Stepwise Metric Adaptation Based on Semi-Supervised Learning for Boosting Image Retrieval Performance

Hong Chang & Dit-Yan Yeung
Department of Computer Science
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
{hongch,dyyeung}@cs.ust.hk

Abstract

For a specific set of features chosen for representing images, the performance of a content-based image retrieval (CBIR) system depends critically on the similarity measure used. Based on a recently proposed semi-supervised metric learning method called *locally linear metric adaptation* (LLMA), we propose in this paper a stepwise LLMA algorithm for boosting the retrieval performance of CBIR systems by incorporating relevance feedback from users collected over multiple query sessions. Unlike most existing metric learning methods which learn a global Mahalanobis metric, the transformation performed by LLMA is more general in that it is linear locally but nonlinear globally. Moreover, the efficiency problem is well addressed by the stepwise LLMA algorithm. We also report experimental results performed on a real-world color image database to demonstrate the effectiveness of our method.

1 Introduction

Content-based image retrieval (CBIR) has gained a lot of research interests over the last decade [15], due largely to the emergence and increased popularity of the World Wide Web. Image retrieval based on content is extremely useful in many applications. CBIR is usually performed on a query-by-example basis. The retrieval performance depends on both the features used to represent the images and the distance function used to measure the dissimilarity between the query image(s) and the images in the database. Given a specific feature representation, the retrieval performance depends critically on the distance function used to measure the dissimilarity between the query image(s) and the images in the database. Many different distance functions have been proposed for CBIR applications based on various features. However, they are not very effective in capturing the semantic (dis)similarity between images. To enhance the retrieval performance of CBIR systems, one promising direction that has aroused a great deal of research interests in recent years is to learn or adapt the distance function automatically based on images in the database.

As in traditional information retrieval, *relevance feedback* from users on the retrieval results is considered as a powerful tool to bridge the gap between low-level features and high-level semantics in CBIR systems [14]. When displayed images retrieved in response to the query image(s), the user is allowed to label some or all of the retrieved images as either relevant or irrelevant. Based on the relevance feedback, the system modifies either the query or the distance function and then carries out another retrieval attempting to improve the retrieval performance. Most existing systems only make use of relevance feedback within a single query session [14, 4, 5, 3, 16]. More recently, some methods have been proposed for the so-called *long-term learning* by accumulating relevance feedback from multiple query sessions which possibly involve different users [7, 6, 8, 11]. However, [6] and [8] are based on the assumption that the feature vectors representing the images form a Riemannian manifold in the feature space. Unfortunately this assumption may not hold in real-world image databases. Moreover, the log-based relevance feedback method [11] is expected to encounter the scale-up problem as the number of relevance feedback log sessions increases.

In machine learning, some researchers have recently proposed semi-supervised metric learning methods based on pairwise constraints, e.g., similarity or dissimilarity side-information, [17, 1, 10]. Most of these methods try to learn a global Mahalanobis metric through linear transformation. In particular, relevant component analysis (RCA) [1, 10] has been applied to enhance image retrieval performance. However, due to large variations between images in both content and style, image distribution in the feature space can be highly nonlinear. As a consequence, global metric learning is not desirable for CBIR tasks as it is not flexible enough in allowing different local metrics at different locations of the feature space.

In this paper, based on a recently proposed semi-supervised metric learning method [2], we present a new method for boosting image retrieval performance by adapting the distance metric in a stepwise manner based on relevance feedback. The metric learning method is more general as it is linear locally but nonlinear globally (Section 2). Metric adaptation is applied in a stepwise manner to make use of relevance feedback from multiple query sessions (Section 3). We perform experiments based on a real-world image database to compare our metric learning method with others for CBIR and demonstrate that continuous improvement in retrieval performance can be achieved via the stepwise learning procedure (Section 4). Finally, some concluding remarks will be given in the last section.

2 Locally Linear Metric Adaptation

Recently, Chang and Yeung [2] proposed a metric learning method called *locally linear metric adaptation* (LLMA). While the original method is based on an iterative optimization procedure, we propose here a more efficient, non-iterative version of the method.

2.1 Basic Idea of LLMA

Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of data points corresponding to n feature vectors in some d-dimensional feature space. Since the Euclidean metric in this space may not best characterize the (dis)similarity between points, the goal is to modify the metric based on pairwise constraints. As in [1], LLMA only uses pairwise similarity constraints, which can be represented in the form of a set of similar point pairs. LLMA seeks to transform the original data points to a new space so that similar points will get closer after the transformation. However, to preserve the topological relationships between points, we should not only transform the points in the similar point pairs. Instead, other points should also be affected, though to different degrees depending on their locations in the feature space.

To keep the computational demand relatively low, LLMA applies linear transformation to each local neighborhood. Since different transformations are applied to different local neighborhoods, nonlinearity can still be achieved globally. In this sense, LLMA generalizes previous metric learning methods that are based on applying linear transformation globally [1, 10, 17].

2.2 More Detailed Formulation of LLMA

Let $\mathcal{S}$ denote the set of all similar point pairs available as side information for metric learning. For each point $\mathbf{x}_r$ involved in some similar point pair, say $(\mathbf{x}_r, \mathbf{x}_s)$, a linear transformation $\mathbf{F}_r(\cdot; \mathbf{A}_r, \mathbf{b}_r)$ is applied to $\mathbf{x}_r$ as well as every data point $\mathbf{x}_i$ in the neighborhood set $\mathcal{N}_r$ of $\mathbf{x}_r$, where $\mathbf{A}_r$ and $\mathbf{b}_r$ denote the rotational and translational parameters of the transformation. Since each data point $\mathbf{x}_i$ may belong to multiple neighborhood sets corresponding to different points involved in $\mathcal{S}$, the new location $\mathbf{y}_i$ of $\mathbf{x}_i$ is the result of the combined effects of possibly all points involved in all similar pairs:

$$\mathbf{y}_i = \mathbf{x}_i + \sum_{\mathbf{x}_r : (\mathbf{x}_r, \cdot) \vee (\cdot, \mathbf{x}_r) \in \mathcal{S}} \pi_{ri} \mathbf{F}_r(\mathbf{x}_i; \mathbf{A}_r, \mathbf{b}_r), \tag{1}$$

where π_{ri} is defined as a Gaussian window function.

To estimate the parameters of each locally linear transformation $\mathbf{F}_r$, the metric adaptation problem is formulated as an optimization problem with

$$J = d_{\mathcal{S}} + \lambda P$$

as the minimization criterion, where $d_{\mathcal{S}} = \sum_{(\mathbf{x}_r, \mathbf{x}_s) \in \mathcal{S}} \|\mathbf{y}_r - \mathbf{y}_s\|^2$ is the sum of squared Euclidean distances for all similar pairs in the transformed space, P is the penalty term that constrains the degree of transformation, and λ is a regularization parameter that specifies the relative importance of the penalty term in the objective function. Different from [2], we define the penalty term to preserve the locally linear relationships between nearest neighbors, as in a nonlinear dimensionality reduction method called *locally linear embedding* (LLE) [13]. Specifically, we seek to find the best reconstruction weights for all data points, represented as an $n \times n$ weight matrix $\mathbf{W} = [w_{ij}]$, by minimizing the following cost function

$$\mathcal{E} = \sum_i \|\mathbf{x}_i - \sum_{\mathbf{x}_j \in \mathcal{N}_i} w_{ij} \mathbf{x}_j\|^2 = \mathrm{Tr}[\mathbf{X}(\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W}) \mathbf{X}^T]$$

with respect to $\mathbf{W}$ subject to the constraints $\sum_{\mathbf{x}_j \in \mathcal{N}_i} w_{ij} = 1$, where $\mathcal{N}_i$ denotes the set of K nearest neighbors of $\mathbf{x}_i$, Tr is the trace operator, and $\mathbf{X}$ is the matrix with $\mathbf{x}_i$'s being its columns. This can be solved as a constrained least squares problem. Similar to $\mathbf{X}$, let $\mathbf{Y}$ denote the matrix with $\mathbf{y}_i$'s being its columns. With the optimal weight matrix $\mathbf{W}$

found, the penalty term P is defined to ensure that points $\mathbf{y}_i$'s in the transformed space preserve the local geometry of the corresponding points $\mathbf{x}_i$'s, i.e.

$$P = \mathrm{Tr}[\mathbf{Y}(\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})\mathbf{Y}^T],$$

subject to the constraints $\frac{1}{n}\sum_i \mathbf{y}_i = \frac{1}{n}\mathbf{1}^T\mathbf{Y}^T = 0$ and $\frac{1}{n}\sum_i \mathbf{y}_i\mathbf{y}_i^T = \frac{1}{n}\mathbf{Y}\mathbf{Y}^T = \mathbf{I}_d$, where $\mathbf{1}$ represents a vector of 1's and $\mathbf{I}_d$ is the $d \times d$ identity matrix.

2.3 Optimization Based on a Spectral Approach

Equation (1) can be written as $\mathbf{Y} = \mathbf{X} + \mathbf{F\Pi} = (\mathbf{X\Pi}^+ + \mathbf{F})\mathbf{\Pi} = \mathbf{L\Pi}$, with $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \ldots)$, $\mathbf{\Pi} = [\pi_{ri}]$ and $\mathbf{\Pi}^+$ is its pseudoinverse. Thus, J can be expressed as

$$\begin{aligned} J &= \mathrm{Tr}[\mathbf{YUY}^T] + \lambda\mathrm{Tr}[\mathbf{Y}(\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})\mathbf{Y}^T] \\ &= \mathrm{Tr}[\mathbf{L\Pi}(\mathbf{U} + \lambda(\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W}))\mathbf{\Pi}^T\mathbf{L}^T], \end{aligned} \quad (2)$$

subject to constraints $\frac{1}{n}\mathbf{1}^T\mathbf{\Pi}^T\mathbf{L}^T = 0$ and $\frac{1}{n}\mathbf{L\Pi\Pi}^T\mathbf{L}^T = \mathbf{I}_d$. $\mathbf{U}$ is an $n \times n$ matrix with u_{ij} defined as $u_{ij} = u_{ji} = \tau_{ij}\sum_{r=1}^{n} s_{ir} - (1 - \tau_{ij})s_{ij}$. $\tau_{ij} = 1$ if $i = j$ and 0 otherwise, and $s_{ij} = 1$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}$ and 0 otherwise.

The solution to the optimization problem can be obtained by solving a generalized eigenvalue problem based on minimizing the criterion in Equation (2). This spectral approach is more efficient than the iterative optimization approach used in [2].

After estimating all the transformation parameters, the data points in the original space are then projected to a new space based on the locally linear transformation specified by the transformation parameters. The Euclidean metric in the transformed space thus corresponds to a modified metric in the original space to better characterize the implicit (dis)similarity relationships between data points.

3 Stepwise LLMA for Image Retrieval

The LLMA algorithm incorporates pairwise similarity constraints into metric learning. Similarity constraints can be obtained from users' relevance feedback, with each relevant image and the query image forming a similar pair.

We accumulate the similarity constraints over multiple query sessions before applying LLMA once. Experimental results show that more pairwise constraints can lead to greater improvement. However, this also implies higher computational demand. As a compromise, we perform stepwise LLMA by incorporating the pairwise constraints in reasonably small, incremental batches each of a certain size ω. Whenever the batch of newly collected pairwise constraints reaches this size, LLMA will be performed with this batch to obtain a new metric. The batch of similarity constraints is then discarded. This process will be repeated continuously with the arrival of more relevance feedback from users. In so doing, knowledge acquired from relevance feedback in one session can be best utilized to give long-term improvement in subsequent sessions. This stepwise metric adaptation algorithm is summarized in Figure 1.

```
Input: Image database $\mathcal{X}$, maximum batch size $\omega$
Begin
    Set Euclidean metric as initial distance metric
    Repeat {
        Obtain relevance feedback from new query session
        Save relevance feedback to current batch
        If batch size $= \omega$
            Adapt distance metric using LLMA
            Clear current batch of feedback information
    }
End
```

Figure 1: Stepwise LLMA algorithm for boosting image retrieval performance

4 Experimental Results

In this section, we compare the image retrieval performance of LLMA with several other distance learning methods and then apply the stepwise LLMA algorithm to further improve the retrieval performance continuously.

4.1 Image Database and Feature Representation

We perform image retrieval experiments on a subset of the Corel Photo Gallery containing 1010 images of 10 different classes. The 10 classes include bear (122), butterfly (109), cactus (58), dog(101), eagle (116), elephant (105), horse (110), penguin (76), rose (98), and tiger (115). The image classes are defined by human based on high-level semantics. We first represent the images in the HSV color space and then compute the color coherence vector (CCV) [12] as the feature vector for each image. In our experiments, we quantize each image to $8 \times 8 \times 8$ color bins, and then represent the image as a 1024-dimensional CCV, $(\alpha_1, \beta_1, \ldots, \alpha_{512}, \beta_{512})^T$, with α_i and β_i representing the numbers of coherent and non-coherent pixels, respectively, in the ith color bin. The CCV representation gives finer distinctions than the use of color histograms. Thus it usually gives better image retrieval results. For computational efficiency, we apply principal component analysis (PCA) to retain the 60 dominating principal components.

4.2 Comparative Study of Distance Learning Methods

We compare several distance learning methods for CBIR. Euclidean distance without distance learning serves as a baseline method. Besides Euclidean distance, we also repeat the experiments using distance functions learned by Xing et al's method [17], RCA [1], DistBoost [9], and LLMA.[1] Both Xing et al's method and RCA change the feature space by a globally linear transformation. DistBoost is a nonmetric distance learning algorithm by boosting the hypothesis over the product space.

Cumulative neighbor purity curves are used as performance measure in our experiments. Cumulative neighbor purity measures the percentage of correctly retrieved images

[1] The MATLAB code for RCA and DistBoost was downloaded from the web page of the authors of [1, 9].

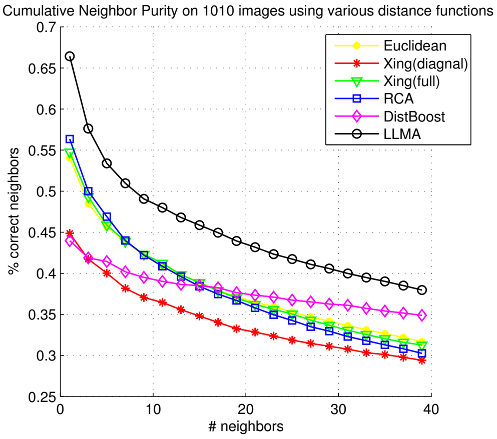


Figure 2: Retrieval results on the Corel image database.

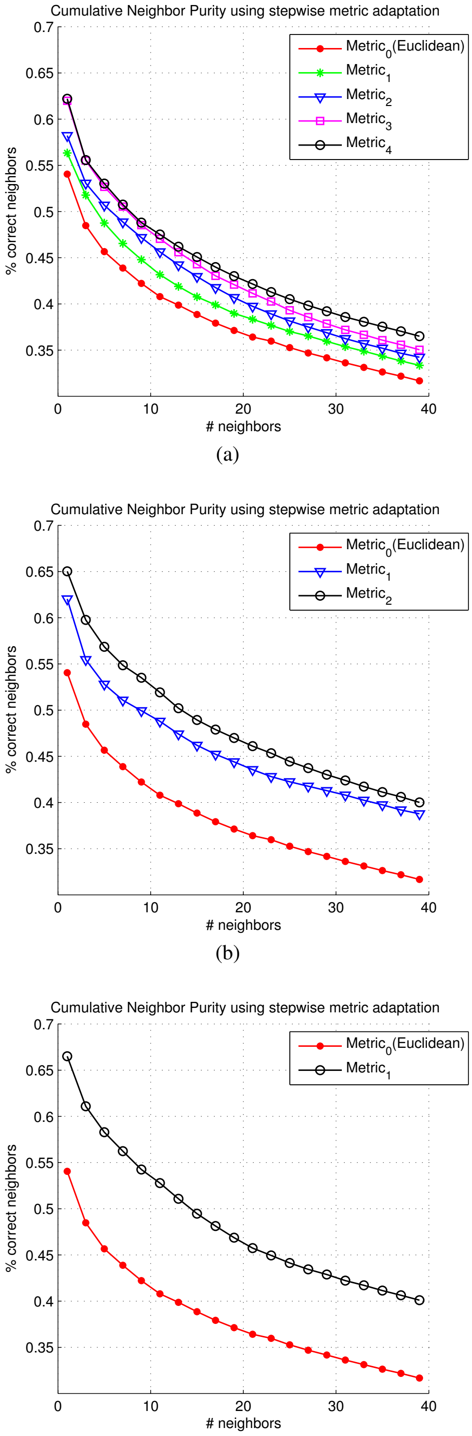
in the k nearest neighbors of the query image, averaged over all queries, with k up to some value K ($K = 40$ in our experiments). For each retrieval task, we compute the average performance statistics over 5 randomly generated $\mathcal{S}$ sets. The number of similar image pairs in $\mathcal{S}$ is set to 150, which is only about 0.3% of the total number of possible image pairs in the database.

Figure 2 shows the retrieval results using various distance functions. We can see that LLMA significantly outperforms other distance learning methods.

4.3 Experiments on Stepwise LLMA

To evaluate the stepwise LLMA algorithm described above, we devise an automatic evaluation scheme to simulate a typical CBIR system with the relevance feedback mechanism implemented. More specifically, for a prespecified maximum batch size ω, we randomly select ω images from the database as query images. In each query session based on one of the ω images, the system returns the top 20 images from the database based on the current distance function, which is Euclidean initially. Of these 20 images, five relevant images are then randomly chosen, simulating the relevance feedback process performed by a user. LLMA is performed once after every ω sessions.

Figure 3 shows the cumulative neighbor purity curves for the retrieval results on the Corel image database based on stepwise LLMA with different maximum batch sizes ω. As we can see, long-term metric learning based on stepwise LLMA can result in continuous improvement of retrieval performance. Moreover, to incorporate the same amount of relevance feedback from users, it seems more effective to use larger batch sizes. For example, after incorporating 40 query sessions from the same starting point, the final metric (metric$_4$) of Figure 3(a) is not as good as that (metric$_2$) of Figure 3(b), which in turn is (slightly) worse than that of Figure 3(c). Thus, provided that the computational resources permit, one should perform each LLMA step using relevance feedback from more query sessions.



(a)

(b)

(c)

Figure 3: Retrieval results based on stepwise LLMA with different maximum batch sizes. (a) $\omega = 10$ sessions; (b) $\omega = 20$ sessions; (c) $\omega = 40$ sessions.

5 Concluding Remarks

In this paper, we have proposed a stepwise metric adaptation method for boosting the retrieval performance of CBIR systems. Our method is based on relevance feedback from users accumulated over multiple query sessions. Experimental results on a real-world color image database demonstrate the effectiveness of the method. Our contributions can be summarized as follows. First, unlike most previous metric learning methods which learn a Mahalanobis metric corresponding to performing linear transformation globally, the transformation performed by LLMA is more general in that it is linear locally but nonlinear globally. LLMA is also more general in that it does not rely on the manifold assumption of the images in the feature space. Second, unlike most existing relevance feedback methods which only improve the retrieval results within a single query session, we propose a stepwise metric adaptation algorithm to boost the retrieval performance continuously by accumulating relevance feedback collected over multiple query sessions. Finally, the efficiency problem is well addressed. On one hand, we propose an efficient, non-iterative spectral method to solve the optimization problem in LLMA. On the other hand, the stepwise LLMA algorithm only requires temporary storage of the relevance feedback as pairwise constraints up to a user-specified maximum batch size.

Acknowledgments

The research described in this paper has been supported by two grants, CA03/04.EG01 (which is part of HKBU2/03/C) and HKUST6174/04E, from the Research Grants Council of the Hong Kong Special Administrative Region, China.

References

[1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In *Proceedings of the Twentieh International Conference on Machine Learning*, pages 11–18, 2003.

[2] H. Chang and D.Y. Yeung. Locally linear metric adaptation for semi-supervised clustering. In *Proceedings of the Twenty-First International Conference on Machine Learning*, pages 153–160, 2004.

[3] A. Dong and B. Bhanu. A new semi-supervised EM algorithm for image retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 662–667, 2003.

[4] A. Doulamis and N. Doulamis. Performance evaluation of Euclidean/correlation-based relevance feedback algorithms in content-based image retrieval systems. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 737–740, 2003.

[5] A. Doulamis, N. Doulamis, and T. Varvarigou. Efficient content-based image retrieval using fuzzy organization and optimal relevance feedback. *International Journal of Image and Graphics*, 3(1):1–38, 2003.

[6] X. He. Incremental semi-supervised subspace learning for image retrieval. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 2–8, 2004.

[7] X. He, O. King, W.Y. Ma, M. Li, and H.J. Zhang. Learning a semantic space from user's relevance feedback. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):39–48, 2003.

[8] X. He, W.Y. Ma, and H.J. Zhang. Learning an image manifold for retrieval. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 17–23, 2004.

[9] T. Hertz, A. Bar-Hillel, and D. Weinshall. Learning distance functions for image retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 570–577, 2004.

[10] T. Hertz, N. Shental, A. Bar-Hillel, and D. Weinshall. Enhancing image and video retrieval: learning via equivalence constraints. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 668–674, Madison, WI, USA, 18–20 June 2003.

[11] C.H. Hoi and M.R. Lyu. A novel log-based relevance feedback technique in content-based image retrieval. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 24–31, 2004.

[12] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *Proceedings of the Fourth ACM International Conference on Multimedia*, pages 65–73, 1996.

[13] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[14] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[15] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[16] D. Tao and X. Tang. Random sampling based SVM for relevance feedback image retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 647–652, 2004.

[17] E.P. Xing, A.Y. Ng, M.I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 505–512. MIT Press, 2003.