DISTANCE-FROM-BOUNDARY AS A METRIC FOR TEXTURE IMAGE RETRIEVAL

Guodong Guo, Hong-Jiang Zhang, and Stan Z. Li

Microsoft Research China
No. 49, Zhichun Road, Haidian District
Beijing 100080, P. R. C.

Abstract

A new metric is proposed for texture image retrieval, which is based on the signed distance of the images in the database to a boundary chosen by the query. This novel metric has three advantages: 1) the boundary distance measures are relatively insensitive to the sample distributions; 2) same retrieval results can be obtained with respect to different (but visually similar) queries; 3) retrieval performance can be improved. The boundaries are obtained by using a statistical learning algorithm called support vector machine (SVM), and hence the boundaries can be simply represented by some vectors and their combination coefficients. Experimental results on the Brodatz texture database indicate that a significantly better retrieval performance can be achieved as compared to the traditional Euclidean distance based approach. This technique can be further developed to learn pattern similarities among different texture classes and used in relevance feedback.

Keywords: *Content-based image retrieval, texture indexing, distance from boundary, support vector machines, learning similarity.*

1. INTRODUCTION

Content based image retrieval is emerging as an important research area with application to digital libraries and multimedia databases [1] [2] [3] [4]. Texture, as a primitive visual cue, has been studied for over twenty years. Various techniques have been developed for texture segmentation, classification, synthesis, and so on. Recently, texture analysis has made a significant contribution to the area of content based retrieval in large image and video databases. Using texture as a visual feature, one can query a database to retrieve similar patterns based on textural properties in the images.

In conventional texture image retrieval, the Euclidean or Mahalanobis distances between the images in the database and the query image are calculated and used for ranking. The smaller the distance, the more similar the pattern to the query. But this kind of metric has some limitations: 1). The retrieval results corresponding to different queries may be much different although they are visually similar. 2). Retrieval performance is sensitive to the sample topology. 3). Retrieval accuracy is low.

These problems can be illustrated in Fig. 1 (a), where we use class to represent the similar images. Queries "a", "b" and "c", for example, belong to class 1, but the retrieval results are much different when the Euclidean distance metric is utilized. In addition, query "a" may retrieve more patterns belong to class 3, while "b" may retrieve more patterns belong to class 2. In fact, these problems are classical and well recognized, but not much effort has been made to address these issues in the context of image database browsing. In this paper, we try to address these difficulties.

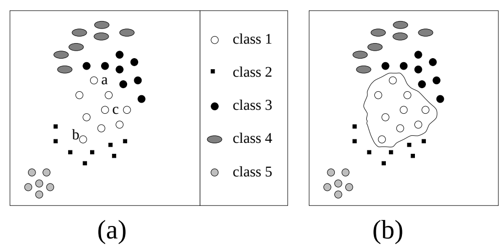


Fig. 1. (a) examples of 2-D image features belonging to five different classes. Queries a, b and c are from class 1. (b) a nonlinear boundary separates the samples of class 1 from others in classes $2 \sim 5$.

The distance metric can be termed as similarity measure, which is the key component in content-based image retrieval [5] [6]. In this paper, we propose a new metric called distance-from-boundary (DFB) to measure image similarities. The basic idea is that a (non-linear) boundary separates similar images from the remaining (dissimilar ones). This non-linear boundary encloses the same patterns inside. In Fig. 1 (b), a non-linear boundary separates patterns in class 1 from others (in classes $2 \sim 5$). The signed distances to this nonlinear boundary can be used to rank the images.

The boundary can be learned from training examples before we construct an image database. A proper learning algorithm for application in content based image retrieval should have two properties: 1) good generalization; 2) simple computation. The first one is a common requirement for any learning strategy, while the second is very important for large image database browsing. We choose to use the support vector machine (SVM) [7] to learn the boundary. The foundations of SVM have been developed by Vapnik [7], which embodies the Structural Risk Minimization (SRM) principle, and has been shown to be superior to traditional Empirical Risk Minimization (ERM) principle employed by conventional artificial neural networks [7].

In [8], a Voronoi Tessellation is used to partition the feature space, however, the non-linear boundaries are difficult to represent. Instead, they use the center to represent each cell and the retrieval is based on the nearest center criterion. On the contrary, the boundaries learned by the SVMs have better generalization property and their representation are simple — some support vectors and their combination coefficients. Our distance-from-boundary (DFB) based similarity measure has four advantages: 1) retrieval performance is relatively insensitive to the sample distributions; 2) same results can be obtained with respect to different (but visually similar) queries; 3) the DFB metric can improve the retrieval accuracy.

The paper is organized as follows. In Section 2, we describe the basic theory of SVM. In Section 3, we present the ranking scheme using the boundary distance metric. And the retrieval performance is evaluated in Section 4. Then, the DFB metric is extended to learn pattern similarities and do interactive learning in Section 5. Finally, Section 6 gives the conclusions.

2. SUPPORT VECTOR MACHINES

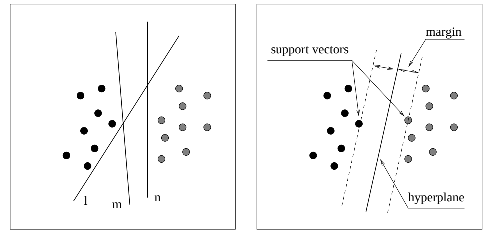
Given a set of training vectors belonging to two separate classes, $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)$, where $\mathbf{x}_i \in R^n$ and $y_i \in \{-1, +1\}$, one want to find a hyperplane $\mathbf{wx} + b = 0$ to separate the data. In Fig. 2 (a), there are many possible hyperplanes, but there is only one (shown in Fig. 2 (b)) that maximizes the margin (the distance between the hyperplane and the nearest data point of each class). This linear classifier is termed the optimal separating hyperplane (OSH).

The SVM solve the optimization problem of

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \parallel \mathbf{w} \parallel^2 - \sum_{i=1}^{l} \alpha_i \left\{ y_i \left[(\mathbf{w} \cdot \mathbf{x}_i) + b \right] - 1 \right\} \tag{1}$$

where α_i are the Lagrange multipliers. And the solution is

$$\mathbf{w}^* = \sum_{i=1}^{l} \alpha_i^* y_i \mathbf{x}_i, \quad b^* = -\frac{1}{2} \mathbf{w}^* \cdot [\mathbf{x}_r + \mathbf{x}_s] \tag{2}$$



(a) (b)

Fig. 2. Classification between two classes using hyperplanes: (a) arbitrary hyperplanes l, m and n; (b) the optimal separating hyperplane with the largest margin identified by the dashed lines, passing the two support vectors.

where $\mathbf{x}_r$ and $\mathbf{x}_s$ are any two support vectors, which satisfy $\alpha_r^*, \alpha_s^* > 0, \; y_r = 1, \; y_s = -1$.

The SVM can realize non-linear discrimination by kernel mapping [7]. In Fig. 3, the samples in the input space can not be separated by any linear hyperplane, but can be linearly separated in the non-linear mapped feature space. There are three typical kernel functions: 1) Polynomial $K(\mathbf{x}, \mathbf{y}) = ((\mathbf{x} \cdot \mathbf{y} + 1))^d$, where d is the degree of the polynomial. 2) Gaussian Radial Basis Function $K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{y})^2}{2\sigma^2}\right)$, where σ is the width of the Gaussian function. 3) Multi-Layer Perception $K(\mathbf{x}, \mathbf{y}) = \tanh\left(scale.(\mathbf{x} \cdot \mathbf{y}) - offset\right)$, with parameters $scale$ and $offset$. The GRBF performs best in our experimental comparisons, hence is chosen in our image retrieval experiments.

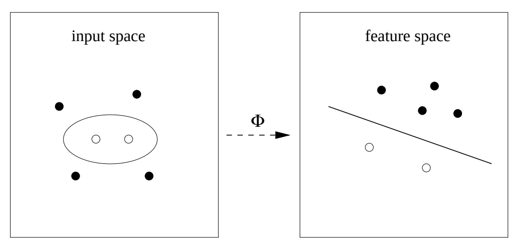


Fig. 3. The feature space is related to input space via a non-linear map $\mathbf{\Phi}$, causing the decision surface to be nonlinear in the input space.

Cortes and Vapnik [9] introduced slack variables $\xi_i \geq 0$ and a penalty function, $F(\xi) = \sum_{i=1}^{l} \xi_i$ to solve the non-separable problem. The solution is identical to the separable case except for a modification of the Lagrange multipliers with constraints $0 \leq \alpha_i \leq C, \; i = 1, \ldots, l$.

3. BOUNDARY DISTANCE METRIC AND RANKING

Recall the pair $(\mathbf{w}, b)$ defines a *separating hyperplane* or boundary of equation $\mathbf{w} \cdot \mathbf{x} + b = 0$. In kernel mapping, the boundary $(\beta^*, \mathbf{x}^*, b^*, K)$ equation is,

$$\sum_{j=1}^{m} \beta_j^* K(\mathbf{x}_j^*, \mathbf{x}) + b^* = 0 \tag{3}$$

where $\mathbf{x}_j^*(j = 1, \cdots, m; m \leq l)$ are support vectors, $\beta_j^* = \alpha_j^* y_j$ are the combination coefficients or weights, b^* is a constant, and $K(\cdot, \cdot)$ is the kernel function. Thus we have

Definition 1 (signed distance with kernel):

The signed distance $D(\mathbf{x}; \beta^*, \mathbf{x}^*, b^*, K)$ from point $\mathbf{x}$ to the boundary $(\beta^*, \mathbf{x}^*, b^*, K)$ with kernel function $K(\cdot, \cdot)$ is defined as

$$D(\mathbf{x}; \beta^*, \mathbf{x}^*, b^*, K) = \frac{\sum_{j=1}^{m} \beta_j^* K\left(\mathbf{x}_j^*, \mathbf{x}\right) + b^*}{\| \sum_{j=1}^{m} \beta_j^* \mathbf{x}_j^* \|} \tag{4}$$

In the case of c groups of similar images, we need c boundaries.

Definition 3 (signed distance to the k^{th} boundary):

If the boundary $(\beta_k^*, \mathbf{x}_k^*, b_k^*, K)$ separates class k from others, the signed distance of pattern $\mathbf{x}$ to this boundary is computed by

$$D(\mathbf{x}; \beta_k^*, \mathbf{x}_k^*, b_k^*, K) = \frac{\sum_{j=1}^{k_m} \beta_{kj}^* K\left(\mathbf{x}_{kj}^*, \mathbf{x}\right) + b_k^*}{\| \sum_{j=1}^{k_m} \beta_{kj}^* \mathbf{x}_{kj}^* \|} \tag{5}$$

where $\mathbf{x}_{kj}^*(j = 1, \cdots, k_m)$, are the support vectors to construct the k^{th} boundary, and β_{kj}^* are the optimal coefficients, and b_k^* are some constants, $k = 1, \cdots, c$.

In our boundary distance measure, the patterns within the same class have positive distances to their enclosing boundary, while other patterns have negative distance to this boundary.

Definition 3 (boundary distance ranking):

The signed distance of all images to the k^{th} boundary $D(\mathbf{x}_i; \beta_k^*, \mathbf{x}_k^*, b_k^*, K)$ are calculated and sorted in decreasing order, thus to rank the patterns $\mathbf{x}_i$, $i = 1, \cdots, N$ in the database with respect to the k^{th} boundary.

In retrieval, when a query image pattern $\mathbf{q}$ is given, a boundary index k^* is first found by

$$k^* = \arg \max_{1 \leq k \leq c} D(\mathbf{q}; \beta_k^*, \mathbf{x}_k^*, b_k^*, K) \tag{6}$$

Then, equation (5) is used to calculate the signed distances of all images to the k^{*th} boundary. According to Definition 3, the signed distance values are used to rank the images in the database with respect to the query.

4. RETRIEVAL PERFORMANCE

The Brodatz texture database [10] with 112 texture classes is used to evaluate the DFB metric for texture image retrieval. Each of the 512×512 images is divided into 49 overlapping sub-images of 128×128 pixels, centered on a 7×7 grid over the original image. The first 33 sub-images are used as the training set and the last 16 for retrieval. This kind division is similar to [11]. Thus a database of 3696 texture images is formed for learning, and another 1792 texture images for testing the retrieval performance.

Texture features are calculated by using the Gabor filter banks as in [4], with four scales and six orientations. Applying these Gabor filters to an image results in 24 filtered images. The mean and standard deviation of each filtered image are calculated and taken as a feature vector

$$\bar{f} = [\mu_{00}, \mu_{01}, \cdots, \mu_{35}, \sigma_{00}, \cdots, \sigma_{35}] \tag{7}$$

where the subscripts represent the scale $(0, \cdots, 3)$ and orientation $(0, \cdots, 5)$. The feature vector dimension is 48.

Before retrieval, 112 nonlinear boundaries are learned by the SVMs using the training set. The GRBF kernel is used with $\sigma = 0.3$ and $C = 200$. When a query is given, one boundary index is first found by (6). The distance values of all patterns in the database to this boundary are calculated and sorted in decreasing order. In the ideal case, all the top 15 retrieved are from the same original image. To measure the retrieval performance, we calculate the average retrieval accuracy [4] defined as the average percentage number of patterns belonging to the same image as the query in the top 15 matches.

On average, 87.61% of the correct patterns are in the top 15 retrieved images. For comparison, the retrieval accuracy obtained by using the Euclidean distance measure [4] is also provided, which is 79.37% and lower than the boundary distance measure. The retrieval performance increases to 96.70% if the top 100 (about 6% of the entire database) retrievals are considered. The comparison of the two measures with respect to the number of top matches are shown in Fig. 4. It is obvious that the performance of DFB metric is consistently better than the Euclidean distance based approach [4].

5. OTHER RELATED ISSUES

In previous Section, we evaluate the DFB metric for texture image retrieval. The goal is to retrieve the same class images on the top matches. However, there are many visually similar textures (but usually in different classes) can not be retrieved even on the top 100 matches. Ma and Manjunath [11] discussed this problem and used the Learning Vector Quantization to learn similarity. Our DFB metric can be extended naturally to solve this problem. The basic idea is

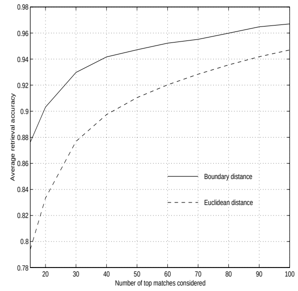


Fig. 4. The retrieval performance comparison between the Euclidean distance and the boundary distance measures.

to partition the original feature space into clusters of visually similar patterns. For Brodatz texture database, the 112 classes can be grouped into 32 perceptual similar clusters. See [12] for detailed discussion and results based on the boundary distance metric.

The DFB metric can also be used for relevance feedback [13], where the positive and negative examples submitted by the user's feedback are used to learn and refine the boundary corresponding to the query. The database consists of about 3100 natural images, part from the Corel photos and others collected from the internet. We manually classified the images into 72 classes. The number of images in each class is not equal and varies from 20 to 10. The Gabor filters are also used to extract the features in R, G, and B channels separately, and the resulted feature dimension is 144.

Some preliminary results are shown in Fig. 5. After 15 interactions, the retrieval accuracy can reach 90%. Further work is to compare the performance of the DFB based relevance feedback with some classic approaches, such as [13], and also to try a much larger database with more than 20,000 images, which is ongoing.

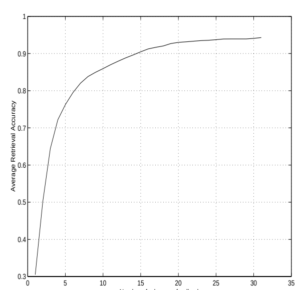


Fig. 5. Retrieval performance vs. the number of relevance feedback.

6. CONCLUSIONS

We have presented a new metric called distance-from-boundary (DFB) for texture image retrieval. The boundaries can be learned effectively by the support vector machines (SVMs) and the boundary can be simply represented. The retrieval performance is significantly improved. Further more, The boundary distance metric can be developed to learn pattern similarities and do interactive learning.

7. REFERENCES

[1] W. Niblack and *et al*, "The qbic project: querying images by content using color, texture, and shape," *Proc. of SPIE, Storage and Retrieval for Image and Video Databases-Π*, vol. 1908, pp. 173–187, Feb. 1993.

[2] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: tools for content based manipulation of image databases," *Proc. of SPIE, Storage and Retrieval for Image and Video Databases-Π*, , no. 2185, pp. 34–47, Feb. 1994.

[3] T. P. Minka and R. W. Picard, "Interactive learning using a society of models," *Technical Report*, , no. 349, 1995.

[4] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE PAMI*, vol. 18, no. 8, pp. 837–842, 1996.

[5] S. Santini and R. Jain, "Similarity measures," *IEEE PAMI*, vol. 21, no. 9, pp. 871–883, 1999.

[6] J. Puzicha, J. M. Buhmann, Y. Rubner, and C. Tomasi, "Emperical evaluation of dissimilarity measures for color and texture," *Proc. ICCV*, pp. 1165–1172, 1999.

[7] V. N. Vapnik, *Statistical learning theory*, John Wiley & Sons, New York, 1998.

[8] D. L. Swets and J. Weng, "Hierarchical discriminant analysis for image retrieval," *IEEE PAMI*, vol. 21, no. 5, pp. 386–401, 1999.

[9] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[10] P. Brodatz, *Textures: A Photographic Album for Artists & Designers*, New York: Dover, 1966.

[11] W. Y. Ma and B. S. Manjunath, "Texture features and learning similarity," *Proc. CVPR*, pp. 425–430, 1996.

[12] G. D. Guo, S. Z. Li, and K. L. Chan, "Learning similarity for texture image retrieval," *Proc. ECCV*, pp. 178–190, 2000.

[13] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, Sept. 1998.