

The optimal projection

The problem (2.8) is rewritten as follows:

$$arg\ min_{\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}}\ trace(\mathbf{U}^T\mathbf{C}\mathbf{U}) \quad (2.10)$$

where $\mathbf{C} = \mathbf{C}_{\mathbf{x}} + \mathbf{C}_{\mathbf{m}}$

The problem (2.8) is rewritten as follows:

$$arg\ max_{\mathbf{U}^T\mathbf{U}=\mathbf{I}}\ trace(\mathbf{U}^T\mathbf{B}\mathbf{U}) \quad (2.11)$$

where $\mathbf{B} = \mathbf{B}_{\mathbf{x}} + \boldsymbol{B}_{\mathbf{m}}$

From the objective function (2.10) and (2.11), the problem of finding the projection $\mathbf{y} = \mathbf{U}^T\mathbf{x}$ will be brought to the following optimal problem:

$$\mathbf{U} = arg\ max_{\mathbf{U}} \frac{trace(\mathbf{U}^T\mathbf{B}\mathbf{U})}{trace(\mathbf{U}^T\mathbf{C}\mathbf{U})} \quad (2.12)$$

So matrix $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ is the d largest vector corresponding to the eigenvalues $\wedge = diag\ (\lambda_1, \lambda_{2,} \dots, \lambda_d)$ of the matrix $(\mathbf{C}^{-1}.\mathbf{B})$ provided that C is invertible.

Therefore, to embed a query image $\mathbf{q}^{(\mathrm{x})} \in \mathbb{Q}$, we map it to the manifold by $\mathbf{q}^{(\mathrm{y})} = \boldsymbol{U}^{\boldsymbol{T}}\mathbf{q}^{(\mathrm{x})}$. Find nearby points of $\mathbf{q}^{(\mathrm{y})}$ using Euclidean distance, and will be ranked at the top of the returned list.

Algorithm 2.1. Semantic Class Discriminant Projection (SCDP).

Input: $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N\} \in \mathbb{R}^n$ include N images with R, IR, UL ⊂X, R : images with positive label, IR: images with negative label, UL: images without label, d: number of dimensions in projection space and k, α: parameters.

Output: Projection matrix $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$

Step 1: $w_{ij}^F \leftarrow \begin{cases} e^{-\frac{\sigma^2(\mathrm{x}_i, \mathrm{x}_j)}{\tau}}, & if\ \mathbf{x}_i \in k - NN(\mathbf{x}_j) \\ & or\ \mathbf{x}_j \in k - NN(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases}$

Step 2:

$$w_{ij}^{R} \leftarrow \begin{cases} \alpha, \text{if } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 1, \text{if } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{UL} \wedge \mathbf{x}_j \in \mathbf{UL}) \\ 0, \text{otherwise} \end{cases}$$

$$w_{ij}^{IR} \leftarrow \begin{cases} 1, \text{if } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{IR}) \\ \quad \text{or } (w_{ij}^{F} > 0) \wedge (\mathbf{x}_i \in \mathbf{IR} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 0, \text{ otherwise} \end{cases}$$

$$s\_s_{ij} \leftarrow \begin{cases} 1, \text{ if } \mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R} \\ 0, \text{ otherwise} \end{cases}$$

Step 3:

$\mathbf{B} \leftarrow (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T + (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T$ with $\mathbf{x}_i, \mathbf{x}_j \in w_{ij}^{IR}$ and $\mathbf{m}_i = \sum_j \mathbf{x}_j w_{ij}^{R}$

$\mathbf{C} \leftarrow (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T + (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T$ with $\mathbf{x}_i, \mathbf{x}_j \in w_{ij}^{R}$ and $\mathbf{m}_i = \sum_j \mathbf{x}_j w_{ij}^{R}$

Step 4: **U**= $arg\ max_{\boldsymbol{U}} \frac{trace(\mathbf{U}^T\mathbf{BU})}{trace(\mathbf{U}^T\mathbf{CU})}$ với $(\mathbf{U}^T\mathbf{CU}) = \mathbf{I}$

$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ with each column as an eigenvector corresponding to the eigenvalue $\lambda_1 > \lambda_2 > \cdots > \lambda_d$.

The complexities of SCDP algorithm is $O((n + d)n^2$ where n is the number of features, d is the number of dimensions in the projected space.

2.4. Semantic class discriminant projection for image retrieval

Algorithm 2.2. Semantic class discriminant projection for image retrieval (SCDPIR).

Input: **DB**: Set of images, **q:**The initial query image, ***N:*** Number of images returned at each iteration, d: number of dimensions in projection space

Output: **S*:*** Set of result images

Step 1: X ←Retrieval-Init(**q**, **DB**, *N*>;

Step 2: Repeat

Step 2.1: $\mathbf{IR} \leftarrow \boldsymbol{Feedback}(\mathbf{X}, -1)$;

Step 2.2 $\mathbf{R} \leftarrow \boldsymbol{Feedback}(\mathbf{X}, 1)$;

Step 2.3 $\mathbf{UL} \leftarrow \mathbf{X} - (\mathbf{IR} \cup \mathbf{R})$

Step 2.4 $\mathbf{U} \leftarrow \boldsymbol{SCDP}(\mathbf{X}, \mathbf{R}, \mathbf{IR}, d, k, \alpha)$;

Step 2.5 $\mathbf{DB}^{(\mathbf{y})} \leftarrow \boldsymbol{Mapping}(\mathbf{DB}, \mathbf{U})$;

$\mathbf{q}^{(\mathbf{y})} \leftarrow \boldsymbol{Mapping}(\mathbf{q}, \mathbf{U})$

Step 2.6 $\mathbf{S} \leftarrow$Retrieval< $\mathbf{q}^{(\mathbf{y})}$, $\mathbf{DB}^{(\mathbf{y})}$, N>;

until (Người dùng dừng phản hồi);

Step 3. Return S;

The complexities algorithm is là $O(l + (\mathrm{n} + \mathrm{d})\mathrm{n}^2)$ where l is the number of images, n is the number of dimensions in the original space and d is the number of dimensions in the projected space.

2.5. Experimental results with semantic class discriminant projection for image retrieval

2.5.1. Precision image retrieval

Thesis compared our proposed image retrieval method (SCDPIR) with baseline, MMP, DSSA (the discriminative semantic subspace analysis) and DAG-DNE with parameters k=12, $\alpha = 50$.

Results of image dataset COREL

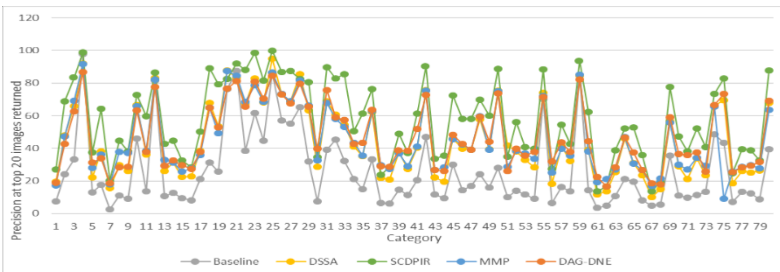
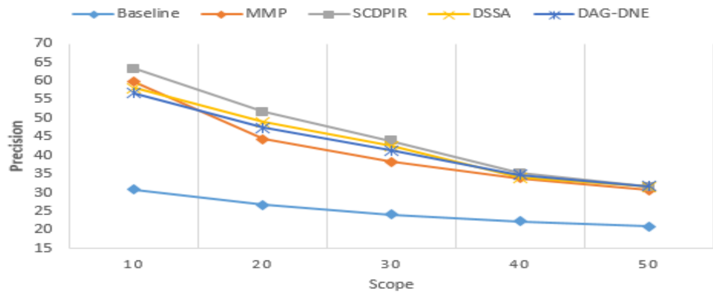
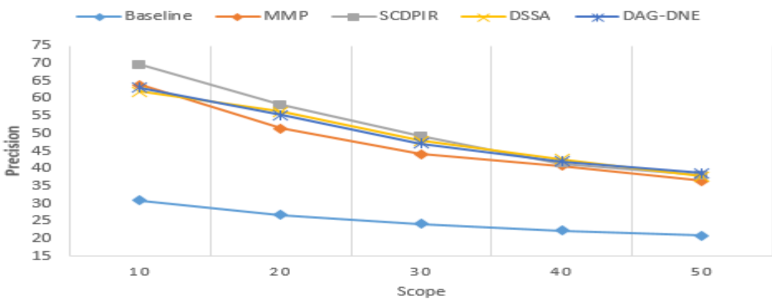


Fig. 2.8. Precision of five methods at the top 20 returned images



(a) the precision for the first feedback iteration



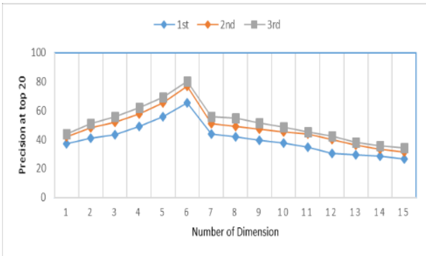
(b) the precision for the second feedback iteration

Fig 2.9. Average precision-scope curves of the different methods: (a) the precision for the first feedback iteration, (b) the precision for the second feedback iteration.

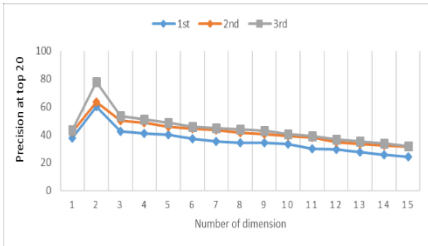
Results of image dataset COREL

With the Corel 10800 dataset, we observed a significant improvement in the performance of the proposed method. However, visualizing the projection results on the Corel dataset is not optimal due to the large number of images. Therefore, in this section, experiments are conducted on the SIMPLIcity image dataset, which consists of 1000 images, to present the visualization of the results from four methods: MMP, DSSA, DAG-DNE, and SCDPIR.

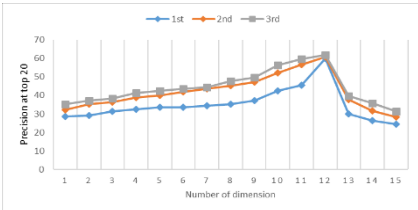
2.5.2. Dimension of semantic class discriminant projection space



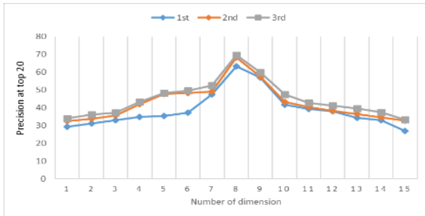
(a) Dimension of SCDP space



(b) Dimension of MMP space



(c) Dimension of DAG-DNE space



(d) Dimension of DSSAspace

Figure 2.11. The performance of four methods term of dimensions

We observe that the performance of MMP is always the best at two dimensions (Figure 2.11 (b)), the performance of SCDP is always the best at six dimensions (Figure 2.11 (a)), DSSA achieves the best performance at a very large number of dimensions which is 8 dimensions (Figure 2.11 (d)), and DAG-DNE achieves the best performance at a very large number of dimensions which is 12 dimensions (Figure 2.11 (c)). Therefore, the optimal projection dimension of SCDPIR is higher than that of MMP but lower than that of DAG-DNE and DSSA. However, the performance of SCDPIR is much higher than that of MMP when it is at a relatively low number of dimensions and this can be acceptable in practical applications. In addition, with the DAG-DNE algorithm, the best performance is achieved with a relatively high number of dimensions and it will suffer from overfitting when applied in real applications.

2.6. Conclusion of chapter 2

In this chapter, the thesis presents the SCDP (Semantic Class Discriminant Projection) method, which is capable of discovering the nonlinear structure of data in the original space to find an projection matrix. Additionally, experimental evaluations were conducted on two datasets, Corel 10K8 and SIMPLIcity, demonstrating that the proposed method has achieved improved and reliable accuracy

CHAPTER 3. BALANCED THE SET OF SAMPLE FEEDBACK AND COMBINE MULTI-ASPECT IMAGE RETRIEVAL

3.1. Introduction

Relevance feedback problems are quite different from traditional classification problems because the feedback provided by users is often limited in practical image retrieval systems. Therefore, small sample learning methods hold promise for relevant feedback. However, most existing approaches do not consider unlabeled images that are highly informative for the relevant feedback process or dimensionality reduction to improve retrieval accuracy. Besides, they overlook the balance between positive and negative samples in the feedback set

Chapter 3 presents balanced the set of sample feedback and combine multi-aspect image retrieval (CIR) method[CT4], which performs the following tasks: (a) supplementing positive samples to construct a balanced sample feedback set (BSFG - balanced sample feedback based on the graph), (b) leveraging geometric information for efficient dimensionality reduction (SCDP - discussed in Chapter 2), and (c) utilizing object aspects to build a strong classifier (CMAC).

3.2. Balanced the set of sample feedback technique using graph-based semi-supervised learning

Let nearest neighbor graph $\mathrm{G} = (\mathrm{X}, \mathrm{S})$ is also an undirected weighted graph with vertices set $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_N\} \in R^n$. These N vertices (images) are the result of the previous query

Assume the graph G is weighted, that is each edge is created by two vertices x_i and x_j carrying a nonnegative $\mathrm{s}_{ij} \geq 0$. The weighted adjacency matrix of the graph, constructed above, is the matrix $S = (s_{ij})_{i,j=1,\dots N}$.

Call $kNN(\mathrm{x}_i)$ is the k nearest neighbors of point x_i. If $\mathrm{x}_i \in kNN(\mathrm{x}_j)$ (or $\mathrm{x}_j \in kNN(\mathrm{x}_i)$), $\mathrm{s}_{ij} = 1$. Otherwise, $\mathrm{s}_{ij} = 0$. Since G is undirected, we require $\mathrm{s}_{ij} = \mathrm{s}_{ji}$.

Assume that there are m points already labeled by the user (including the original query image) $LX = \{\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_m\} \in R^n$ and $N - m$ points have not been labeled by the user $UX = \{\mathrm{x}_{N-m+1}, \mathrm{x}_{N-m+2}, \dots \mathrm{x}_{N-m}\} \in R^n$. To determine the temporary label of point x_i, where the density of the positive class around that point is high, we construct graph G^{label}.

In the graph G^{label}, its vertices are those of the graph G and its weight matrix is S^{label}. Let $label(\mathrm{x}_i)$ be the label of the point x_i (this label is either relevant or irrelevant). For each point x_i, set $kNN^{label}(\mathrm{x}_i)$ include its neighboring points that have the same class or not labeled. The reason for this is that we look at the points closest to x_i to be similar to x_i, i.e.,

$$kNN^{label}(\mathrm{x}_i) = \{\mathrm{x} |\ label(\mathrm{x}) == label(\mathrm{x}_i)\ hoặc\ \mathrm{x} \in UX\} \quad (3.1)$$

We define S^{label} the following:

$$s_{ij}{}^{label} = \begin{cases} \beta, \text{ if } label(x_i) = label(x_j) \\ 1, \text{if } x_i \in UX,\ x_i \in k-NN^{label}(x_j) \\ \quad \text{or } x_j \in UX,\ x_j \in k-NN^{label}(x_i) \\ 0, \text{ otherwise} \end{cases} \quad (3.2)$$

In Equation (3.2), the value of β is large, it implies that the two images have the same label

On the graph G^{label}, the degree of node $\mathrm{x}_i \in X$ is defined as follows:

$$d_i{}^{label} = \sum_{j=1}^{N} s_{ij}{}^{label} \quad (3.3)$$

For each point $\mathrm{x}_i \in UX$, find the point with the highest degree $d_i{}^{label}$ of the neighboring points $kNN^{label}(\mathrm{x}_i)$ and get the label of that point as the temporary label of x_i. Thus, the temporary label of x_i will be the label of x^*, where x^* defined as follows:

$$x^* = \underset{x_j \in kNN^{label}(x_i)}{\operatorname{argmax}} \left(d_j^{\,label}\right) \tag{3.4}$$

This temporary labeling process is illustrated in Figure 3.5.

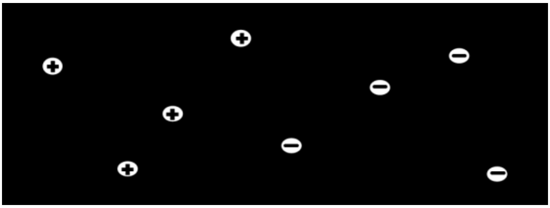


Figure 3.5. This temporary labeling process

The idea of identifying the last label of a point x_i is presented as follows. First, we divide the graph G^{label} into two classes: negative and positive class. Then, we check whether the point xi belongs to any of the two classes

Figure 3.6. The graph G^{label} is divided according to the Ncut.

Precision of BSFG

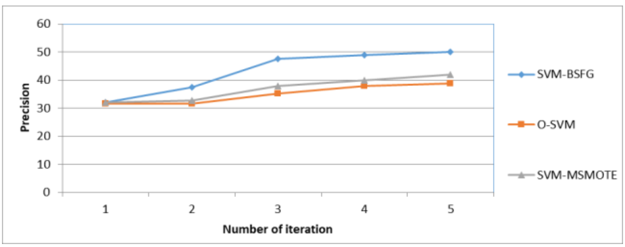


Figure 3.7. The precision of O-SVM, SVM-MSMOTE, and SVM-BSFG method.

3.3. Combine multi aspect classifiers technique

The issue of sample balancing has been solve through supervised graph-based learning. However, it has not explored the statistical properties of data classification. With the belief that there is no classifier that can represent all the useful aspects of the object. Given the various useful aspects of the object under consideration, these classifiers can be independently trained on a training set of aspects. Thesis expect that a generalized classifier, which is combined with many aspect classifiers with using the majority voting technique. In this paper, we chose five aspects, including color moment, color histogram, color correlation, gabor features and wavelet features

The problem is formulated as the Combine Multiple Aspect Classiers - CMAC.

Algorithm 3.2 Combining multiple aspect classifiers algorithm (CMAC)

Input: reduced_Aspect$_i$, $i = 1,\ldots, k$ **:** Aspect training set with reduced-dimensional

Output: β: classifiers:

Step 1: For i=1, ..., k

$$C^i \leftarrow \text{Aspect Classifiers (reduced\_Aspect}_i\text{)};$$

Step 2: $\beta(x) = \underset{y\in\{-1,\ 1\}}{\operatorname{argmax}} \sum_b \delta_{sgn\left(c^i(x)\right),y}$

Precision of CMAC

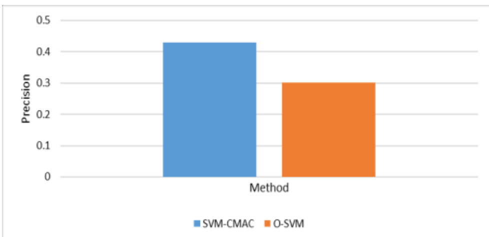


Figure 3.8. The precision of O-SVM and SVM-CMAC

3.4. Combine semantic class discriminant multiple aspect projection for image retrieval method.

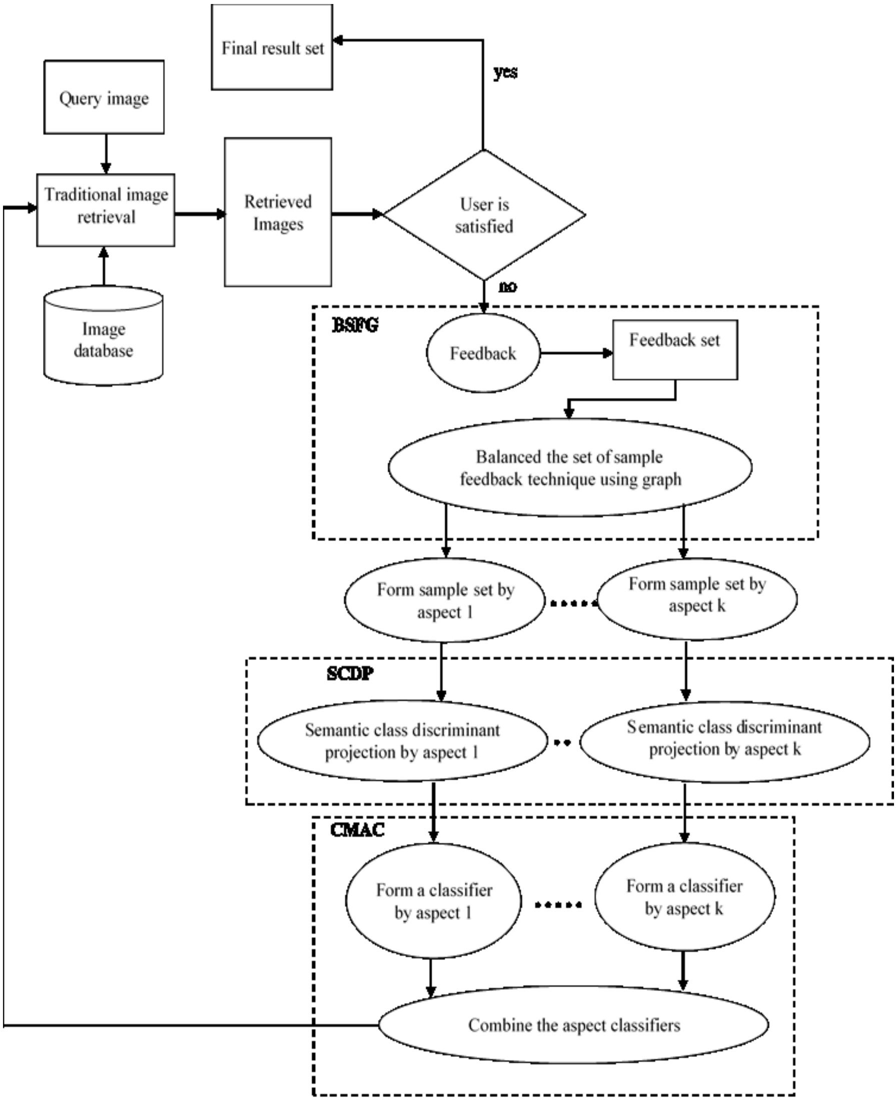


Figure 3.9. Combine semantic class discriminant multiple aspect projection for image retrieval method diagram

3.5. Evaluation of the precision of the CIR method

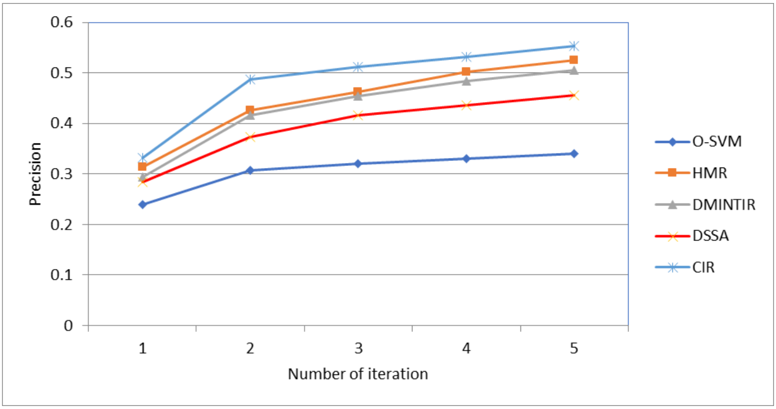


Figure 3.10. The precision of five method

3.6. Conclusion of chapter 3

In this thesis, the CIR method has been proposed to improve the precision of the RF-based retrieval system. The CIR method can: (1) utilize the information of unlabeled samples to create good training sample set;(2) exploiting the geometric properties of the data manifold for reducing the dimensionality of the feature and, and (3) utilizing the various useful aspects of objects. Experimental results on the Corel dataset have demonstrated that our proposed SEMAL significant improvement in retrieval precision.

CONCLUSIONS

The precision of content-based image retrieval systems has been an ongoing research focus in the community. Many methods have been proposed in recent years. However, the discrepancy between low-level image features and users' visual perception of image content still creates a gap between the accuracy of image retrieval systems and users' needs. The main contributions of this thesis also aim to bridge this gap by utilizing relevant feedback mechanisms

The main contributions of this thesis:

(1) Propose an method to find optimal projection matrix finding using a manifold learning approach [CT5]. This method considers the local structures of positive and negative samples in two different neighborhoods to learn a projection that can effectively separate the data in the projected space, leading to improved accuracy for image retrieval

(2) Propose an method to automatic augmenting positive samples to solve the issue of training data imbalance [CT4]. This method can (a) add additional positive samples to the training set and (b) leverage different aspects of objects to create a strong classifier.

Some future research directions include:

- Research convolutional neural networks to improve retrieval accuracy on larger image datasets.
- Research the application of hashing mechanisms to enhance retrieval speed.
- Gradually applying the system to various domains in real-life scenario.

NOVEL CONTRIBUTIONS OF THE THESIS

The novel contributions of this thesis:

(1) Propose an method to find optimal projection matrix finding using a manifold learning approach [CT5]. This method considers the local structures of positive and negative samples in two different neighborhoods to learn a projection that can effectively separate the data in the projected space, leading to improved accuracy for image retrieval

(2) Propose an method to automatic augmenting positive samples to solve the issue of training data imbalance [CT4]. This method can (a) add additional positive samples to the training set and (b) leverage different aspects of objects to create a strong classifier.

LIST OF PUBLICATIONS

[CT1] Cu Viet Dung, Nguyen Huu Quynh, An Hong Son, Dao Thi Thuy Quynh, Improved image retrieval through a combination of random subspace classifiers, *Fundamental and Applied Information Technology*, **2018,** 72- 78

[CT2] Cu Viet Dung, Nguyen Huu Quynh, Ngo Quoc Tao, Tran Thi Minh Thu, A images retrieval method base representation and manifold learning for dimensionality reduction with information from users, *Fundamental and Applied Information Technology*, **2019,** 307-314

[CT3] Cu Viet Dung, An Hong Son, Nguyen Huu Quynh, Ngo Quoc Tao, Dao Thi Thuy Quynh, The graph-based semi-supervised learning method builds a balanced set for image retrieval, *Fundamental and Applied Information Technology*, **2021,** 143-149

[CT4] Nguyen Huu Quynh, Cu Viet Dung, Dao Thi Thuy Quynh, Ngo Quoc Tao, Phuong Van Canh, Graph-based semisupervised and manifold learning for image retrieval with SVM-based relevant feedback, *Journal of Intelligent & Fuzzy Systems(SCIE,IF=1.637)*, **2019**, *37*, 711–722

[CT5] Nguyen Huu Quynh, Cu Viet Dung, Dao Thi Thuy Quynh, (2021), Semantic class discriminant projection for image retrieval with relevance feedback. *Multimedia Tools and Applications (SCIE, IF = 2.313, Q1),* **2021,** *80*, 15351–15376