Biased Discriminant Analysis with Feature Line Embedding for Interactive Image Retrieval

Yu-Chen Wang
Dept. of Computer Science & Information Engineering,
National Central University
Taoyuan, Taiwan
m09502060@chu.edu.tw

Chin-Chuan Han*
Dept. of Computer Science & Information Engineering,
National United University
Miaoli, Taiwan
cchan@nuu.edu.tw

Chang-Hsing Lee
Dept. of Computer Science & Information Engineering,
Chung Hua University
Hsinchu, Taiwan
chlee@chu.edu.tw

Kuo-Chin Fan
Dept. of Computer Science & Information Engineering,
National Central University
Taoyuan, Taiwan
kcfan@csie.ncu.edu.tw

Abstract

The problem of content based image retrieval is to narrow down the gap between low-level image features and high-level semantic concepts. In this paper, a biased discriminant analysis with feature line embedding (FLE-BDA) is proposed for performance enhancement in the relevance feedback scheme. We try to maximize the margin between relevant and irrelevant samples at local neighborhoods. In the reduced subspace, relevant images would be closed as possible; while irrelevant samples are far away from relevant samples. The evaluation results on dataset SIMPLIcity are given to show the performance of the proposed method.

1. Introduction

The main approach in Content based image retrieval (CBIR) is to extract low-level visual features from images, and calculate the similarity degrees between them to find the most similar ones. The image contents are usually characterized by various low-level visual features, e.g., colors, shapes, textures, etc. However, it is very hard to represent human high-level semantic concept by low-level visual image features. To narrow down this gap in CBIR, extracting meaningful information from high dimensional feature spaces is proposed in [1]. Relevance feedback (RF)[2-3] also bridges the semantic gap between low-level visual features and high-level semantic concepts for performance improvement. In the past decades, RF based methods are classified into two categories: query movement and biased subspace learning (BSL). In the early stage, query movement and re-weighting techniques were developed. Relevant and irrelevant samples labeled from users give the new weights of features[2] or revise the query features to obtain a new representation[3]. In the second category, biased subspace learning methods are proposed with the observation 'all positive examples are alike; each negative example is negative in its own way. [4]' Various alternative forms are been proposed, e.g., maximum margin projection (MMP)[5], biased maximum margin analysis (BMMA)[6], and biased discriminant analysis (BDA)[7].

The biased discriminant analysis[7] which is extended from MFA[10]. Another alternative criterion is designed based on maximum margin criterion (MMC). He et al. [5] proposed a semi-supervised method extended from local preserving projection [8] for dimensionality reduction, called MMP. The MMP algorithm maximizes the margin between relevant and irrelevant samples at local neighborhoods. Zhang et al. [6] propose a BMMA and a semi-supervised BMMA (SemiBMMA) for SVM-based RF schemes. All their methods are operated on the graph embedding framework to extract the intrinsic geometry structure. They aim to the same goals: maximize the margin between relevant and irrelevant samples at local neighborhoods using the manifold-learning approach.

In this paper, we propose a FLE-BDA method, in which point to line metric is embedded into the transformation. Feature lines [9] are considered as the linear combination of image features in feature spaces. This subspace learning method discovers the intrinsic manifold structure from feedback data. The contribution of FLE-BDA method is to maximize the margin between query sample and irrelevant feature lines, and minimize the margin between query sample and relevant feature lines. In the RF scheme, users label the relevant or irrelevant samples for subspace learning. The within-class and between-class weighting graphs are constructed from users' feedbacks.

2. The proposed method

In this section, the proposed subspace learning algorithm FLE-BDA constructs the relationship of points to feature lines. Positive and negative labels are given from user's feedbacks. In the reduced subspace, the positive samples are clustered together; while the query sample is separated from the negative samples by a maximum margin.

2.1 The BDA objective function for dimensionality reduction

In [10], the geometric structures of feature lines are shown in Fig. 1. Sample points x_i, x_u, and x_v represent the relevant images from user's feedbacks. Those 'beach' images are composed of blue sky, sand, and buildings. Their features are partially overlapped. For example, the color features in images x_i and x_u are similar, and the textural features of sand in image x_i and x_v are alike. The feature line $L(x_u, x_v)$ is considered to be the approximation of two images in feature space. The projection point $P_{u,v}(x_i)$ is composed of the linear combination of feature points x_u and x_v.

Given m sample points $X = (x_1, x_2, \ldots\ldots x_m) \in \Re^{n \times m}$

collected from user's interaction, the first m^+ samples represent the positive (relevant) sample with a label $l(x_i)=1$, $1 \le i \le m^+$; the next m^- samples represent the negative (irrelevant) samples with a label $l(x_i)=-1$, $m^+ +1 \le i \le m^+ + m^-$. Similar to the approach in [5], two graphs, a within-class graph G^w and a between-class graph G^b, are constructed from the positive and negative samples, respectively. They are also called the intrinsic and penalty graphs in the graph embedding framework[6]. The discriminant vector $\vec{V} = x_i - P_{u,v}(x_i)$ is used to construct the within-class and between-class scatters. Both scatters with point to line embedding metric characterize the local geometry relationship of data in manifold structure.

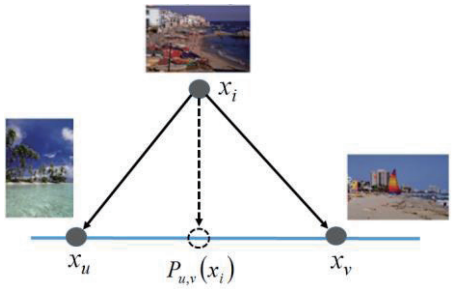


Figure 1. The geometric structure of nearest feature line.

Let m points be the corresponding points in the reduced subspace $Y=(y_1, y_2, \ldots\ldots, y_m)$. Now, the BDA with feature line embedding transforms the original space into the reduced subspace in which positive samples are clustered together, while negative samples are separated from positive sample by a maximum margin after the projection. The within-class scatter minimization and the between-class scatter maximization are both taken into account as follows:

$$S^w = \sum_i \left(\sum_{i \ne u \ne v} \left(y_i - P^w_{u,v}(y_i)\right)^2 Z^w_{u,v}(y_i) \right); \quad (1)$$

$$S^b = \sum_i \left(\sum_{i \ne u \ne v} \left(y_i - P^b_{u,v}(y_i)\right)^2 Z^b_{u,v}(y_i) \right); \quad (2)$$

$$Z^w_{u,v}(y_i) = \begin{cases} 1, \text{if } y_i \text{ connects to line } L(y_u, y_v), \\ \quad , l(y_i) = l(y_u) = l(y_v) = 1. \\ 0, \text{otherwise} \end{cases}; \quad (3)$$

$$Z^b_{u,v}(y_i) = \begin{cases} 1, \text{if } y_i \text{ connects to line } L(y_u, y_v), \\ \quad , l(y_i) = -1 \text{ and } l(y_u) = l(y_v) = 1 \\ 0, \text{otherwise} \end{cases}; \quad (4)$$

where the $P^w_{uv}(y_i)$ is the within-class projection point on the feature line $L(y_u, y_v)$ for point y_i after mapping into the new feature space. The weight $Z^w_{u,v}(y_i)$ (being 0 or 1) represents connectivity relationship from point y_i to a feature line $L(y_u, y_v)$ that passes through two points y_u and y_v. On the other hand $P^b_{uv}(y_i)$ is the between-class projection point.

In the reduced subspace, the objective function in (1) aims to minimize the distances of within-class graph to let all relevant points in the new subspace are close as possible, while the objective function in (2) attempts to maximize the distance of between-class graph to separate the query sample from irrelevant points far apart by a maximum margin.

2.2 Optimal linear embedding

In this section, the subspace learning algorithm FLE-BDA solves the objective functions (1) and (2) simultaneously. It considers the local geometric structures from points to feature lines. Let A be a projection matrix which projects vectors in the original space into a reduced subspace by $Y = A^T X$. Some simple algebraic steps are applied to reduce the objective functions in (1) and (2) as follows:

$$\begin{aligned} S^w &= \sum_i \left(\sum_{i \ne u \ne v} \left\| y_i - P^w_{u,v}(y_i) \right\|^2 Z^w_{u,v}(y_i) \right) \\ &= \sum_i \sum_{i \ne u \ne v} \left\| y_i - t_{u,v} y_u - t_{v,u} y_v \right\|^2 Z^w_{u,v}(y_i) \\ &= \sum_i \left\| y_i - \sum_j M^w_{i,j} y_j \right\|^2 = tr\left(Y \left(I - M^w\right)^T \left(I - M^w\right) Y \right) \\ &= tr\left(A^T X \left(D^w - W^w\right) X^T A \right) = tr\left(A^T X L^w X^T A \right). \end{aligned} \quad (5)$$

$$\begin{aligned} S^b &= \sum_i \left(\sum_{i \ne u \ne v} \left\| y_i - P^b_{u,v}(y_i) \right\|^2 Z^b_{u,v}(y_i) \right) \\ &= \sum_i \sum_{i \ne u \ne v} \left\| y_i - t_{u,v} y_u - t_{v,u} y_v \right\|^2 Z^b_{u,v}(y_i) \\ &= \sum_i \left\| y_i - \sum_j M^b_{i,j} y_i \right\|^2 = tr\left(Y \left(I - M^b\right)^T \left(I - M^b\right) Y \right) \\ &= tr\left(A^T X \left(D^b - W^b\right) X^T A \right) = tr\left(A^T X L^b X^T A \right). \end{aligned} \quad (6)$$

From the consequences in [10], the discriminant vector $\vec{V}$ is represented as $y_i - \sum_j M^w_{i,j} y_j$, in which two values in the ith row in matrix M^w are set as $M^w_{i,u} = t_{v,u}$, $M^w_{i,v} = t_{u,v}$, $t_{u,v} = (y_i - y_u)^T (y_u - y_v) / (y_u - y_v)^T (y_u - y_v)$, and $t_{v,u} + t_{u,v} = 1$. The other values in the ith row are set as zero. The mean squared distance in Eq. (5) for all relevant/irrelevant points to their NFLs is next obtained as $tr\left(A^T X L^w X^T A\right)$, in which $L^w = D^w - W^w$, and matrix D^w is a matrix of the column sums of the similarity matrix W^w. Matrix W^w is defined as $W^w_{i,j} = \left(M^w + \left(M^w\right)^T - \left(M^w\right)^T M^w\right)_{i,j}$ when $i \ne j$, and zero otherwise; $\sum_j M^w_{i,j} = 1$ from the consequences of Yan et al. [11]. Matrix L^w in Eq. (5) is represented as a Laplacian matrix. Similar to Eq. (5), the objective function (2) is reduced to Eq. (6). According to the constraint of Laplacian matrix, we define as follows:

$$Y^T D Y = 1 \Rightarrow A^T X D^w X^T A = 1. \quad (7)$$

Therefore, the objective function in (1) is written as:

$$\min_A 1 - A^T X W^w X^T A. \quad (8)$$

Equivalently,

$$\max_A A^T X W^w X^T A \quad (9)$$

The same as the objective function in (2) is written as:

$$\max_{A} A^T X L^b X^T A \tag{10}$$

Two objective functions in (1) and (2) are fused:

$$\arg \max_{A^T X D^w X^T A=1} A^T X\left(\alpha L^b + (1-\alpha) W^w\right) X^T A, \tag{11}$$

where α is a weighted value for the fusion of within-class and between-class graphs, and $0 < \alpha < 1$. Value α is set to be 0.05 in the experiments. To maximize (11), the solution is solved from the general eigenvalue problem. The optimal transformation matrix is obtained from the eigenvectors with the corresponding d largest eigenvalues. Every original feature vector is mapped into the reduced subspace by $x_i \to y_i = A^T x_i$.

3. The CBIR System Using Algorithm FLE-BDA

The proposed BDA with feature line embedding method is applied on the CBIR systems. Specifically, in the implementation, we have investigated five low-level features with color and texture descriptors from MPEG-7 standard are investigated for image content representation.

3.1 Low-level feature for image representation

Feature extraction is the crucial step in CBIR. Images are usually represented by low-level visual features such as color, texture, and shape. In this study, three color descriptors (color layout descriptor(CLD), color structure descriptor(CSD), and scalable color descriptor(SCD)) and two textural descriptors (homogeneous texture descriptor(HTD) and edge histogram descriptor(EHD)) are adopted for image representation. All of them are included in the MPEG-7 standard. Five low-level visual features are catenated to generate a new vector of length 718. The global and local features are both integrated in this new vector for representing the low-level semantic contents.

3.2 The relevance feedback scheme

In this study, users' feedbacks, the given labels, are used to calculate the within-class and between-class scatters. Two labels are assigned to the top rank images according to users' preferences. Feedbacks with relevant or irrelevant labels represent users' preferences, e.g. the similar textures or colors. The within-class scatter is calculated from the image samples with positive labels, while the between class scatter is calculated from those with negative labels. Based on these assigned labels, the within-class and between class weighted graphs are constructed for maximizing the margin of relevant and irrelevant samples. Let A be the optimal transformation matrix, the gallery and query samples in the low dimensional space are generated by the transformation $y_i = A^T x_i$ and $q^{new} = A^T q$, respectively. The new distance between y_i and q^{new} in the subspace is computed in (10). The RF scheme effectively finds the semantic and geometrical structures of images from users' preference. It is favorable to connect low-level image features and high-level semantic concepts using users' feedbacks.

$$\begin{aligned} dist(y_i, q) &= \sqrt{\left(y_i - q^{new}\right)^T \left(y_i - q^{new}\right)} \\ &= \sqrt{(x_i - q)^T A A^T (x_i - q)}. \end{aligned} \tag{12}$$

4. Experimental Results

In this section, some experimental results are conducted to evaluate the performance of the proposed method. The relevance feedback driven image retrieval is evaluated by a benchmark dataset SIMPLIcity[12] which was accessed from the website: http://www.wang.ist.psu.edu/docs/related. This SIMPLIcity dataset is composed of one thousand images of ten categories in 256 by 384 or 384 by 256 pixels. All images in this dataset are manually classified and labeled with the corresponding class IDs. Then, the proposed FLE-BDA method generates the optimal manifold structure from users' feedbacks. The query sample in the original space is projected into the subspace by the learned transformation in which the query sample and relevant samples are clustered together; while the query sample is separated from the irrelevant samples by a maximum margin.

4.1 Experimental design

In [5], the precision scope curves and the precision rates are adopted to be the performance indices of the CBIR algorithms. From the definition, the precision-scope is specified as the number of top-ranked images *N* which is presented to the user, and the precision rate is the ratio of the number of relevant images presented to the user in the scope *N*. The precision-scope curve represents the precision with various scopes and gives the overall performance of algorithms. On the other hand, the precision rate emphasizes the precision at a specified scope. According to the consequences in [5], the top 20 images shown on a screen are the suitable layout for IR systems. Therefore, the number of top rank images is set to be 20, i.e., *N* = 20, the number of labelled images for users. In the experiment, five-fold cross validation is used to evaluate the algorithms. At each run, one subset is selected to be the query set, and the other four subsets are used to be the gallery images for retrieval. The precision-scope curves and precision rates are calculated by averaging the results from the five-fold cross validation.

4.2 The results on dataset SIMPLIcity

In this section, three state-of-the-art subspace learning algorithms, LDA, MMP[5] and LPP[8], are compared to demonstrate the effectiveness of the proposed method. The fused features of length 718 are extracted from the images in dataset SIMPLIcity, i.e., three color features and two textural features. Since the dimension of features is very high, PCA was performed to find the best representation for avoiding the small-sample-size problem. The features are reduced to the same dimension by all the compared algorithms. In the experiment, the fea-

tures are reduced to the new vectors of length 38 by PCA for preserving more than 99% information. In addition, the final feature dimensions of algorithm FLE-BDA, LDA, MMP, and LPP are set as 4, 1, 2, and 2, respectively. The retrieval results of the compared algorithms are shown in Fig. 2. The average precision-scope curves are generated from the precision values at top 10, 20, 30, 40, and 50 retrieving images in the first feedback iteration. The results of curve 'baseline' are obtained from the retrieving results using the original features without any feedbacks. From this figure, the proposed FLE-BDA method outperforms the other algorithms LDA, MMP, and LPP. In addition, the retrieval results are sufficiently improved by relevance feedback scheme.

Furthermore, the retrieval results after various feedback iterations are also given in Fig. 3. The top 10 and 20 retrieving images after three feedback iterations are shown in Figs. 3(a) and 3(b), respectively. The retrieval results of the proposed method are better than the other algorithms LDA, MMP, and LPP.

5. Conclusions

In this paper, a novel subspace learning algorithm FLE-BDA is proposed for image retrieval. The proposed method achieves the significantly higher precision values than the other algorithms. From the experimental results, all algorithms FLE-BDA, MMP, LDA, and LPP significantly outperform the baseline algorithm. The RF scheme could enhance the retrieval results.

Acknowledgments: This work is supported by Ministry of Science and Technology, Taiwan under the grant numbers NSC 102-2221-E-239-023 and MOST 103-2221 -E-239 -017.

References

[1] Y. D. Chun, et al.: "Content-Based Image Retrieval Using Multi-Resolution Color and Texture Features," *IEEE Transaction on Multimedia*, vol. 10, pp. 1073-1084, 2008.

[2] Y. Wu, et al.: "A Feature Re-Weighting Approach for Relevance Feedback in Image Retrieval," *IEEE International Conference on Image Processing*, pp. 581-584, September 2002.

[3] Y. Rui, et al.: "Content-Based Image Retrieval with Relevance Feedback in MARS," *IEEE International Conference on Image Processing*, pp.815-818, 1997.

[4] X. Zhou, et al.: "Small Sample Learning during Multimedia Retrieval Using BiasMap," *IEEE International Conference on computer vision and pattern recognition*, pp. 11-17, 2001.

[5] X. F. He, et al.: "Learning a Maximum Margin Subspace for Image Retrieval," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 2, pp. 189-201, 2008.

[6] L. Zhang, et al.: "Semisupervised Biased Maximum Margin Analysis for Interactive Image Retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2294-2308, 2012.

[7] D. Xu, et al.: "Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content-Based Image Retrieval," *IEEE Transactions on Image Process*, vol. 16, no. 11, pp. 2811-2821, 2007.

[8] X. He, et al.: "Face Recognition Using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, Mar. 2005.

[9] S. Z. Li, et al.: "Face Recognition Using the Nearest Feature Line Method," *IEEE Transactions on Neural Network*, vol. 10, no. 2, pp. 439-443, 1999.

[10] Y. N. Chen, et al.: "Face Recognition Using Nearest Feature Space Embedding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 6, pp. 1073-1086, 2011.

[11] S. Yan, et al.: "Graph Embedding and Extensions: A Framework for Dimensionality Reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp.40-51, Jun. 2007.

[12] J. Z. Wang, et al.: "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 9, pp. 947-963, 2001.

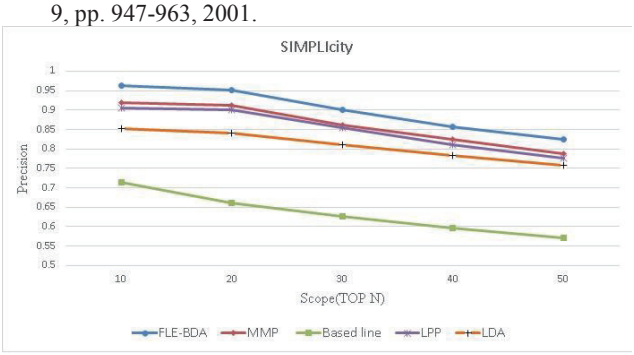
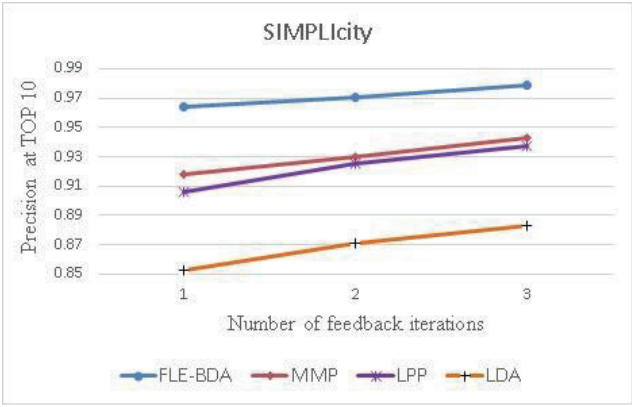
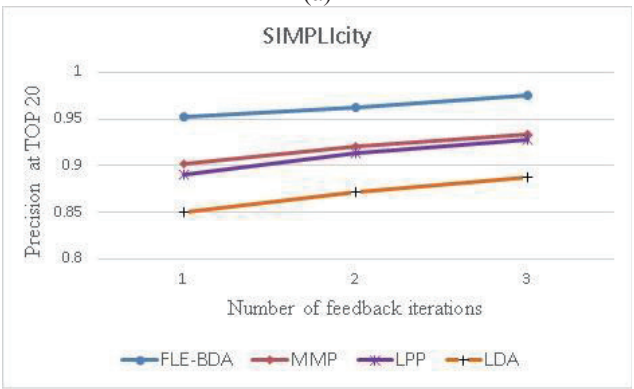


Figure 2. The precision-scope curves of various compared algorithms after the first feedback iteration.



(a)



(b)

Figure 3. The precision values at the (a) top 10, (b) top 20 retrieval results of four algorithms.