Improve the Effectiveness of Image Retrieval by Combining the Optimal Distance and Linear Discriminant Analysis

Phuong Nguyen Thi Lan[1]
Thai Nguyen University – Lao Cai Campus
Lao Cai, Vietnam

Tao Ngo Quoc[2]
Institute of Information Technology
Vietnam Academy of Science and Technology
Hanoi, Viet Nam

Quynh Dao Thi Thuy[3]
Faculty of Information Technology
Posts and Telecommunications Institute of Technology
HaNoi, Viet Nam

Minh-Huong Ngo[4]
Institute of Sciences of Digital, Management and Cognition
University of Lorraine, France

***Abstract*—In image retrieval with relevant feedback, classification and distance calculation have a great influence on image retrieval accuracy. In this paper, we propose an image retrieval method, called ODLDA (Image Retrieval using the optimal distance and linear discriminant analysis). The proposed method can effectively exploit user's feedback from relevant and irrelevant image sets, which uses linear discriminant analysis to find a linear projection with an improved similarity measure. The experimental results performed on the two benchmark datasets have confirmed the superiority of the proposed method.**

Keywords*—*Content-based image retrieval; deep learning; similarity measures; Mahalanobis metric distance; linear discriminant analysis

I. Introduction

Due to the need to efficiently process huge and rapidly increasing amounts of multimedia data, content-based image retrieval (CBIR) has received a lot of attention from researchers over the past few decades. Many CBIR systems have been developed, including QBIC [21], Photobook [22], MARS [23], PicHunter [24], Blobworld [25], SIMPLIcity [26].

In a typical CBIR system, low-level visual features include color, texture, and shape, which are automatically extracted and represented as feature vectors. It should also be added that feature vectors are good if they are of the high semantic meaning of the image and serve well for image comparison. To find the desired images, the user gives a sample image and the system returns a list of similar images based on the extracted features. When the system presents a list of images that are similar to the query image, the user marks the images most relevant to the given query image to get a feedback list. The system relies on this feedback list to learn a representation or similar measure to improve the accuracy of the image retrieval.

Therefore, the representation of the image by the feature vector and the similarity measure are the two main factors that influence the efficiency of the CBIR system. Improving the effectiveness of the CBIR system is a challenging issue in research. To improve efficiency, we need to reduce semantic gaps in CBIR. The semantic gap implies the difference between the image represented by the low-level feature that is automatically extracted and the semantics of the human perceived image. To reduce this semantic gap, we need to incorporate machine learning into the image retrieval process.

Recently, there are good results due to the application of CNNs to CBIR. It has been shown that if a CNN is trained in a full surveillance context on a large set of object recognition tasks, the features extracted from the CNN can address a variety of tasks such as object image classification, scene recognition, attribute detection, and image retrieval [27,28]. Research in [29] has shown that the performance of CBIR systems using CNNs is competitive even when CNNs are trained for an unrelated classification task. To improve efficiency right from the process of building an image representation feature set, the proposed method will use CNN to build a high semantic feature set. Besides, the proposed method will incorporate similarity metrics learning techniques to have an improved similarity measure more consistent with the data.

The idea of learning similarity metrics is to find an optimal distance measure that minimizes the distance between pairs of similar images and maximizes the distance between pairs of dissimilar images. This optimal distance measurement is then used to re-rank the entire set of images and return better results. In this paper, we propose an effective image retrieval technique, called ODLDA (Image Retrieval using the optimal distance and linear discriminant analysis). The proposed method is more accurate than some state of the art methods because the feature representation is highly semantic and the similarity metrics being learned is consistent with the data. By experimenting with two databases, we will show the accuracy of the proposed method.

The remainder of the paper is organized as follows. Section 2 reviews some related studies. We present in detail the proposed method in Section 3. Section 4 describes and analyzes our experimental results. Setion 5 concludes this paper.

II. Related Work

Learning similar metrics in content-based image retrieval has received the attention of the research community [6,9,13,14,15,16,17,18]. In image retrieval with relevant feedback, the input data of distance learning algorithms are often divided into two groups: the first group consists of pairs of similar images; the second group consists of pairs of similar images and the pairs of images are not similar.

The idea of adjusting the weights of the distance function has been included in some content-based image retrieval methods such as SRIR [19]. These methods often take advantage of information from pairs of similar images and consider the scattering of the data on each dimension to construct an improved Euclidean distance function.

The MCML method [4] learns a Mahalanobis distance measure so that samples of the same class will be mapped to the same point. The distance metric learning problem is referred to as the convex optimization problem and is solved by the Gradient Descent method. However, the limitation of this method is the large computational complexity because it uses the Gradient Descent method to solve the convex optimization problem.

The idea of the LMNN [5] method is to minimize the distance of the samples of the same label in K-Nearest Neighbor and to maximize the distance of the samples that are not of the same label by a larger margin. It uses the Mahalanobis distance function. This idea is expressed as an optimization problem and solved by the SDP method [3] to find the improved distance metric.

Online Algorithm for Scalable Image Similarity learning (OASIS) [18] is specifically designed to work with pair constraints. However, they are based on strong assumptions about the input data or the structure of the constraints (requiring the input data to be sparse vectors). Therefore, it is difficult to apply in practice.

The idea of the Xing method [20] is to attribute to the convex optimization problem that minimizes the total distance of similar image pairs with the constraint that the total distance of pairs of images that are not similar reaches the maximum. In the initial phase, the method using the Euclidean distance function is improved with $A = I$. The Xing method presents an improved distance function where A is the result of the convex optimization problem. However, Xing's method has a large computational complexity due to the use of the Gradient Descent method and has not yet exploited information of similar image pairs.

The idea of the RCA method [8] is to use only similar pairs, find a data transformation based on a matrix of variance that is generated from pairs of similar images. From there it improved the Mahalanobis distance function by altering the weighting matrix. Although this method has lower computational complexity than that of the Xing method, however, the RCA method is limited to only considering the same set of images.

From analyzing the limitations of the above-related works, we propose an improved image retrieval method with an improved distance function. Improvement of the distance function which is based on maximizing the quotient between the total distance of dissimilar image pairs and the total distance of similar image pairs. Here, we look at both similar and dissimilar image sets to find the weight matrix and improve the efficiency of the retrieval method.

III. Proposed Image Retrieval Method

In this section, we will briefly present our proposed method. First, our proposed method builds deep features for representing images. Next, on the result set of the initial retrieval phase that uses deep features, the user marks up the images that are related to the query image to obtain the relevant image set (including relevant samples and samples are not irrelevant to the query image). Based on the relevant sample set, the proposed method is to train the model to find the linear projection. This linear projection satisfies the condition that the variance between samples in the same relevant set is minimized while maximizing the variance between the relevant and irrelevant samples. Besides, our proposed method also builds an improved Mahalanobis similarity metric by finding the optimal matrix M in the improved similarity metric formula.

A. Overview of the Proposed Method

A diagram of the proposed ODLDA, method is shown in Fig. 1. The method of using the CNN model has been trained on an ImageNet data set to extract the deep feature (high-level feature). When a user submits a query image, the method of extracting the deep feature of the query image is in the same way as performing an extraction with a database image. It then compares the similarity between the query image feature vector and the feature vector set of the image database which uses the Euclidean distance to return the initial result set to the user. Users conduct feedback by marking the images that are relevant and irrelevant to the query image to obtain the feedback image set. Then the feedback image set is used as input to the weight optimization and distance metric learning algorithm. Next, all images that are in the image database are re-ranked, which are based on the value of the improved Mahalanobis distance function. If the user is not satisfied with the result set, the feedback process will be repeated. If the user is satisfied, the system returns the final result set to the user.

B. Represent Image Features using Deep Learning

In recent years, CNN network has brought great results in the field of machine vision such as image classification problem, object identification, semantic segmentation. On that basis, there are many studies on content-based image retrieval using CNN and have obtained good results.

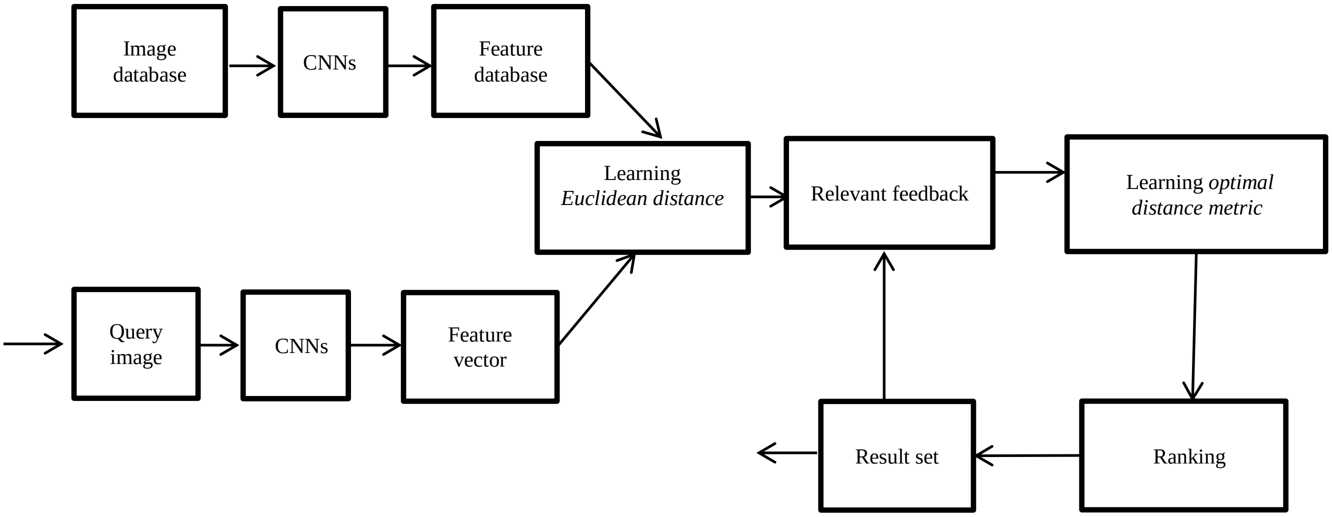

Fig. 1. Diagram of ODLDA Method.

In the document [1,2,7] has shown several approaches to improve the efficiency of a CBIR system using deep learning in building a more semantic feature set: 1) uses a pre-trained CNN model to construct an image feature set with an L_2 distance to compare the similarity measures between feature vectors; 2) it still uses the pre-trained CNN model to build the feature set, but improves it by using distance metric learning (DML) to obtain a similarity metric that is better suited to the data; 3) With a specific data set, retraining the CNN model associated with a specific classifier, then using the metric as 1) or 2) approaches is to complete a retrieval method.

Assuming we have two images in the database, I_i and I_j, the deep features are extracted using a pre-trained CNN model on the Imagenet dataset. The high-level feature of the two images I_i and I_j is denoted by x_i and x_j. The similarity metric used to compare these two features is L_2:

$$similarity(x_i, x_j) = \|x_i - x_j\|_2$$

$$= \sqrt{(x_i - x_j)^T(x_i - x_j)} \qquad (1)$$

Formula (1) shows the similarity between images I_i and I_j, the greater the similarity, the more similar images I_i and I_j are.

Similarity metric using approach 2) to compare two feature vectors of the image calculated by the formula L_T:

$$similarity(\, x_i, x_j) = \|x_i - x_j\|_T$$

$$= \sqrt{(x_i - x_j)^T T (x_i - x_j)} \qquad (2)$$

With a matrix, T obtained from learning the similarity metric which satisfies the condition T is a positive defined matrix, because the similarity metric must be positive, and the similarity metric has the smallest value when $x_i = x_j$.

The similarity metric here is that in approach 1) when the matrix T is a unit matrix $T = I$. In other words, it is a special case when we consider the correlation between the feature components in approach 1). Furthermore, each feature component has a different similarity, so it is often the similarity metric with approach 2) to get higher efficiency.

The proposed method is to build feature sets based on deep learning. After performing the K-NN procedure to obtain a list of initialization results and return them to the user, the user will mark the images that are related to the query image to obtain the feedback set. Next, it constructs an improved similarity metric by utilizing the positive sample set, which is inspired by approach 2) to construct the matrix T in the similarity metric formula (2). Matrix M is a complete matrix, which reflects the correlation of data on each feature and between features.

In the proposed method, we use a pre-trained CNN model on a very large data set. It then uses the model to extract high-level features, also known as image representation learning. The main reason we choose this approach is that a large enough data set is not available to train a CNN, Also, to train a CNN model, we will need a lot of time. CNNs are commonly used for image classification problems, in which an image is propagated across the network and the final probability is taken from the bottom layer of the network. However, in the process of learning a representation, instead of allowing the image to propagate over the entire network, we can stop the transmission at an arbitrary layer, for example, the final fully connected layer, and extracts the values from the network at this point, then uses them as feature vectors.

In the proposed method, we only use convolutional layers to extract features. The aim is to generalize a pre-trained CNN in learning the specific features of the image in the data set. The pre-trained model is used to obtain more powerful feature vectors than some algorithms such as SIFT, GIST, HOG, etc. We exploit the ability of a widely known convolutional neural network model, called ImageNet, pre-trained in ILSVRC 2012 with 1.2 million images and 1000 concepts to acquire outstanding features of the image. It consists of convolutional layers, pooling layers, and fully connected layers. The preceding layers are usually Convolutional layers combined with nonlinear activation functions and pooling layers (collectively referred to as ConvNet). The last layer is a fully-connected layer and is usually a softmax regression (see Fig. 2). The number of units in the last layer is equal to the number of layers (with ImageNet is 1000). So the output near the last layer can be considered as a useful feature vector and Softmax Regression is the classifier used. The model uses a fixed size 256 x 256 input, while the data set used in the proposed method has a variable size of images. Therefore, the images are preprocessed by converting them to 256 x 256 size. When using the network to extract the fixed feature, we cut the network at a point before the last fully connected layer. Therefore, we obtained a feature vector of 1000 dimensions for each image.

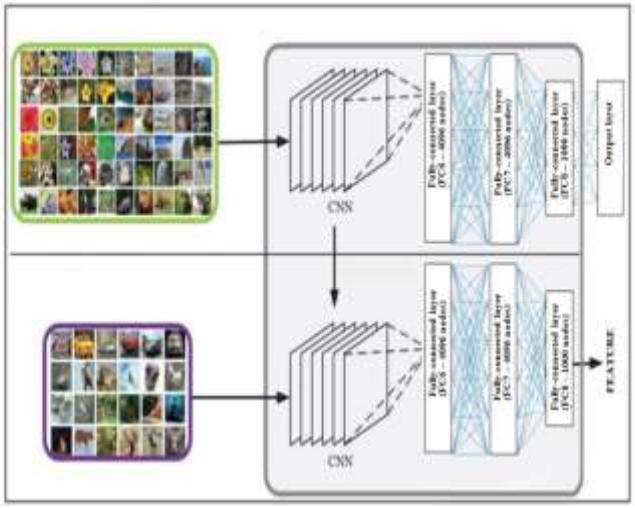
CNN
Fully-connected layer (FC6 – 4096 nodes)
Fully-connected layer (FC7 – 4096 nodes)
Fully-connected layer (FC8 – 1000 nodes)
Output layer
CNN
Fully-connected layer (FC6 – 4096 nodes)
Fully-connected layer (FC7 – 4096 nodes)
Fully-connected layer (FC8 – 1000 nodes)
FEATURE

Fig. 2. Representational Learning Architecture is based on the Pre-Training of the CNN Model.

C. An Improved Distance Metric

Up to now, there have been several different distance learning methods that exploit the properties of the user feedback set during image retrieval. However, existing methods generally consider only the positive sample set but ignore the negative sample set. The basic idea of linear discriminant analysis (LCA) is to find an optimal transformation leading to an optimal distance function, which is accomplished by maximizing the sum of variance between samples of different classes (negative or positive) and minimize the variance of data in the same class (negative or positive).

Assume that the initial resulting set consists of N images: $X = \{x_i\}_{i=1}^{N}$. The initial result set is returned to the user's feedback and is divided into two distinct sets: a positive sample set and a negative sample set. To achieve the goal, we need to define two matrices of variance, S_b and S_w. Where, S_b is the distance between the expectations of the different classes and S_w is the distance between the expectations and the samples of each class. These two matrices are calculated by the formula:

$$S_b = \frac{1}{n_b}\sum_{j=1}^{2}\sum_{i\in D_j}(m_j - m_i)(m_j - m_i)^T \tag{3}$$

$$S_w = \frac{1}{n}\sum_{j=1}^{2}\frac{1}{n_j}\sum_{i=1}^{n_j}(x_{ji} - m_i)(x_{ji} - m_i)^T \tag{4}$$

Where n_b is the total number of samples of the two sets of positive and negative samples, m_j is the center of class j, x_{ji} is the ith vector of class j, each D_j is a class. In this problem, we have 2 classes: positive class and negative class. Center m_j of class j is calculated by the formula: $m_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ji}$.

The LDA process is referred to as the optimal problem as follows:

$$T = argmax_T \frac{|T^T S_b T|}{|T^T S_w T|} \tag{5}$$

Matrix T is the optimal transformation matrix, which we need to find. When we obtain the optimal transformation T, we get the optimal weight of the Mahalanobis distance function: $M_o = T^T$T.

According to the Fisher theory [11,12], the optimization problem (5) is equivalent to maximizing the total expected distance of different classes ($\hat{C}_b$) and minimizing the total expected distance in the same class (S_w) [10]. To find the solution to the problem (5), we propose to apply algorithm 1.1 below. This algorithm is also used to solve for previous studies on LDA [22].

D. Image Retrieval Algorithm

Algorithm 1.1, called ODLDA (Image Retrieval using the optimal distance and linear discriminant analysis) describes an effective image retrieval algorithm based on the optimal distance and linear discriminant analysis.

Algorithm 1.1. ODLDA

```
Input:
    Image set: DB
    Initialization query image: Q
    Returned image number for each iteration: N
Output:
    Result set: R
1. S ←IRL<DB,M>;
2. $S_q$ ←IRL<Q,M>;
3. $Result_{Initial}$(Q)←Retrieval $_{Initial}$ $(S_q, S, N)$
4. R←$Result_{Initial}$(Q);
5. Repeat
    5.1. $< F_{feature}, F^{+}_{label}, F^{-}_{label} >$ )←Feedback $(R)$ ; relevant feedback
    5.2. $A = LDA(F_{feature}, F^{+}_{label}, F^{-}_{label})$; Find the optimal transformation T
    5.3. $M_o = T^T T$ ; The optimal weight of the Mahalanobis distance function
    5.4. $R$←Ranking $(S, M_o, N)$; Rerank the set of images according to the Mahalanobis distance function with the optimal weight
  until (User stops responding);
6. Return R;
```

The ODLDA algorithm is implemented as follows: Each image in the DB image set is represented by a feature vector in multidimensional feature space (Step 1). When the user introduces an image of the initialization query Q, the algorithm represents the query image into a feature vector S_q (Step 2). The initialization query is performed in Step 3 by $\text{Result}_{\text{Initial}}(\text{Q}) \leftarrow \textit{Retrieval}_{\textit{Initial}} \left(S_{\text{q}}, S, N\right)$, where S_q is the representation of the query image, S is the representation set of the database image set and N is the number of images to be retrieved in set S after each iteration. The retrieval result with the initialization query $\text{Result}_{\text{Initial}}(\text{Q})$ is assigned to R (Step 4).

On the $\text{Result}_{\text{Initial}}(\text{Q})$ set returned by the initialization query, the user responds through the function $Feedback(R)$ to get the feature set $F_{feature}$ and the label set $F_{Label} = \{F^{+}_{label}, F^{-}_{label}\}$ (Step 5.1). The user's feedback, including the relevant and irrelevant feedback set, is then fed into LDA (Step 5.2) to find projection A. Finding the projection A is done by solving the optimization problem (5). The results of this projection matrix were included to construct the optimal weight matrix to improve the weight of the Mahalanobis distance function (Step 5.3). At this point, we obtain the following improved Mahalanobis distance function:

$$d_M\left(F_i, F_j\right) = \left\|F_i - F_j\right\|_M = \sqrt{(F_i - F_j)^T M (F_i - F_j)}$$

The retrieval process reclassifies the entire image set in the image database by the function Ranking $(S, \text{M},)$ N, and takes N images as the result set returned to the user (Step 5.4).

IV. Experimental Results

A. *Experimental Environment*

1) Image Dataset COREL: The image set that we used for our experiment is Corel Photo Gallery with 10800 images Fig. 3. Some of the topics for this set[1] include bonsai, castle, cloud, autumn, aviation, dog, primate, ship, stalactite, fire, tiger, elephant, iceberg, train, waterfall, Each image in this set contains a prominent foreground object. Each topic consists of about 100 images. The size of the images is 120 * 80 or 80 * 120.

2) Ground truth for evaluating the precision of the CBIR: Ground truth set is used to evaluate the precision of the CBIR system, i.e., the relevant or irrelevant images identified under this set. Accordingly, the image retrieval system considers the images that are related to the query image as images with the same subject. This set consists of 3 columns (titled: Query Image ID, Image ID, and Relation) and consists of 1,981,320 rows.

3) Image Dataset SIMPLIcity: To demonstrate the performance of the proposed method, in addition to experimenting on Image Dataset COREL, we also conducted experiments on Dataset SIMPLIcity. This is a small data set with a thousand images and 10 categories. Each image in this set is 256×384 or 384×256. Some samples in this image database are shown in Fig. 4. We represent each image by two features, that is, color and edge features. The color feature is represented by the color structure descriptors with a 128-dimensional vector, while the edge feature is the edge histogram descriptors with the 150-dimensional vector. A vector of 278 dimensions, composed of two color and edge features, represents an image. The precision of the Baseline method is calculated based on the Euclidean distance between the 278-dimensional feature vector of the query image and the images in the database.

Fig. 3. Some Samples in the Corel Photo Gallery.

Fig. 4. Some Samples in the Image Dataset SIMPLIcity.

[1] https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval (Download lúc 6:32 AM ngày 25/12/2016)

B. *Execute Query and Evaluation*

In the experiment, the proposed method is compared with five image lookup methods using different distance metrics: (1) Euclidean; (2) improved Euclidean: weighted Euclidean metric of each feature dimension; (3) Xing: improved Euclidean distance function and weight matrix, which is the solution of the convex optimization problem; (4) RCA: the RCA distance metric improved from the Mahalanobis distance [8]; and (5) MCML: MCML distance metric is improved from Mahalanobis distance whose weight set is the result of data transformation with label constraints. In the experiment, our proposed method (ODLDA) performs retrieval on the deep feature set combined with the optimal Mahalanobis distance function. Results were obtained over three scopes of 50, 100, and 150. Note that the value of each scope is the top of the images returned by each retrieval loop. The reason we take these three scopes is that users often don't have the patience to choose more than 150 responses.

The average precision of the methods is shown in Table I. In this table, we find that the method using the original Euclidean metric has the lowest precision. The three methods, including Xing, RCA, and MCML, have similar precision. Our proposed method has the highest precision.

The average precision-scope curves of the Improved Euclidean, Xing's distance, RCA, MCML and ODLDA are shown in Fig. 5. These are the precision values of the top 50, 100, and 150 images after the first two iterations of feedback. In addition, in Fig. 5, we also draw the Baseline's precision for comparison purposes. According to these results, our proposed method outperforms better than the remaining methods. Thus, on two benchmark data sets, the precision of our proposed method is higher than that of the Improved Euclidean, Xing's distance, RCA, MCML and ODLDA methods. This reinforces that the idea of the proposed method is very effective.

TABLE I. COMPARISON OF AVERAGE PRECISION OF METHODS IN THE 50, 100, AND 150 SCOPES ON THE COREL DATASET

Method	**Average precision by scopes**		
	50	**100**	**150**
Euclidean	0.2887	0.3065	0.3199
Improved Euclidean	0.3135	0.42658	0.4846
Xing	0.3324	0.47658	0.5125
RCA	0.3424	0.48058	0.5015
MCML	0.3328	0.47958	0.4925
ODLDA	**0.4836**	**0.5065**	**0.5199**

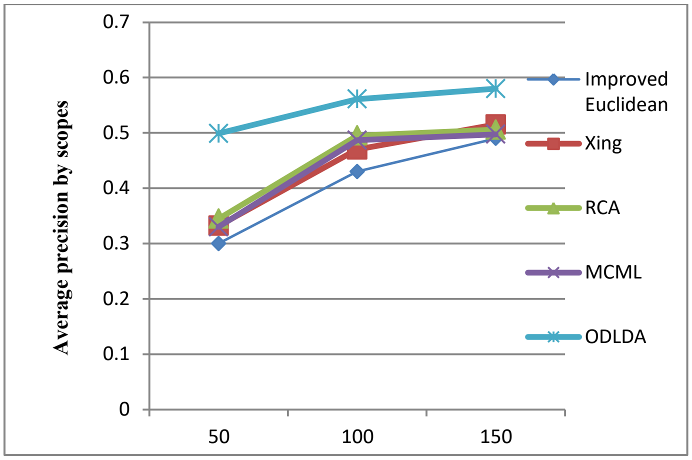

Fig. 5. Comparison of Average Precision of Methods in the 50, 100, and 150 Scopes on the SIMPLIcity Dataset.

V. CONCLUSION

This paper presents the ODLDA method, an effective image retrieval technique for improving the performance of multipoint image retrieval systems. ODLDA effectively exploits the user's information through the relevant and irrelevant sample set, which performs learning an optimal projection to separate irrelevant images and narrow the distance of related images. The proposed method finds the optimal weight matrix of the Mahalanobis distance function and uses this improved distance function to rank the entire database image set and return the result set to the user. Experimental results on two databases have proven that ODLDA provides much greater precision than the Euclidean, improved Euclidean, RCA, and OASIS methods.

ACKNOWLEDGMENT

The author gratefully acknowledges the many helpful suggestions of the anonymous reviewers during the preparation of the paper. This research was supported by the research support program "Offer for senior researchers in 2021" under grant no. NVCC02.01/21-21 and "Improve the efficiency of content-based image retrieval through metric learning" under grant no. VAST01.07/19-20.

REFERENCES

[1] Andre B, Vercauteren T, Buchner AM, Wallace MB, Ayache N (2012). Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Transactions on Medical Imaging. 31(6):1276–88.

[2] Ruigang Fu, Biao Li, Yinghui Gao, Ping Wang, (2016). Content-Based Image Retrieval Based on CNN and SVM, 2nd IEEE International Conference on Computer and Communications, 638-642.

[3] Monique Laurent, Franz Rendl, "Semidefinite Programming and Integer Programming", Report PNA-R0210, CWI, Amsterdam, April 2002.

[4] A. Globerson and S. Roweis. Metric learning by collapsing classes. Advances in Neural Information Processing Systems, 18:451, 2006.

[5] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. Advances in Neural Information Processing Systems, 18:1473, 2006.

[6] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In ICML, pages 11–18, 2003.

[7] J. Wan,D. Wang,S. C. H. Hoi, and et al, "Deep learning for content-based image retrieval: A comprehensive study," ACM International Conference on Multimedia,pp. 157-166,2014.

[8] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, Learning a Mahalanobis Metric from Equivalence Constraints, in Journal of Machine Learning Research (JMLR), 2005.

[9] C. Domeniconi, J. Peng, and D. Gunopulos. Locally adaptive metric nearest-neighbor classification. IEEE Trans. Pattern Anal. Mach. Intell., 24(9):1281–1285, 2002.

[10] Q. Liu, H. Lu, and S. Ma. Improving kernel fisher discriminant analysis for face recognition. IEEE Trans. on Circuits and Systems for Video Technology, 14(1):42–49, 2004.

[11] G. McLachlan. Discriminant Analysis and Statistical Pattern Recognition. John Wiley, 1992.

[12] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In Proc. IEEE NN for Signal Processing Workshop, pages 41–48, 1999.

[13] M. Guillaumin, J. J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In ICCV, pages 498–505, 2009.

[14] J.-E. Lee, R. Jin, and A. K. Jain. Rank-based distance metric learning: An application to image retrieval. In CVPR, 2008.

[15] A. S. Mian, Y. Hu, R. Hartley, and R. A. Owens. Image set based face recognition using self-regularized non-negative coding and adaptive distance metric learning. IEEE Transactions on Image Processing, 22(12):5252–5262, 2013.

[16] Z. Wang, Y. Hu, and L.-T. Chia. Learning image-to-class distance metric for image classification. ACM TIST, 4(2):34, 2013.

[17] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In NIPS, 2005.

[18] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. Journal of Machine Learning Research, 11:1109–1135, 2010.

[19] D. T T Quynh, N H Quynh, PV Canh, NQ Tao, An efficient semantic – Related image retrieval method, Expert Systems with Applications, Volume 72, pp. 30-41, 2017.

[20] E. Xing, A. Ng, and M. Jordan. Distance metric learning with application to clustering with side-information. In NIPS, 2002.

[21] Flickner, M., Sawhney, H., Niblack, W., et al., (1995). Query by image and video content: The QBIC system. IEEE Computer Magazine 28 (9), 23–32.

[22] A. Pentland, R. W. Picard, and S. Sclaroff (1996). Photobook: content-based manipulation for image databases.International Journal of Computer Vision, 18(3):233–254.

[23] M. Ortega-Binderberger and S. Mehrotra (2004). Relevance feedback techniques in the MARS image retrieval systems. Multimedia Systems, 9(6):535–547.

[24] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N.Yianilos (2000). The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments. IEEE Transactions on Image Processing, 9(1):20–37.

[25] C. Carson, S. Belongie, H. Greenspan, and J. Malik (2002). Blobworld: image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(8):1026–1038, 2002.

[26] J. Z. Wang, J. Li, and G. Wiederhold, (2001). “SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 23, no. 9, pp. 947-963.

[27] A. S. Razavian,H. Azizpour,1. Sullivan,and et aI,"Cnn features offthe-shelf: An astounding baseline for recognition," IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 512-519,2014.

[28] J. Donahue, Y. Jia, O. Vinyals, and et aI, "Decaf: A deep convolutional activation feature for generic visual recognition," Computer Science. vol. 50,pp. 815-830,2013.

[29] A. Babenko, A. Slesarev, A. Chigorin, and et aI, "Neural codes for image retrieval," vol. 8689,pp. 584-599,2014.