Asymmetric Learning and Dissimilarity Spaces for Content-based Retrieval

Bruno, Eric; Moenne-Loccoz, Nicolas; Marchand-Maillet, Stéphane

How to cite

BRUNO, Eric, MOENNE-LOCCOZ, Nicolas, MARCHAND-MAILLET, Stéphane. Asymmetric Learning and Dissimilarity Spaces for Content-based Retrieval. In: Image and Video Retrieval : 5th International Conference, CIVR 2006. Proceedings. Tempe (AZ). [s.l.] : Springer, 2006. p. 330–339. (Lecture Notes in Computer Science) doi: 10.1007/11788034_34

This publication URL: https://archive-ouverte.unige.ch//unige:47730

Publication DOI: 10.1007/11788034_34

Asymmetric Learning and Dissimilarity Spaces for Content-based Retrieval

Eric Bruno, Nicolas Moenne-Loccoz, Stéphane Marchand-Maillet ⋆

Viper group
Computer Vision and Multimedia Laboratory, University of Geneva
eric.bruno@unige.ch

Abstract. This paper presents novel dissimilarity space specially designed for interactive multimedia retrieval. By providing queries made of positive and negative examples, the goal consists in learning the positive class distribution. This classification problem is known to be asymmetric, *i.e.* the negative class does not cluster in the original feature spaces. We introduce here the idea of Query-based Dissimilarity Space (QDS) which enables to cope with the asymmetrical setup by converting it in a more classical 2-class problem. The proposed approach is evaluated on both artificial data and real image database, and compared with state-of-the-art algorithms.

1 Introduction

Determining semantic concepts by allowing users to iteratively and interactively refine their queries is a key issue in multimedia content-based retrieval. The relevance feedback loop allows to build complex queries made out of positive and negative documents as examples. From this training set, a learning process has to create a model of the sought concept from a set of data features so as to provide relevant documents to the user. The success of this search strategy relies mainly on the representation spaces where data is embedded as well as on the learning machine operating in those spaces.

Various aspects of these problems have been studied with success for the last few years. This includes works on machine learning strategies such as active learning [3], imbalance classification algorithms [13], automatic kernel setting [12] or automatic labelling of training data [10]. All these studies have in common to consider feature spaces to represent knowledge on the multimedia content.

An alternative solution is to represent documents according to their similarities (related to one or several features) to the other documents rather than to a feature vector. Considering a collection of documents, the similarity-based representation, stored in (dis)similarity matrices or some distance-based indexing structures [4], characterizes the content of an element of the collection relatively to a part of or the whole collection. Recent studies have been published for

⋆ This work is funded by the Swiss NCCR (IM)2 (Interactive Multimodal Information Management).

document retrieval and collection browsing by using pre-computed similarities ([1], [7]) The idea is to index elements relatively to their closest neighbours, *i.e.* those who have the best probabilities to belong to the same class providing then a sparse association graph structuring the multimedia collection and allowing fast retrieval of data. As pointed out by authors, the similarity approach provides a convenient way for multimodal data fusion, since adding new features simply consists in adding new distances to the same representation framework. It is also noted that the off-line computation of similarities enables fast accesses and scalable content-based multimedia retrieval systems.

In [2], we proposed a similarity-based representation that goes further the nearest-neighbour model by allowing non-linear mapping of the low-level distance measures to the high-level concept space. Based on Dissimilarity Spaces (DS) introduced by Pekalska *et al* [8], we have defined representation spaces adapted to the *query-by-example* paradigm. These Query-based Dissimilarity Spaces (QDS) have the advantages to be of low-dimension, to allow the direct use of modern non-linear learning techniques (such as SVM or Adaboost) and to ease the fundamental problem of fusion of multimodal sources (*eg* multimodal similarities).

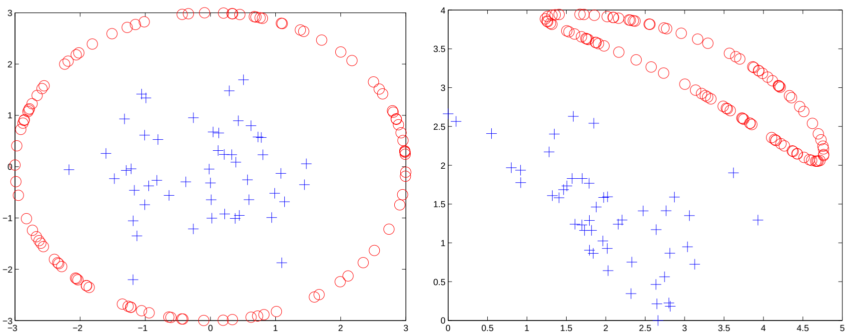
In this paper, we discuss another nice property of the QDS making the approach attractive for content-based retrieval. We demonstrate indeed how QDS overcomes the famous problem of asymmetrical classification due to the ill-definition of the negative class during retrieval. This theoretical study is supported by experimental comparisons with a kernel-based technique and the dedicated Biased Discriminant Analysis approach proposed by Zhou *et al* [13]. The overall results obtain on artificial data and collection of images indicate the validity and the efficiency of QDS for treating asymmetrical classification problem.

2 Query by example and asymmetric classification

In a query by example retrieval system, users formulate complex queries by iteratively providing positive and negative examples in a Relevance Feedback (RF) loop. From this training data, the aim is to perform, at each step of the RF loop, a real-time classification that will select the most relevant documents. Denoting the query as the set T of positive and negative training examples, respectively noted $\mathcal{P}$ and $\mathcal{N}$ with $T = \mathcal{P} \cup \mathcal{N}$, $p = |\mathcal{P}|$ and $n = |\mathcal{N}|$, the problem is to estimate (learn) a ranking function $f(x|T)$ allotting a rank r_i for each element x_i relatively to its relevance to the sought concept.

Because the training set is provided manually by user, through a graphical interface for instance, the number of examples (positive and negative) remains usually small. As a consequence, the learning may be severely undetermined, especially when it consists in estimating complex distributions in high-dimensional space. Moreover, the ill-determination is enforced by the asymmetric nature of the classification problem: To retrieve a concept out of a collection of documents, it is generally assumed that, on the average, the positive elements (representing the sought concept) are close to each other, thus conforming a specific class distribution. On the other hand the negative examples, drawn from the "rest of the world", follow some unknown and complex distributions hardly estimable from the available sparse sampling . The figure 1.a displays an example of such setup. Learning the negative classes (the circular distribution is viewed as an undetermined number of classes) becomes an under-constrained optimization problem when the training sample is small, limiting the efficiency of traditional two-class learning machines.

Dedicated algorithms have been proposed to address the asymmetrical classification [5,11]. Among all of them, an interesting approach, named *Biased Discriminant Analysis* (BDA) [13], consists in maximizing a criterion which tends to enforce compactness of the positive class while pushing apart negative examples from the positive centroid. It results in a discriminative subspace where query is processed by retrieving nearest elements in the Euclidean neighborhood of the positive centroid. In the following sections, BDA is considered for comparison with the dissimilarity-based solution studied.



(a) (b)

Fig. 1. The $1+x$ class problem in feature space (left) and 2D dissimilarity space (right) where the representation objects are two points from the central class (cross)

3 Dissimilarity space

Let $d(\mathbf{x}_i, \mathbf{x}_j)$ be the distance between elements i and j according to their descriptors $\mathbf{x} \in \mathcal{F}$. $\mathcal{F}$ expresses the original feature space. The dissimilarity space $\mathcal{D}_\Omega$ is defined relatively to a subset $\Omega \subset \mathcal{F}$ by the mapping $\mathbf{d}(\mathbf{x}, \Omega) : \mathcal{F} \to \mathbb{R}^N$

$$\mathbf{d}(\mathbf{x}, \Omega) = [d(\mathbf{x}, \mathbf{x}_1), d(\mathbf{x}, \mathbf{x}_2), \dots\ d(\mathbf{x}, \mathbf{x}_N)].$$

The representation set $\Omega = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ is a subset of N objects from which any elements of the collection will be evaluated. The new "features" of an input element are now its dissimilarity values with the representation objects. As a consequence, learning or classification tools for feature representations are also available to deal with the dissimilarities.

The dimensionality of the dissimilarity space is equal to the size of Ω, which controls the approximation made on the original feature space (such an approximation could be computed using projection algorithms like classical scaling [6]). Increasing the number of elements in Ω increases the representation accuracy. On the other hand, a well-chosen space of low dimension would be more effective for learning processes as it avoids the *curse of dimensionality* problem and reduces the computation load. The selection of a "good" representation set may be driven by considerations on the particular learning problem we are dealing with, as shown in the next section.

4 Query-based Dissimilarity Space

In this section, we look at how the selection of the set Ω may offer us the possibility to turn the asymmetrical classification problem into a more classical formulation. As stated in section 2, we are facing a $1+x$ class setup where 1 class corresponds to positives while an unknown number x of classes are associated to negative examples. In BDA, this statement consists in finding a subspace where the distances from negatives to positives (*between* scatter) are maximized while inter-positives distances (*within* scatter) are minimized. This may be achieved by seeking some linear or non-linear projections of the original space where the following ratio will be maximized

$$J = \frac{\sum_{i \in \mathcal{P}, j \in \mathcal{N}} d(\mathbf{x}_i, \mathbf{x}_j)^2}{\sum_{i,j \in \mathcal{P}} d(\mathbf{x}_i, \mathbf{x}_j)^2}, \tag{1}$$

Then, defining the Query-based Dissimilarity Space (QDS) $\mathcal{D}_{\mathcal{P}}$ by the mapping

$$\mathbf{d}(\mathbf{x}, \mathcal{P}) = [d(\mathbf{x}, \mathbf{x}_1^+), d(\mathbf{x}, \mathbf{x}_2^+), \ldots\ d(\mathbf{x}, \mathbf{x}_p^+)] \tag{2}$$

and noting that, in QDS, the norm is

$$||\mathbf{d}_i||^2 = \sum_{j \in \mathcal{P}} d(\mathbf{x}, \mathbf{x}_j)^2,$$

the quotient J may be simply rewritten as the ratio between the sum of the negative and the positive vector norms

$$J = \frac{\sum_{i \in \mathcal{N}} ||\mathbf{d}_i||^2}{\sum_{i \in \mathcal{P}} ||\mathbf{d}_i||^2}.$$

As a matter of fact, selecting $\mathcal{P}$ as the representation set naturally embeds the data in an intrinsic discriminative space where the criterion to classify elements is simply the vector norms of elements. Therefore, optimizing any learning machines in that space to separate positive from negative samples will optimize the

BDA criterion. In other word, in $\mathcal{D}_{\mathcal{P}}$, the $(1+x)$-class learning is transformed in a classical binary setup. From a geometrical point of view, the learning task does not consist anymore in estimating a complex distribution composed of x negative classes but a simpler (eventually non-linear) function separating the positive class (close to the origin) to the rest of the space (Figure 1.b).

5 Experiments and evaluations

5.1 Kernel SVM, BDA and QDS

This experimental section proposes qualitative and quantitative assessment of the retrieval efficiency when operated in QDS, in original feature space and through the BDA algorithm. For QDS and feature space, we have to choose machine learning strategy that will estimate the ranking function introduced in section 2. In both cases, a SVM algorithm is used, where the rank of every element is obtained by sorting the SVM decision function

$$f(\mathbf{x}) = \sum_i \alpha_i k(\mathbf{x}, \mathbf{x}_i) \tag{3}$$

with $\mathbf{x}_i$ the support vectors and α_i their respective weights. The kernel $k(\mathbf{x}, \mathbf{y})$ is chosen linear for QDS in order to facilitate comparison with BDA, but is non-linear (rbf Gaussian kernel) for feature space so as to cope with the $(1+x)$ classification setup. For all the following retrieval experiments, the Gaussian scale parameter is set by cross-validation.

As far as BDA is concerned, we follow the algorithm presented in [13], where the ranking function is obtained by sorting the euclidean distances between elements and the positive centroid in the discriminative subspace.

5.2 Artificial data

A toy example The toy example depicted in figure 2.a gives an illustration of how perform the three retrieval approaches considered. In this 2D example, 3 positive and 4 negative examples are provided (`*` markers) to determine a decision function enabling to retrieve the positive samples (`+` markers) and discarding negative elements (`o` markers). Because the problem is 2D, BDA implicitly works within a 1-dimensional subspace, leading to a linear decision function not suited for the problem (figure 2.d). On the other hand, rbf SVM in feature space estimate a non-linear function, but because the $1+x$ class setup and the small number of training data, the SVM is not able to model well the positive class with respect to the negative one (figure 2.c). For linear SVM in QDS, the use of a Euclidean distance as dissimilarity measure leads also to a non-linear decision function in feature space, but because applied to a 2-class problem, the SVM is able in that case to provide a better estimation of the positive class distribution (figure 2.b).

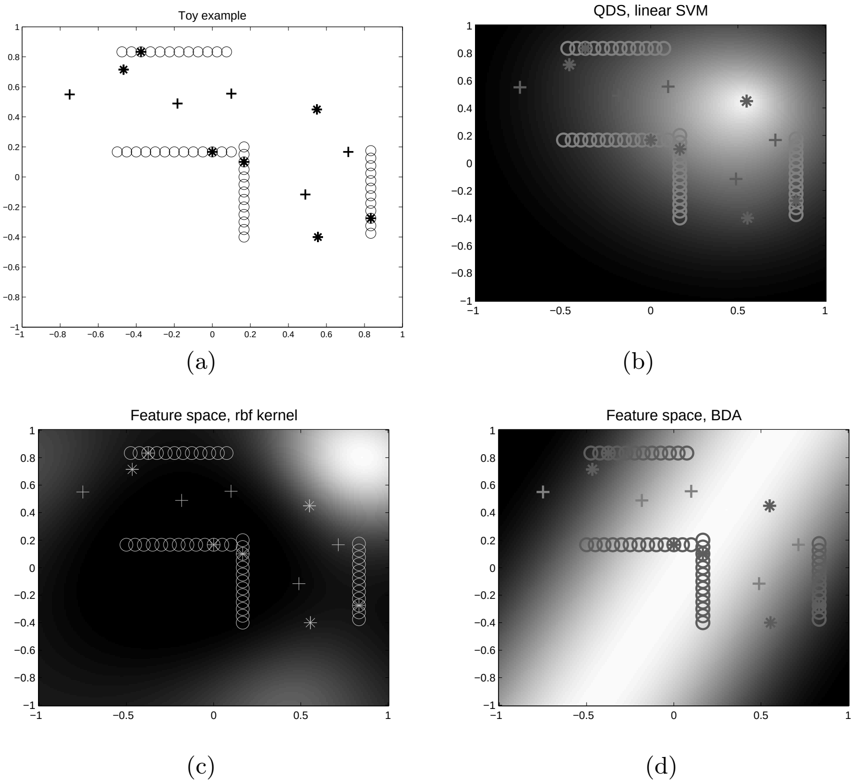
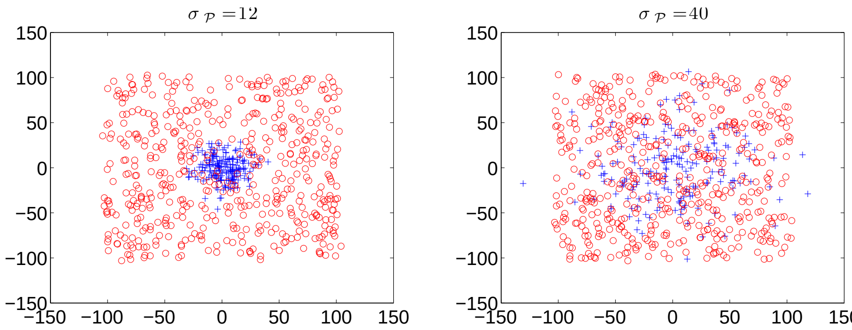


Fig. 2. Decision function on a toy example

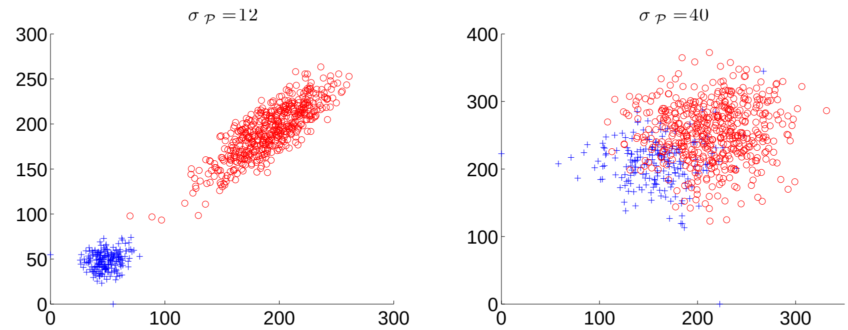
High-dimensional $1 + x$ class problem A multidimensional feature space is generated with a positive class of elements x_i^+ drawn from a centered Gaussian distribution $N(0, \sigma_{\mathcal{P}})$ and a negative class x_j^- uniformly distributed with the constraint $|x_j^-| > 10, \forall j$. The set is composed of 250 positive and 750 negatives elements.

In this setup, the positive class is effectively surrounded by negative elements uniformly distributed within the space. The positive scale $\sigma_{\mathcal{P}}$ defines how the two classes overlap each other, making the discrimination more or less difficult to be achieved. The figure 3.a displays a 2D slice of the a 50-dimension feature space for two values of $\sigma_{\mathcal{P}}$.

For the experimentation, an equal number of positive and negative examples is randomly drawn from the two classes. From this training set, an Euclidean QDS is generated by taking positive examples as representation set. The figure 3.b shows QDS built from two positive samples for the corresponding feature spaces. It is worth recalling that the dimensionality of QDS is equal to the number of positive examples p.



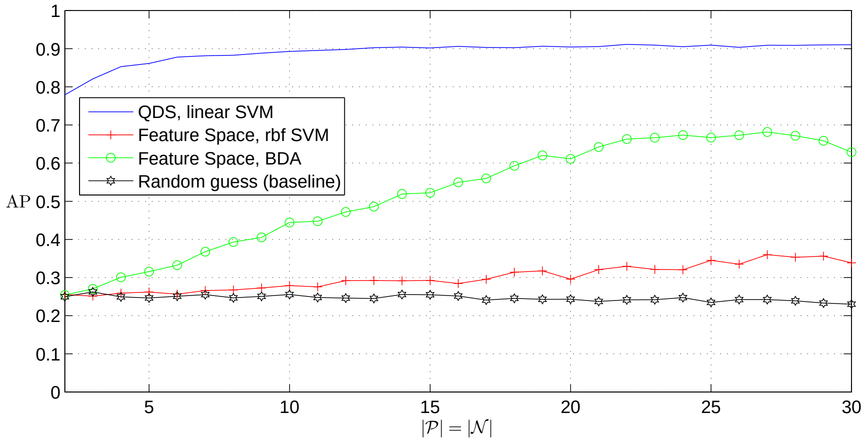
(a) Feature space



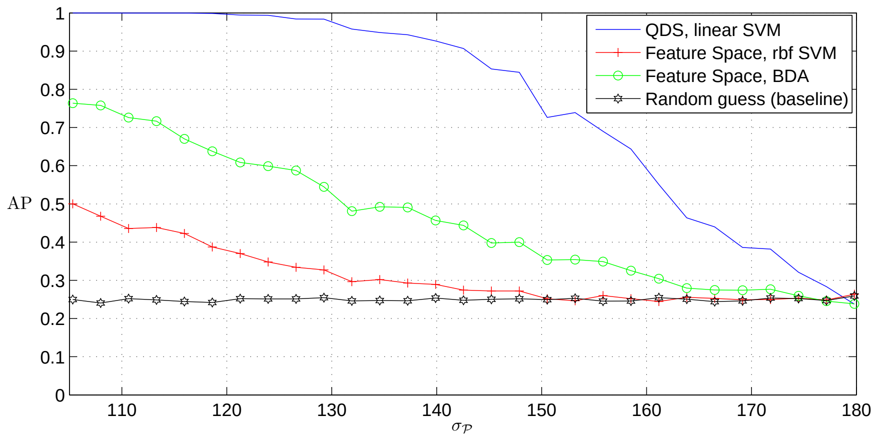
(b) Dissimilarity space

Fig. 3. Artificial data composed of positive (+ markers) and negative (o markers) elements in a) Feature space and b) the corresponding QDS build from two positive examples

The retrieval performance is measured using the Average Precision (AP) [9] computed over the entire ranked list. The measure is repeated 10 times and an averaged value of AP is given for each experimental conditions given below. The figure 4.a presents the AP measures for the three retrieval algorithms and a comparison with the baseline performance given by a random guess of elements. The artificial data are embedded in 50-dimension space and the positive class bandwidth is set to $\sigma_{\mathcal{P}} = 140$. An overlap so important between the two classes does not permit the rbf-SVM to provide results significantly better than the baseline, even when the number of examples becomes large. On the other and, BDA and QDS are able to cope with the asymmetric class setup. However, BDA suffers from the high-dimensionality of the space, especially when the small size of training set leads to a miss-estimation of the *within* and *between* covariance matrices. For QDS, the linear SVM, trained in low-dimensional space, is able to provide an efficient retrieval whatever the number of examples involved.



(a) Retrieval performance with $\sigma_{\mathcal{P}} = 140$



(b) Retrieval performance with $|\mathcal{P}| = |\mathcal{N}| = 10$

Fig. 4. Results on artificial data when a.) the number of examples increases and b.) the overlap between the positive and the negative samples growth.

The second experiment (Figure 4.b) tests the discriminative efficiency as the classes become more and more intricated. In that experiment, 10 positive and 10 negative examples are provided to the machine learning algorithms. Unsurprisingly, the QDS approach outperforms both BDA and rbf SVM. After a certain point however, the three approaches perform just like the baseline, indicating that positive samples are totally scattered within the negative elements.

5.3 Image retrieval

A last evaluation is conducted on a Corel image subset. The feature space consists in a 64 RGB histogram and embeds 18521 images annotated by several keywords. Symmetrized Kullback-Leibler divergence is taken as the dissimilarity measure for QDS. We get interested by successively retrieving images annotated with the 6 following keywords: `'whale'`, `'ice'`, `'wave'`, `'tulip'`, `'sunset'`, `'mountain'`. These keywords, somehow correlated with the low-level color descriptors extracted, have been selected to conform with the $1 + x$ classification setup.

For every keyword, 50 queries are made by selecting randomly an equal number of positive (labeled by keyword) and negative (not labeled by keyword) images. The overall evaluation is obtained by taking the mean AP over all queries for all keywords (MAP, [9]).

Figure 5 gives the MAP scores for an increasing training set. The result obtained with QDS outperforms BDA and non-linear SVM, especially when the training set becomes very small ($\backsim 1-2$ examples per class). This behavior is particularly interesting for retrieval with the RF paradigm because the very first positive examples are generally tediously gathered to build the query.

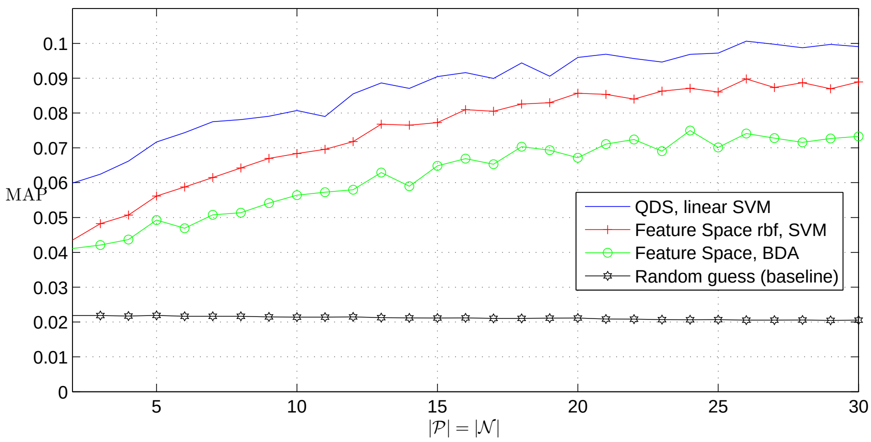


Fig. 5. Results on the Corel image set

6 Conclusion

We have presented a new similarity-based representation space for content-based multimedia retrieval. The proposed Query-based Dissimilarity Space (QDS) is adapted to cope with asymmetrical classification problems generally encountered when dealing with query by example and relevance feedback paradigms. The idea of QDS is to consider data solely from the point of view of their similarities with the positive examples provided by user. As a consequence, and as shown by experimental evaluations, learning is simplified to a binary classification problem in a low-dimensional space, leading to a more robust and efficient retrieval of relevant documents.

For the sake of evaluation, learning in QDS has been done through a simple linear-SVM. However, in order to build an effective multimedia retrieval system as the one we presented in [2], non-linear approaches and more sophisticated strategies may be enlisted to cope with real world non-linearly distributed multimodal documents.

References

1. Liudmila Boldareva and Djoerd Hiemstra. Interactive content-based retrieval using pre-computed object-object similarities. In *Conference on Image and Video Retrieval, CIVR'04*, pages 308–316, Dublin, Ireland, 2004.
2. Eric Bruno, Nicolas Moenne-Loccoz, and Stphane Marchand Maillet. Learning user queries in multimodal dissimilarity spaces. In *Proceedings of the 3rd International Workshop on Adaptive Multimedia Retrieval, AMR'05*, Glasgow, UK, July 2005.
3. E. Y. Chang, B. Li, G. Wu, and K. Go. Statistical learning for effective visual information retrieval. In *Proceedings of the IEEE International Conference on Image Processing*, 2003.
4. E. Chávez, G. Navarro, R. Baeza-Yates, and J.L. Marroquin. Searching in metric spaces. *ACM Computing Surveys*, 33(3):273–321, September 2001.
5. Y. Chen, X.S. Zhou, and T.S. Huang. One-class svm for learning in image retrieval. In *IEEE International Conference on Image Processing*, 2001.
6. T.F. Cox and M.A.A. Cox. *Multidimensional scaling.* Chapman & Hall, London, 1995.
7. D Heesch and S Rueger. Nnk networks for content-based image retrieval. In *26th European Conference on Information Retrieval*, Sunderland, UK, 2004.
8. E. Pekalska, P. Paclík, and R.P.W. Duin. A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research*, 2:175–211, December 2001.
9. A. F. Smeaton, P. Over, and W. Kraaij. Trecvid: Evaluating the effectiveness of information retrieval tasks on digital video. In *Proceedings of the ACM MM'04*, pages 652–655, New York, NY, 2004.
10. R. Yan, A. Hauptmann, and R. Jin. Negative pseudo-relevance feedback in content-based video retrieval. In *Proceedings of ACM Multimedia (MM2003)*, Berkeley, USA, 2003.
11. Ruofei Zhang and Zhongfei (Mark) Zhang. Stretching bayesian learning in the relevance feedback of image retrieval. In *proceedings of the 8th European Conference on Computer Vision (ECCV)*, Prague, Czech Republic, 2004.
12. X.S. Zhou, A. Garg, and T.S. Huang. A discussion of nonlinear variants of biased discriminant for interactive image retrieval. In *Proc. of the 3rd Conference on Image and Video Retrieval, CIVR'04*, pages 353–364, 2004.
13. X.S. Zhou and T.S. Huang. Small sample learning during multimedia retrieval using biasmap. In *Proceedings of the IEEE Conference on Pattern Recognition and Computer Vision, CVPR'01*, volume I, pages 11–17, Hawaii, 2004.