Kernel Discriminant Analysis using Triangular Kernel for Semantic Scene Classification

M. A. Tahir, J. Kittler, F. Yan and K. Mikolajczyk
Centre for Vision, Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH, UK
E-mail: {m.tahir,j.kittler,f.yan,k.mikolajczyk}@surrey.ac.uk

Abstract

Semantic scene classification is a challenging research problem that aims to categorise images into semantic classes such as beaches, sunsets or mountains. This problem can be formulated as multi-labeled classification problem where an image can belong to more than one conceptual class such as sunsets and beaches at the same time. Recently, Kernel Discriminant Analysis combined with spectral regression (SR-KDA) has been successfully used for face, text and spoken letter recognition. But SR-KDA method works only with positive definite symmetric matrices. In this paper, we have modified this method to support both definite and indefinite symmetric matrices. The main idea is to use LDL^T decomposition instead of Cholesky decomposition. The modified SR-KDA is applied to scene database involving 6 *concepts. We validate the advocated approach and demonstrate that it yields significant performance gains when conditionally positive definite triangular kernel is used instead of positive definite symmetric kernels such as linear, polynomial or RBF. The results also indicate performance gains when compared with the state-of-the art multi-label methods for semantic scene classification.*

1 Introduction

The image/video database retrieval problem involves finding in the database, instances of multimedia content that is similar to the content of interest, specified by the user. Semantic scene classification is a challenging research problem that aims to categorise images into semantic classes such as beaches, sunsets or mountains [4]. This problem can be formulated as multi-labeled classification [24] where an image can belong to more than one conceptual class such as sunsets and beaches at the same time.

Multilabel classification methods can be divided into two different groups [24]: i) *problem transformation* methods ii) *algorithm adaptation* methods. Problem transformation methods aim to transform multilabel classification problem into one or more single-label classification [4, 9], or label ranking [5] tasks. Algorithm adaptation methods extend traditional learning classifiers in order to handle multi-label directory [8, 10, 27]. This paper deals with the former group of methods.

Binary relevance (BR) learning is the most widely-used problem transformation method and is adopted in this paper. It considers the prediction of each label as an independent binary classification task [5]. The original data set is divided into $|Y|$ data sets where $Y = \{1, 2, ..., q\}$ is the finite set of labels. BR learns one binary classifier $h_a : X \rightarrow \{\neg a, a\}$ for each concept $a \in Y$.

Linear Discriminant Analysis (LDA) [12] is one of the most widely used statistical methods that has proved successful in many classification problems. LDA as well as SVM are linear classifiers learned by optimizing different criteria. Nonlinear extensions through implicit mapping of the original nonlinear observations into a high dimensional feature space (kernel trick) are widely used and the same mapping functions (kernels) can be used for both methods. Kernel Fisher Discriminant Analysis [17] and Generalized Discriminant Analysis [2] are the most popular kernel-based extensions of LDA. However, the main issue with these approaches is the singularity and the complexity of eigen-value decomposition, in particular for large datasets in image or video retrieval. Regularization techniques [2] or generalised singular value decomposition [19] can handle singularities while greedy approximation [18] or QR decomposition [26] can speed-up eigen-decomposition but more promising approach was introduced in [6] in the context of multi-class face, text and spoken letter recognition. They have proposed to use Spectral Regression which combines the spectral graph analysis and regression for an efficient large matrix decomposition in KDA. It has been demonstrated in [6] that KDA using Spectral Regression (SR-KDA) can achieve an order of magnitude speedup over the eigen-decomposition while producing smaller error rate when compared with state-of-the-art classifiers such as SVM.

SR-KDA proposed by Cai et al [6] requires positive definite matrix as Cholesky decomposition is used for complex eigen-value decomposition. That means positive definiteness has to be checked for kernels to be suitable in this method. But there are alternative kernels called conditional positive definite or indefinite kernels that have drawn attention during the last decade and proved successful in image recognition [3].

LDL^T decomposition is applicable to symmetric matrices which are not positive definite. In this paper, we propose to use LDL^T decomposition instead of Cholesky decomposition and thus the modified SR-KDA can be used for positive definite, conditionally positive definite and indefinite symmetric matrices. The modified SR-KDA is then applied to scene database involving 6 concepts. We validate the advocated approach and demonstrate that it yields significant performance gains when conditionally positive definite triangular kernel is used instead of positive definite symmetric kernels e.g linear, polynomial or RBF. The results also indicate performance gains when compared with the state-of-the art multi-label methods for semantic scene classification. The other strong point of using SR-KDA for multilabel classification using positive definite kernel such as RBF is that the time complexity scales linearly with respect to the number of labels $|Y|$. The main computationally intensive operation is either Cholesky or LDL^T decomposition, which is actually independent of $|Y|$.

The paper is organised as follows. In section 2, we review state-of-the-art methods for semantic scene classification. Section 3 discusses kernel discriminant analysis using spectral regression along with proposed modification to support indefinite symmetric matrices. Experiments are discussed in Section 4 followed by the results and discussion in Section 5. Section 6 concludes the paper.

2 Related Work

In this section, we briefly review state-of-art-methods for semantic scene classification. Since, we have used two sets of measures for the performance study, methods for these evaluation metrics are discussed separately. The first measure is an image ranking measure used in classical image retrieval for individual concepts. The second set of measures are the label rank based measures used frequently to compute the performance of multi-label learning system [21]. These evaluation metrics are described in section 4.3.

2.1 Individual Concept Ranking Measure

Most image/video database methods in the literature adopt the detection of each concept as a binary classification problem and use discriminative machine learning solutions to learn individual concepts. For example in [22], the goal is to detect the presence of 101 semantic concepts in videos. In this discriminative setting, SVMs are employed. [14] considers the situation where each image/video can take multiple class labels, i.e., the multi-label problem. However, the underlying probabilistic model is still discriminative. More examples of using discriminative machine learning techniques for image/video classification can be found in [7, 11, 15, 20].

2.2 Multi-label Ranking Measures

Multilabel scene classification is a rapidly developing field and numerous methods have been proposed to solve this problem [4, 9, 5, 8, 10, 27]. In this section, we briefly review some state-of-the-art methods used in this paper for comparison. Binary Relevance (BR) and Label Powerset (LP) learning are the most widely transformation methods for multilabel classification. BR considers the prediction of each label as an independent binary classification task while LP considers each different subset of Y as a single label and then learns only a single classifier. In [23], two variants of the k-nearest neighbour classifier (BRkNNa) and (BRkNNb) are proposed with BR transformation. In [13], multilabel classification is performed via calibrated label ranking. The key idea in their method (CMLPC) is to introduce an artificial calibration label that, in each example, separates the relevant label from the irrelevant labels. Linear Perceptron was used as base classifier in CMLPC. In [27], a lazy learning approach (ML-KNN) is proposed.

3 Kernel Discriminant Analysis using Spectral Regression (SR-KDA)

Kernel Discriminant Analysis is a nonlinear extension of LDA [12] which maps the original nonlinear measurements into a higher dimensional space thus using the "kernel trick" in a similar way to SVM. Let $\mathbf{x}_i$ be training vectors $\mathbf{x}_i \in \mathcal{R}^d, i = 1, \cdots, m$. K is an $m \times m$ kernel matrix and q is the number of class labels. If ν denotes a projective function in the kernel feature space, then the objective function for KDA is

$$\max_{\nu} \quad D(\nu) = \frac{\nu^T C_b \nu}{\nu^T C_t \nu} \tag{1}$$

where C_b and C_t denote the between-class and total scatter matrices in the feature space respectively. Equation 1 can be solved by the eigen-problem $C_b = \lambda C_t$. It is proved in [2] that equation 1 is equivalent to

$$\max_{\alpha} \quad D(\alpha) = \frac{\alpha^T K W K \alpha}{\alpha^T K K \alpha} \tag{2}$$

where $\alpha = [\alpha_1, \alpha_2, ...\alpha_m]^T$ is the eigen-vector satisfying $KWK\alpha = \lambda KK\alpha$. K is the kernel matrix and $W = (W_l)_{l=1,...,n}$ is a $(m \times m)$ block diagonal matrix of labels arranged such that upper block corresponds to positive examples and lower one to negative examples of the class. Each eigenvector α gives a projection function ν in the feature space.

It is proved in [6] that instead of solving the eigen-problem in Eqn. 2, the KDA projections can be obtained by the following two linear equations

$$\begin{aligned} W\phi &= \lambda\phi \\ (K + \delta I)\alpha &= \phi \end{aligned} \tag{3}$$

where ϕ and α are the eigenvectors of eigen-problem, I is the identity matrix and $\delta > 0$ is the regularisation parameter. Eigen-vectors ϕ are obtained directly from Gram-Schmidt method. Since $(K + \delta I)$ is positive definite, the Cholesky decomposition is used to solve the linear equations from Eqn 3. Thus, SR-KDA only needs to solve a set of regularised regression problems [6] and there is no eigenvector computation involved. This results in great improvement of computational cost and allows to handle large kernel matrices.

3.1 SR-KDA for indefinite symmetric matrices

One of the assumptions in SR-KDA is that the matrix $(K + \delta I)$ must be positive definite as Cholesky decomposition is used to solve the linear equations from Eqn 3. In this paper, we propose to use LDL^T decomposition instead of Cholesky decomposition. LDL^T-decomposition is applicable to symmetric matrices which are not positive definite. As opposed to Cholesky decomposition, which is applicable only for symmetric positive definite matrices, LDL^T-decomposition exists any each symmetric matrix [1].

In LDL^T decomposition, D is a diagonal matrix, L is a lower triangular matrix with ones on its diagonal and LT denotes its transpose. If $(K + \delta I)$ is non singular, so is L [16]. In this case L and D are uniquely determined by the matrix $(K + \delta I)$. It should be noted that the LDL^T-decomposition is faster than the LU-decomposition e.g. singular value decomposition used in original KDA but slower than the Cholesky decomposition, so it is recommended to use the latter whenever the matrix is symmetric positive definite.

3.2 SR-KDA for Multi-label Scene Classification

In this paper, we propose to use modified SR-KDA for multi-label scene classification. Binary Relevance (BR) transformation discussed in Sections 1 and 2 is used as transformation method i.e. for each concept, a SR-KDA classifier is learned. Among the strong points of using SR-KDA for multilabel classification is that its time complexity scales linearly with respect to $|Y|$. The main computationally intensive operation is either Cholesky or LDL^T decomposition, which is actually independent of $|Y|$.

3.3 Triangular Kernel

Normally, the classification power of kernel classifiers comes directly from the complexity of the underlying kernels. In this paper, we have used standard RBF kernel along with triangular kernel. These kernels are defined as follows:

$$K_{Triangular}(x_i, x_j) = -||x_i - x_j||^\beta \tag{4}$$

$$K_{RBF}(x_i, x_j) = exp(-\gamma \times ||x_i - x_j||^2) \tag{5}$$

4 Experimental Setup

A real-world Scene data which has been studied in the literature is used for experiments [4]. In scene data set, 6 categories are used. This data set is divided into 1211 training samples and 1196 test samples. The feature vector consists of 294 features. Detailed description about how these features are computed can be found in [4]. Table 1 show the ground truth in Scene data set.

Concept	Ground Truth		Concept	Ground Truth	
	Train	Test		Train	Test
Beach	227	200	Mountain	196	235
Fall Foliage	199	165	Sunset	277	256
Field	197	200	Urban	207	204

Table 1. Ground Truth in Scene Data set.

4.1 Benchmark Methods

The modified SR-KDA method is evaluated using 5 kernels (Triangular, Pseudo-Euclidean, RBF, Polynomial and Linear) and compared with the state of the art SVM classifier. For SVM, all except triangular and pseudo-euclidean kernels are used. Since, conventional SVM also requires positive definite kernels, it is not possible to evaluate it using these kernels. However, in [3], conditionally positive definite kernels are used with SVM under some assumptions and we will investigate then in future. Furthermore, the modified SR-KDA is also compared with state-of-the-art methods for multi-label classification namely BR-SVM [4], CMPLC-LP [13], BR-KNN_a [23], BR-KNN_b [23], ML-KNN [27].

4.2 Implementation

In our experiments we used the SVM implementation from the publicly available machine learning toolbox

SHOGUN[1] and Matlab implementation of Spectral Regression for KDA from [6]. The original matlab implementation of SR-KDA is modified to support kernels other than positive definite. All kernel functions (Linear, Polynomial, RBF, Triangular and Pseudo-Euclidean) are implemented from kernel toolbox KerMet [2]. In addition, multi-label classification methods BR-KNN$_a$, BR-KNN$_b$ and ML-KNN are implemented via publicly available Java package based on WEKA [25] [3].

4.3 Evaluation Measures

We used two sets of measures for the performance study. The first set of measures comprises image ranking measures used in classical image retrieval for individual concepts. The second set of measures are label rank based measures used frequently to compute the performance of multi-label learning system [21].

4.3.1 Individual Concepts Ranking Measure (Average Precision):

The average precision is a single-valued measure that is proportional to the area under a precision-recall curve. This value is the average of the precision over all relevant judged shots. This metric combines precision and recall into one performance value. This measure is computed from the ranking list of all the key frames in the database established by ordering their similarities to a specified concept. Average Precision for each concept (AP) is defined as

$$AP = \frac{1}{|R|} \sum_{k=1}^{|R|} c_k \tag{6}$$

where R is the number of positive samples in a test set and the contribution c_k of the k^{th} element in the ranking list is defined as

$$c_k = \left\{ \begin{array}{lll} \frac{|R \cap M_k|}{k} & if & concept\ true \\ 0 & if & concept\ not\ true \end{array} \right. \tag{7}$$

where $M_k = \{i_1, i_2, ..., i_k\}$ is a ranked list of the top k retrieved samples from the test set.

4.3.2 Multi-label Ranking Measures

Hamming Loss, One Error, Coverage, Ranking Loss and Average Precision are the most popular ranking measures for multi-label classification and are used in this paper for comparison with different methods. The details can be found in [21, 13, 27]. Hamming loss computes the percentage of labels that are misclassified i.e. relevant labels that are not predicted or irrelevant labels that are predicted. One Error evaluates how many times the top-ranked label was not in the set of possible labels. For single label classification problems, the one-error is similar to ordinary error. Coverage assesses performance of a system for all the possible labels of images. For single-label classification problems, the coverage is the average rank of the correct label and is zero if the system does not make any classification errors. Ranking Loss evaluates the average fraction of label pairs that are reversely ordered for the instance. The smaller the values of Hamming Loss, One Error, Coverage and Ranking Loss, the better is the performance. Average Precision is frequently used in information retrieval systems to evaluate the image ranking performance for query retrieval and discussed in section 4.3.1. Nevertheless, it is used here to measure the effectiveness of the label rankings. The higher the value of average precision, the better is the performance.

5 Results and Discussion

Table 2 shows the average precision (AP) for each concept using SR-KDA/SVM classifiers and Triangular, Pseudo-Euclidean, RBF, Polynomial and Linear kernels. From the results, it is cleared that the triangular kernel using SR-KDA performs best in all concepts. Overall in mean average precision (MAP), there is 6.1% improvement when compared with RBF kernel of SR-KDA and 7.7% improvement when compared with polynomial kernel of SVM. It should be noted that the best performance is achieved with SVM using polynomial kernel. It is also worth noting that pseudo-euclidean indefinite symmetric kernel performs quite poorly compared to other kernels, but we would like to stress that it has been shown here to indicate the possibility of using indefinite symmetric kernels using modified SR-KDA.

When SR-KDA and SVM are compared using same kernels, SVM has better performance for polynomial and linear kernels while SR-KDA performs higher for RBF kernel in terms of MAP. Table 2 also indicates that when SVM and SR-KDA are compared using same kernels like RBF, the performance varies for different object categories. For example, using RBF kernel, SR-KDA has better performance in beach, fall foliage, field and sunset categories while SVM performs better in mountain and urban.

Table 3 shows the multi-label evaluation measures using various kernels and SR-KDA/SVM. The results clearly indicate that the triangular kernel using modified SR-KDA outperforms all other kernels in all ranking evaluation criteria. The triangular kernel has a HammingLoss, One-Error, Coverage, Ranking Loss and Average Precision of 0.0769, 0.1840, 0.4423, 0.0667 and 0.8860 respectively while the

[1] http://www.shogun-toolbox.org/
[2] http://lmb.informatik.uni-freiburg.de/people/haasdonk/KerMet-Tools/
[3] http://mlkd.csd.auth.gr/multilabel.html

Concept	SR-KDA					SVM		
	Triangular	PE	RBF	Poly	Linear	RBF	Poly	Linear
	$\beta = 1.0$	$p = 1$	$\gamma = 0.5$	$deg = 2$		$\gamma = 0.5$	$deg = 2$	
Beach	**0.8714**	0.6524	0.8471	0.8242	0.6910	0.7722	0.7852	0.7609
Fall Foliage	**0.9803**	0.8546	0.9667	0.9539	0.8958	0.9600	0.9653	0.9565
Field	**0.9492**	0.5715	0.8985	0.8425	0.6176	0.8733	0.8947	0.8376
Mountain	**0.9252**	0.7840	0.8934	0.8756	0.805	0.8958	0.8995	0.8925
Sunset	**0.6762**	0.4261	0.5870	0.5641	0.4748	0.5619	0.5663	0.5408
Urban	**0.7678**	0.4449	0.6799	0.6234	0.5036	0.6888	0.7184	0.6639
MAP	**0.8617**	0.6223	0.8121	0.7806	0.6646	0.7920	0.8000	0.7754

Table 2. Comparison of Average Precision for Individual Concepts. PE = Pseudo-Euclidean.

Concept	SR-KDA					SVM		
	Triangular	PE	RBF	Poly	Linear	RBF	Poly	Linear
HammingLoss ↓	**0.0769**	0.1421	0.0929	0.0998	0.1364	0.1148	0.0971	0.1168
One-Error ↓	**0.1840**	0.3687	0.2107	0.2258	0.3261	0.2207	0.2174	0.2224
Coverage ↓	**0.4423**	0.8972	0.5477	0.5920	0.7884	0.5059	0.4866	0.5159
RankingLoss ↓	**0.0667**	0.1563	0.0870	0.0961	0.1352	0.0803	0.0765	0.0823
AvgPrecision ↑	**0.8860**	0.7639	0.8645	0.8539	0.7927	0.8642	0.8679	0.8623

Table 3. Comparison of multi-label ranking measures for various kernels using modified SR-KDA and SVM. PE = Pseudo-Euclidean. For each evaluation criterion, ↓ indicates "the smaller the better" while ↑ indicates "the higher the better"

second best is polynomial kernel using SVM with values 0.0971, 0.2174, 0.4866, 0.0765 and 0.8679 respectively. In summary, by using conditional positive definite kernel such as triangular, much better performance is achieved.

Table 4 shows the comparison of modified SR-KDA using triangular kernel with state-of-the art multi-label classification methods. SR-KDA has outperformed all other methods in all evaluation measures. SR-KDA gives an improvement of approx. 15%, 2%, 8.5% and 2% in hamming loss, coverage, ranking loss and average precision respectively when compared with nearest best multi-label method CMPLC-LP. Similarly, there is improvement of approx. 15% for one-error when compared with BR-SVM.

5.1 Execution Time

Table 5 shows the training time for original and modified SR-KDA. As discussed in Section 1, one binary classifier is needed for each label. All the experiments have been performed on a $16 \times 3GHz$ hyperthreaded CPUs and 128 GBytes of memory. Since there are 6 concepts in scene data set, original SR-KDA trains 6 classifiers in approx. 0.24 seconds while modified SR-KDA requires approx. 0.36 s to train 6 classifiers. This additional time is due to the diagonal matrix D in LDL^T decomposition which requires an additional linear equation to solve. The table clearly indicates that the main computationally intensive operation is either Cholesky or LDL^T decomposition for positive definite kernel which is actually independent of the number of labels $|Y|$. For triangular kernel, since diagonal matrix D is not sparse, it requires more time to train each concept. It should be noted that the LDL^T decomposition is slower than the Cholesky decomposition, so it is recommended to use the latter whenever the matrix is symmetric positive definite.

	Original SR-KDA	Modified SR-KDA	
	RBF	RBF	Tria
Decomposition	0.12	0.25	0.25
Training (6 classifiers)	0.24	0.36	4.00
Total	0.36	0.61	4.25

Table 5. Execution Time in seconds for original and modified SR-KDA.

6 Conclusions

Kernel discriminant analysis using spectral regression (SR-KDA) is modified in this paper to support both positive definite and indefinite symmetric matrices. The main idea is to use LDL^T decomposition instead of Cholesky decomposition. The modified SR-KDA using triangular kernel is applied to scene database involving 6 concepts and evaluated using various ranking measures. The results clearly indicate its advantage over other approaches. In this paper, Binary relevance (BR) learning is used as a problem transformation method for multi-label classification. Future research aims to use other problem transformation methods such as Label Powerset (LP) for semantic scene classification.

Acknowledgement: This work was supported by the EU VIDI-Video Project.

	BR-SR-KDA	BR-SVM	CMPLC-LP [13]	BR-KNN$_a$ [23] $k=17$	BR-KNN$_b$ [23] $k=17$	ML-KNN [27] $k=15$
HammingLoss $\downarrow$	**0.0769**	0.0971	0.0903	0.1066	0.1080	0.0974
One-Error $\downarrow$	**0.1840**	0.2174	0.2243	0.2918	0.2918	0.2425
Coverage $\downarrow$	**0.4423**	0.4866	0.4512	0.6279	0.6279	0.5217
RankingLoss $\downarrow$	**0.0667**	0.0765	0.0729	0.1170	0.1170	0.0837
AvgPrecision $\uparrow$	**0.8860**	0.8679	0.8670	0.8232	0.8232	0.8546

Table 4. Comparison of Multi-label Measures for different methods.

References

[1] http://www.alglib.net/linequations/symmetric/ldlt.php.

[2] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 2(12):2385–2404, 2000.

[3] S. Boughorbel, J. P. Tarel, and N.Boujemaa. Conditionally positive definite kernels for SVM based image recognition. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME'05)*, Amsterdam, The Netherlands, 2005.

[4] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown. Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–71, 2004.

[5] K. Brinker, J. Furnkranz, and E. Mullermeier. A unified model for multilabel classification and ranking. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06)*, pages 489–493, 2006.

[6] D. Cai, X. He, and J. Han. Efficient kernel discriminat analysis via spectral regression. In *International Conference on Data Mining (ICDM 07)*, Omaha, NE, October 2007.

[7] E. Chang, K. Goh, G. Sychay, and G. Wu. Cbsa: Content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38, 2003.

[8] A. Clare and R. King. Knowledge discovery in multi-label phenotype data. In *Proceedings of the 5th European Conference on Principles of Data Mining and Knwoledge Discovery (PKDD 2001)*, pages 42–53, Freiburg, Germany, 2001.

[9] S. Diplaris, G. Tsoumakas, P. Mitkas, and I. Vlahavas. Protein classification with multiple algorithms. In *Proceedings of the 10th Panhellenic Conference on Informatics*, pages 448–456, Volos, Greece, 2005.

[10] A. Elisseeff and J. Weston. A kernel method for multi-labelled classification. *Advances in Neural Information Processing Systems*, 14, 2002.

[11] J. Fan, Y. Gao, and H. Luo. Multi-level annotation of natural scenes using dominant image components and semantic concepts. In *Proceedings of ACM Multimedia Conference*, pages 540–547, 2004.

[12] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2, 1990.

[13] J. Frnkranz, E. Hllermeier, E. L. Menca, and K. Brinker. Multilabel classification via calibrated label ranking. *Machine Learning*, 23(2):133–153, 2008.

[14] K. Goh, E. Chang, and B. Li. Using one-class and two-class SVMs for multiclass image annotation. *IEEE Transactions on Knowledge and Data Engineering*, 17(10):1333–1346, 2005.

[15] K. Goh, B. Li, and E. Chang. Semantics and feature discovery via confidence-based ensemble. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 1(2):168–189, 2005.

[16] J. P. May, B. D. Saunders, and D. H. Wood. Numerical techniques for computing the inertia of products of matrices of rational numbers. In *Proceedings of the 2007 international workshop on Symbolic-numeric computation*, pages 125–132, 2007.

[17] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller. Fisher discriminant analysis with kernels. In *IEEE Neural Networks for Signal Processing Workshop (NNSP)*, 1999.

[18] S. Mika, A. Smola, and B. Scholkopf. An improved training algorithm for kernel fisher discriminants. In *AISTATS 2001*. Morgan Kaufmann, 2001.

[19] H. P. Peg Howland. Generalizing discriminant analysis using the generalized singular value decomposition. *PAMI*, 26(8):995–1006, 2004.

[20] K. Sande and T. G. adn C. Snoek. A comparison of color features for visual concept classfication. In *ACM Iternational Conference on Image and Video Retrieval*, 2008.

[21] R. E. Schapire and Y. Singer. Boostexter: A boosting based system for text categorization. *Machine Learning*, 39(2-3):135–168, 2000.

[22] C. G. M. Snoek, M. Worring, J. C. van Gemert, J. M. Geusebroek, and A. W. M. Smeulders. The challenge problem of automated detection of 101 semantic concepts in multimedia. In *Proceedings of the ACM International Conference on Multimedia*, 2006.

[23] E. Spyromitros, G. Tsoumakas, and I. Vlahavas. An empirical study of lazy multilabel classification algorithms. In Springer, editor, *5th Hellenic Conference on Artificial Intelligence (SETN 2008)*, Syros, Greece, 2008.

[24] G. Tsoumakas and I. Katakis. Multi-label classification: An overview. *International Journal of Data Warehousing and Mining*, 3:1–13, 2007.

[25] I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2005.

[26] T. Xiong, J. Ye, and Q. L. amd V. Cherkassky amd R. Janardan. Efficient kernel discriminant analysis via QR decompostion. *Advances in Neural Information Processing Systems*, 1(17), 2004.

[27] M. L. Zhang and Z. H. Zhou. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048, 2007.